DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

ESSAYS ON EDUCATIONAL INTERVENTIONS FOR DISADVANTAGED CHILDREN

PhD dissertation

Søren Albeck Nielsen

Aarhus University, BSS
Department of Economics and Business Economics

2021

PREFACE

This dissertation was written from September 2018 to December 2021 at the Department of Economics and Business Economics at Aarhus University. I would like to thank the department and TrygFonden's Centre for Child Research for providing a truly excellent research environment and for their generous financial support.

First and foremost, I wish to express my sincere gratitude to my supervisors Michael Rosholm and Marianne Simonsen, for their support, kindness, encouragement, and guidance throughout this period. I enjoyed and learned a lot from their various comments and suggestions. You have pushed my projects further than I was willing to go on my own and shown me the pathway into academics. A special thanks to Charlotte for a great cooperation on running the RCT and to Michael for trusting us to handle a large-scaled randomized experiment, which is an opportunity only few get and was a great learning experience.

In the fall of 2021, I was visiting FAIR at NHH. I am grateful for Aline Bütikofer and Kjell Salvanes for hosting me in difficult time and for introducing me to the excellent research environment. A special thanks to Rene and Daniel who made my stay truly pleasurable.

Huge thanks to the secretariat and management team at TrygFonden's Centre for Child research. I enjoy our conversations at lunch and on coffee runs as well as your open door policy where you always are willing to help.

Furthermore, I would like to thanks friends, co-authors, faculty, and fellow PhD students for three good years with a lively work environment and for sharing your experiences and advices. It has been a pleasure to have daily discussions with you about economics and everything else. My office mates Simon deserve special thanks. I have spent many days with you discussing economic questions, machine learning, and cycling. I hope we can continue to have all of this in the future.

My family deserves huge thanks as they supported me through ups and downs. Above all my partner, Katja. You have been invaluable throughout this dissertation - especially when a global pandemic changed my stay aboard from a family trip to me going alone to Bergen with you staying home with the kids. Finally, I would like to thank my daughters Agnes and Johanne for putting everything into perspective. You have kept me busy at work, as the afternoons and weekends have been reserved for important family time.

Søren Albeck Nielsen
Aarhus, December 2021

UPDATED PREFACE

The predefence took place on January 12, 2022. The assessment committee consists of Associate Professor Maria Knoth Humlum (Aarhus University), Professor Mette Ejrnæs (University of Copenhagen), and Associate Professor Joshua Goodman (Boston University). I am grateful for their careful reading of my dissertation and their valuable comments and suggestions. Some of their suggestions for improvement have already been implemented while others remain for future research.

Søren Albeck Nielsen
Aarhus, February 2022

CONTENTS

BRIEF SUMMARY

The aim of this thesis is to investigate educational remedial programs targeting disadvantaged children in lower secondary education. There exists large discrepancies within the OECD countries in the allocation of resources to the school systems with Denmark ranking at the very top. Despite this, Denmark, as well as many other OECD countries, struggles with a large tail of underperforming pupils who do not obtain the necessary level of proficiency in reading and math.

The dissertation comprises three self-contained chapters within economics of education. Each chapter empirically studies different educational interventions targeting disadvantaged children. The first two chapters make use of a field experiment as well as quasi-experimental methods combined with Danish administrative records with the purpose of identifying causal effects of intensive learning camps for grade 8 pupils at-risk of not being able to complete upper secondary education. The third chapter exploits observational methods to evaluate a dyslexia intervention.

The first chapter investigates an intensive learning camp implemented in the regular school hours and shows positive short-run effects on test scores in math, but no short-run effects on test scores in Danish language. Further, we find indications of positive long-run effects on the school-leaving exam in math and on enrollment in post-compulsory education. Finally, the analysis finds no evidence that training non-cognitive skills affects academic outcomes.

The second chapter exploits data from a two-week Danish summer camp with a one-year follow-up program. Using difference-in-differences and triple-differencing methods, I find that the summer camp has positive effects on academic and personal competencies. Furthermore, I find that changing the follow-up program from individual to group mentoring improves personal and social competencies.

The final and third chapter studies a 10-week learning program with 18 months follow-up for children with severe dyslexia using difference-in-differences techniques. Surprisingly, there is limited evidence on the effect of dyslexia learning programs in general, although 8% of the Danish population suffer from dyslexia. I find positive and persistent effects on reading abilities, personality traits, and school well-being, such that participating in the intervention reduces the gap to similar non-dyslexics with up to 80%.

KORT RESUMÉ

Formålet med denne afhandling er at undersøge uddannelsesinterventioner rettet mod udfordrede børn og unge i udskolingen. Der er store forskelle mellem OECD-landene i tildelingen af ressourcer til folkeskolen med Danmark i toppen af fordelingen. På trods af dette kæmper Danmark, såvel som mange andre OECD-lande, med en stor gruppe af underpræsterende elever, der ikke opnår det nødvendige færdighedsniveau i læsning og matematik.

Denne afhandling omfatter tre selvstændige kapitler inden for uddannelsesøkonomi. Hvert kapitel undersøger, ved hjælp af empiriske metoder, forskellige uddannelsesinterventioner rettet mod udfordrede børn og unge. De to første kapitler anvender et lodtrækningsstudie og kvasi-eksperimentelle metoder kombineret med danske administrative registre med det formål at identificere årsagseffekter af intensive læringsforløb for 8. klasses elever, der er i risiko for ikke at kunne gennemføre en ungdomsuddannelse. Det tredje kapitel anvender observationelle metoder til at evaluere en ordblindeindsats.

Første kapitel undersøger effekten af intensiv læringsforløb implementeret i den almindelige skoletid. Analysen af lodtrækningsforsøget viser positive kortsigtede effekter i matematik, men ingen kortsigtede effekter i dansk. Ydermere finder vi indikationer på positive langsigtede effekter på afgangseksamen i matematik og på optag på ungdomsuddannelse. Endelig finder analysen ingen evidens for, at træning af ikke-kognitive færdigheder påvirker akademiske resultater.

Det andet kapitel udnytter data fra en to-ugers sommerskole med et 12 måneders opfølgningsprogram. Ved at bruge en difference-in-difference strategi, der sammenligner sommerskole drenge med den resterende population af drenge i perioden fra 8.klasse til 9.klasse, og triple differencing finder jeg, at sommerskolen har positive effekter på deltagernes faglige og personlige kompetencer. Ydermere viser jeg, at en ændring af opfølgningsprogrammet fra individuel til gruppevejledning forbedrer de personlige og sociale kompe-

tencer.

Det sidste og tredje kapitel undersøger et 10-ugers læringsprogram med 18 måneders opfølgning for børn med svær ordblindhed. Metodisk anvender jeg en difference-in-difference strategi til at sammenligne ordblinde elever der modtager indsatsen med andre ordblinde over tid. Overraskende er der begrænset evidens for effekten af læringsprogrammer rettet mod ordblinde, selvom 8% af den danske befolkning lider af ordblindhed. Jeg finder positive og vedvarende effekter på læseevner, personlighedstræk og skoletrivsel. Deltagelse i interventionen reducerer således afstanden til sammenlignelige ikke-ordblinde med op til 80%.

SUMMARY

There exists large discrepancies within the OECD countries in the allocation of resources to the school systems, with Denmark ranking at the very top. Despite this, Denmark, as well as many other OECD countries, struggles with a large tail of underperforming pupils who do not obtain the necessary level of proficiency in reading and math (OECD, 2019). This is a problem for at least two reasons. First, it signals deficiencies in the educational system, since it does not live up to its purpose. Second, low achieving pupils often come from a disadvantaged background, which exacerbating intergenerational mobility and leads to persistent educational inequality of opportunity.

Recent literature shows that low levels of academic abilities are associated with negative school behavior and low school well-being as well as a range of negative long-run outcomes, such as low educational attainment and reduced earnings, employment, and health (Heckman and Mosso, 2014). Thus, academic low performers are a problem, not only for the children themselves, but also for society. High-quality educational remedial interventions are acknowledged as important tools for decreasing social inequality and ensuring a highly qualified labor force, which is the very foundation for future economic growth and welfare. This dissertation investigates educational remedial programs targeting disadvantaged children in lower secondary education.

The dissertation comprises three self-contained chapters within economics of education. Each chapter empirically studies different educational interventions targeting specific groups of disadvantaged children. It contributes to the literature, initiated by the 2021 Nobel Prize winners Joshua Angrist, Guido Imbens and David Card on causal impact analysis of policy initiatives, especially within educational policies targeting disadvantaged children (Angrist and Keueger, 1991; Card and Krueger, 1994; Imbens and Angrist, 1994). There is an urgent need for evidence-based policy initiatives aimed at disadvantaged children because it is tremendously expensive to

enhance their human capital (Heckman, 2006). The first two chapters make use of a field experiment and methods for observational data combined with Danish administrative records with the purpose of identifying causal effects of intensive learning camps for grade 8 pupils at-risk of not being able to complete upper secondary education. The third chapter employs a difference-in-differences framework to investigate a 10-week learning program with 18 months follow-up for children with severe dyslexia.

The first chapter is titled "An intensive, school-based learning camp targeting academic and non-cognitive skills evaluated in a randomized trial" (co-authored with Charlotte Hvidman, Alexander Koch, Julia Nafziger, and Michael Rosholm) and evaluates school-based intensive learning camps for pupils assessed "not ready" for upper secondary education using a randomized controlled trial involving 264 schools in Denmark. Despite a great number of studies on intensive learning camps, there is little causal evidence of the impacts of this particular type of camp; only few studies rely on experiments or use appropriate methods for causal inference and none of the interventions studied take place in the school. We investigate two camp variants, with the main variant targeting Danish language, math, and non-cognitive skills. The alternative variant drops the non-cognitive skills training for additional lectures of Danish language and math. Both camps consist of two weeks of learning camp and additional eight weeks of short follow-up sessions with the primary teacher. We find positive short-run effects in math but no short-run effects in Danish in which only one of two targeted areas improves. We find weak evidence of positive long-run effects on the school-leaving exam in math and on enrollment in upper secondary education. Finally, we find no evidence that targeting non-cognitive skills in this camp setting improves academic performance. Our results thus provide a new perspective on recent evidence regarding the effects of training non-cognitive skills – by running an intervention with older pupils and in a comparatively high-resource school system.

The second chapter, "Boys Left Behind: The Effects of Summer Camp and Follow-up Strategies on Academic, Personal, and Social Competencies", investigates the consequences of summer camp participation for disadvantaged boys in grade 8. I also study how mentoring strategies in the follow-up program affects outcomes. Previous studies show weak evidence of summer camps effect on academic outcomes for pupils in the transition from lower to upper secondary education (Mariano and Martorell, 2013; Jacob and Lefgren,

2004; Battistin and Schizzerotto, 2019). Yet, there is not much knowledge about the effect on non-academic outcomes. The analysis is based on individual level panel data obtained from the detailed administrative data covering all children in compulsory school. Using a difference-in-differences strategy, I find positive effects on academic and personal competencies. Furthermore, I exploit a structural change in the follow-up program to evaluate how changing from individual to group mentoring in the one-year follow-up program affects outcomes. Using a triple differences strategy, I find that group mentoring dramatically improves personal and social competencies. The results suggest that the camp itself affects academic competencies whereas the format of the follow-up program is crucial for the impacts on non-academic competencies.

In the third and final chapter, called "How to Cope with Dyslexia: The Effects of Special Education on Academic Performance, Personality Traits, and Well-being", I use Danish administrative data to study the effects of a special education intervention for pupils with severe dyslexia. The intervention consists of a 10-week learning program with 18 months follow-up for pupils in grade 4 to 8. Dyslexia is a learning disorder that affects 3-10% of the population (Snowling, 2013). It affects the ability to read and write and has potential long-run consequences through low school grades, poor educational attainment, and behavior problems (Undheim, 2009; Epnion, 2018; Einar et al., 2001). Yet, there exists no causal evidence for programs specifically targeting pupils with dyslexia. I use Danish administrative data combined with the membership list from the Danish Library and Expertise Center for people with print disabilities, which enables me to identify pupils diagnosed with dyslexia who have not participated in the intervention under study. 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-differences strategy, I find positive and persistent effects on reading abilities, personality traits, and school well-being. The intervention increases pupils' outcomes well beyond other dyslexics and significantly reduces the gap to non-dyslexics. The results show that the intervention enables pupils with severe dyslexia to participate in age-appropriate learning with their peers.

This dissertation compounds three self-contained intervention studies that all aim to improve disadvantaged children in lower secondary education life trajectories. Table 1 shows the mean of key standardized variables for each chapters' intervention group prior to intervention. It is evident from the table that the participating children are significantly disadvantaged. Across

Table 1: Mean for the intervention group across chapters

	Chapter 1	Chapter 2	Chapter 3
Language Comprehension (std.)	-0.522	-0.546	-0.823
Decoding (std.)	-0.609	-1.095	-1.475
Text Comprehension (std.)	-0.637	-0.927	-0.862
Conscientiousness (std.)	-0.515	-0.616	-0.387
Agreeableness (std.)	-0.214	-0.527	-0.22
Emotional Stability (std.)	-0.149	-0.103	-0.111

Notes: The table shows mean of key variables for the main intervention group across the three chapters. All variables are standardized to mean zero and standard deviation of zero using the full population. They are measured in the time-period up to treatment.

all variables, they are below the population average of zero with the largest differences for the academic measures. Their reading abilities are between 0.5 and 1.5 standard deviations below, indicating a significant academic backlog.

A common theme in all three chapters concerns the impacts of interventions on non-cognitive skills, which are known to have important effects on later life outcomes of the child (Durlak et al., 2010; Kautz et al., 2014). This is possible due to the existence of high-quality data from compulsory national well-being tests in public schools in Denmark (Andersen et al., 2020). Furthermore, each chapter aims to reduce the inequalities by examining not only the effects of the intervention under study but also exploiting small design variations within the interventions to provide better guidance for policy-makers, i.e. additional training of cognitive skills vs. non-cognitive skills and mentoring strategies in the follow-up program.

The first two chapters of this dissertation investigates how intensive learning camps implemented at different stages with different follow-up programs impact both academic and non-academic outcomes. This is a crucial and important piece of information for policy-makers in their policy decisions, i.e. deciding on a strategy to enhance the abilities of disadvantaged pupils, especially when considered in relation to the large discrepancies in the costs of different interventions. Chapter 1, then, finds that additional hours of non-cognitive skills training relatively to cognitive skills training, do not change the effect of a school-based intensive learning camp in a high-resource lower secondary school system. Chapter 2, in contrast, focuses on the follow-up mentoring strategies and finds that policy-makers must carefully select the follow-up program in order to enhance personal and social competencies. Chapter 3, supports this finding by showing that a 18 months individual

follow-up program to a 10-weeks learning program leads to persistent learning effects and in some cases increasing effects on personality traits and school well-being.

Bibliography

Andersen, S. C., Gensowski, M., Ludeke, S. G., John, O. P., Oct. 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 (5), 925–939.

Angrist, J. D., Keueger, A. B., Nov. 1991. Does Compulsory School Attendance Affect Schooling and Earnings?*. The Quarterly Journal of Economics 106 (4), 979–1014.

Battistin, E., Schizzerotto, A., 2019. Threat of grade retention, remedial education and student achievement: Evidence from upper secondary schools in italy. Empirical Economics 56 (2), 651–678.

Card, D., Krueger, A., Sep. 1994. Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania. American Economic Review 84 (4), 21.

Durlak, J. A., Weissberg, R. P., Pachan, M., 2010. A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. American journal of community psychology 45 (3-4), 294–309, publisher: Wiley Online Library.

Einar, H., Anders, L., Jim, S., Kenneth, H., Jan. 2001. Behaviour problems in children with dyslexia. Nordic Journal of Psychiatry 55 (4), 251–256.

Epnion, Nov. 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.

Heckman, J. J., 2006. Skill Formation and the Economics of Investing in Disadvantaged Children. Science 312, 3.

Heckman, J. J., Mosso, S., 2014. The Economics of Human Development and Social Mobility. Annual Review of Economics (6), 689–733.

Imbens, G. W., Angrist, J. D., 1994. Identification and Estimation of Local Average Treatment Effects. Econometrica 62 (2), 467–475.

Jacob, B. A., Lefgren, L., 2004. Remedial education and student achievement: A regression-discontinuity analysis. Review of economics and statistics 86 (1), 226–244, publisher: MIT Press.

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

Mariano, L. T., Martorell, P., 2013. The academic effects of summer instruction and retention in New York City. Educational Evaluation and Policy Analysis 35 (1), 96–117, publisher: SAGE Publications Sage CA: Los Angeles, CA.

OECD, 2019. Pisa 2018 results (volume i): What students know and can do. OECD Publishing, Paris.

Snowling, M. J., Jan. 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 (1), 7–14.

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 (4).

CHAPTER 1

AN INTENSIVE, SCHOOL-BASED LEARNING CAMP TARGETING ACADEMIC AND NON-COGNITIVE SKILLS EVALUATED IN A RANDOMIZED TRIAL

Charlotte Hvidman
TrygFonden's Centre for Child Research

Alexander Koch
Aarhus University and IZA

Julia Nafziger
Aarhus University and CEPR

Søren Albeck Nielsen
Aarhus University and TrygFonden's Centre for Child Research

Michael Rosholm
Aarhus University, TrygFonden's Centre for Child Research and IZA

Abstract

We evaluate school-based, intensive learning camps for pupils assessed 'not ready' for post-compulsory education, using a stratified cluster randomized trial involving 15,559 pupils in 264 schools in Denmark. Next to Danish and mathematics, the main variant targets non-cognitive skills. The alternative variant uses this time for more training in Danish and math. We find positive short-run effects in the standardized test score in math (effect sizes 0.06 – 0.2) but not in Danish, where only one of two targeted academic areas improves. We find weak evidence of positive long-run effects on the final exams in math in grade 9 and enrollment in post-compulsory education 2.5 years post-intervention. We find no evidence that training of non-cognitive skills affects academic outcomes. Our results provide a perspective on recent evidence regarding the effects of training non-cognitive skills – by running an intervention with older pupils and in a comparatively high-resource school system.

Keywords: Randomized trial, remedial education program, non-cognitive skills

JEL Codes: I21, C21, D91, I28

1.1 Introduction

All over the world, education systems struggle with a large tail of underperforming pupils. In a recent PISA study (OECD, 2019), 23 percent of pupils were below the minimum level of proficiency in reading that all children should have acquired by the end of lower secondary education. In mathematics, the share scoring below this threshold was 24 percent. Such outcomes are a problem for at least two reasons. First, it is a signal that the education system does not live-up to its purpose. Second, low achieving pupils often come from a disadvantaged background, thus exacerbating low intergenerational mobility and persistent educational inequality of opportunity.

Numerous public and private remedial education programs exist that target low-achieving pupils. Interventions that can be delivered at scale at a relatively low cost per pupil are of particular interest to policy makers, which explains the popularity of intensive learning camps where pupils train one or a few subjects intensively during a limited time period (typically 1-4 weeks). Specifically, the disruption to education caused by the Covid-19 pandemic has increased the need for effective remedial education programs that can help a large number of pupils to recover learning losses (Di Pietro et al., 2020).

The aim of this study is to evaluate, using a large stratified cluster randomized trial involving 15,559 pupils in 264 schools in Denmark, the impact of an intensive learning camp that includes novel elements compared to traditional intensive learning camps in terms of (i) being run at the pupils' school by the school's teaching staff during regular school hours and (ii) targeting both academic and non-cognitive skills.

There is reason to believe that combining these elements can strengthen the impact of intensive learning camps. First, running such a camp at a pupil's own school during regular school hours can remove potential barriers to participation. For example, Lee et al. (2006) document that many potential participants face barriers to participation in after-school or summer school programs targeting at risk pupils, most notably getting to and from the program and participation in other activities. Second, being school-based permits teachers to follow-up on the learning camp, both in the associated follow-up program and during regular teaching, and ensures longer-lasting engagement by the pupils. Third, it is widely recognized that non-cognitive skills play an important role for academic performance.[1] Recent evidence points to such non-cognitive skills being malleable over a short time frame: Alan and Ertac (2018) and Alan et al. (2019) use randomized trials to document large and lasting effects – also on academic outcomes – of training elementary school pupils in Turkey

[1] For example, Duckworth et al. (2007) demonstrates the importance of grit for academic achievement. Duckworth and Seligman (2005); Duckworth et al. (2012) shows that self-control is as important as IQ in predicting academic performance and Eskreis-Winkler et al. (2014) observe that grit predicts completing school, among other outcomes. Heckman et al. (2006) and Almlund et al. (2011) examine the predictive power of various personality measures (vs. IQ).

in non-cognitive skills related to patience (the former study) and perseverance (the latter study). Thus, the many and diverse reasons for which pupils fall behind in their learning can potentially better be addressed by complementing the commonly applied 'more of the same' training on academic skills with training on non-cognitive skills.[2]

The target group of the two-week camps were grade 8 pupils assessed as "not ready for further education".[3] The camps were conducted at the school of a participating pupil during the regular teaching hours – replacing the lessons that these pupils would otherwise have received. After the camp, pupils participated in a follow-up program where they met with a camp instructor for 1.5 hours once a week for eight weeks. The camp costs the school approx. 750$ per pupil.

We evaluate two variants of the camp. In addition to math and Danish, the main learning camp variant devoted roughly 30 percent of the time to strengthen the non-cognitive skills of pupils. For example, pupils (and teachers) learned about self-regulation strategies – such as goal setting (Locke and Latham, 1990) and mental contrasting with implementation intentions (Gollwitzer et al., 2011); and they were introduced to the concept of "growth mindsets" (Dweck, 2006), which reflects a view that ability is malleable and that success is driven by effort.

In the alternative variant of the camp, the time spent on training non-cognitive skills in the main variant was instead used for extra training in math and Danish. The motivation for testing two variants of the camp stems from the uncertainty as to whether non-cognitive skills can be manipulated during an intervention as short as the one studied here and which is aimed at 8^{th} graders rather than younger pupils. The two variants allow us to explore if the time spent training non-cognitive skills makes a difference relative to using it for additional training of academic skills.

An advantage of our study is that we can exploit Danish register data to evaluate the effects of the camps using standardized national tests. In doing so, we can circumvent issues that other studies face when relying on teachers' evaluations of pupils' skills (the treatment may affect teachers' evaluations rather than pupils' actual academic level) or tests designed by the researchers (the problem of 'teaching to the test'). The national tests were conducted briefly after the interventions and thus provide short-run effects of the interventions. In each subject, these tests measure

[2] First, there are intellectual disabilities, dyslexia/dyscalculia, and other medical/neurological causes. Second, psychological causes comprise, for example, math anxiety, low self-efficacy, and low levels of conscientiousness/grit (e.g., OECD, 2013). Third, sociological causes include primarily the learning environment and the fact that some children grow up in less advantaged circumstances than others, and this affects their ability to learn. Lastly, didactic factors reflect inadequate teaching methodologies and variations in teacher quality (e.g., Hanushek and Rivkin, 2006; Hanushek, 2011).

[3] Between a quarter and a third of a school cohort are considered to be 'not ready for further education' by the end of lower secondary education (Undervisningsministeriet, 2017a,b). Denmark has 10 years of compulsory schooling and starts counting with grade 0. So grade 8 corresponds to grade 9 in other countries.

the competencies of the pupil in different areas – some of which were targeted in the camp (such as numbers and algebra in math or text and language comprehension in Danish) and some of which were not (such as decoding in Danish or geometry and statistics and probability in math). This allows us to test both for direct effects of the camp on learning and for indirect spillover effects to other academic areas. Further, we measure the long-run effects of the camp using grades in the final exams in grade 9, which were given one year after the intervention and involved external examiners, and by looking at post-compulsory education enrollment.

We observe positive short-run effects on the targeted area in mathematics (numbers and algebra) with effect sizes ranging from 0.07 to 0.19, depending on whether we focus on the entire grade 8 cohort among which some pupils were offered the camp or the pupils actually receiving treatment. The effect sizes for overall performance in the math standardized test range from 0.06 to 0.2. In Danish, the evidence for positive effects is weaker as only one out of the two targeted areas (namely, text comprehension) is positively affected by the camp with effect sizes ranging from 0.07 to 0.18. In particular, there is no significant impact on the overall performance in the Danish standardized test. In terms of long-run effects, we do not find any effects that are consistently statistically significant at conventional levels. There is some weak indication of the camp having a positive impact on math performance in the final exams in grade 9 (effect sizes range from 0.06 to 0.18) and on being enrolled in post-compulsory education 2.5 years after the camp (0 to 8 percentage points, depending on the estimate, which should be compared with the baseline of 19 percent not being enrolled in post-compulsory education in the control group).

We also examine the effects of the camp on non-cognitive skills. Next to self-administered pre- and post-surveys that included several psychological scales, we observe in the register data how school counsellors evaluated, according to schemes provided by the ministry, a pupil's readiness for further education, and use here specifically the evaluation of the personal and social skills. We do not observe any positive effects on any of these outcome measures for the main learning camp that trained non-cognitive skills.

The measurement of non-cognitive skills with surveys and the evaluation of the pupil's readiness for further education has some disadvantages. However, there is some evidence that suggests that the lacking effects on non-cognitive skills are not merely a measurement problem. Firstly, we do not observe any spillover effects to non-targeted academic areas in the standardized national tests. If the intervention at the camps affected non-cognitive skills such as self-control, one could expect the improved non-cognitive skills to broadly enhance academic performance both in targeted and non-targeted areas. This, however, is not the case. Secondly, when comparing the two camps, we primarily find non-significant differences in outcomes between the two, once again suggesting the lacking effect of non-cognitive skills. Overall, the results thus suggest that the main camp did not affect non-cognitive skills.

The latter results are in contrast to the studies by Alan and Ertac (2018) and Alan et al. (2019). Their interventions trained non-cognitive skills for a total duration of 24 hours spread over a course of 12 weeks in Turkish primary schools. When compared to our main camp, a similar amount of time was dedicated to training non-cognitive skills, with 16 hours devoted solely for this purpose during the two-week camp and additional time for revisiting these skills during weekly sessions in the eight-week follow-up program. Thus, our intervention had a similar dosage to that of Alan and Ertac (2018) and Alan et al. (2019). One possible reason for the diverging findings across studies could stem from the different age profiles – the pupils in our study are 15-16 years old, while those in Alan and Ertac (2018) and Alan et al. (2019) are 9-10 years old. Non-cognitive skills may be more malleable in these younger kids than for the pupils in our study (cf. Kautz et al., 2014, for a discussion of differential plasticity of different skills by age). Other differences that we discuss in section 1.5 are that their interventions focused on a more narrow set of non-cognitive skills than our camp and that they implemented their interventions in a school system with fewer resources[4] compared to the Danish system.

The paper is structured as follows. Next, we discuss the related literature. Section 1.2 describes the background and design of the study, such as the teaching materials and randomization procedure. We describe the data in section 1.3 and the results in section 1.4. Section 1.5 provides a discussion of possible caveats and interpretations of our findings. The last section contains conclusions.

1.1.1 Related literature

Our study provides two main contributions. First, we evaluate, using a large randomized trial, a camp that is conducted in school, during school time by the school's regular teaching staff and is, in doing so, distinguished from the prevalent summer camps. Clean evidence for such school-based camps is limited in the literature. Related evidence, however, exists regarding learning camps during summer, small and medium group instruction in schools, and increased instruction time in school, all of which constitute elements in the school-based intensive learning camp in our study. Yet, the existing evidence is quite mixed, so it is hard to have any a priori expectations regarding the effectiveness of such a camp. Second, we contribute to the literature by including non-cognitive skills in the curriculum of an intensive learning camp. As most existing studies examining non-cognitive skills focus on longer duration programs and/or on younger pupils, the question arises whether the non-cognitive skills of middle-school pupils can be affected during an intensive learning camp.

[4] One dimension of measuring the resources used in a school system are monetary expenditures per pupil. Denmark has higher expenditures per pupils than the OECD average, Turkey lower expenditures.

Intensive learning camps. Recent meta-analyses of intensive learning camps, which are primarily summer camps taking place outside the school system, are Lauer et al. (2006), Cooper et al. (2000), and Kim and Quinn (2013).They all report statistically significant effects on academic outcomes (math and/or reading), albeit with small effect sizes.

Despite a great number of studies on intensive learning camps, there is little causal evidence. Only few studies rely on experiments (based on randomization) or use quasi-experimental methods. Among the most convincing existing evaluations of intensive learning camps are studies relying on quasi-experimental methods using a regression discontinuity design (RDD) and data from standardized tests. Mariano and Martorell (2013) exploit test score cutoffs in the assignment to a summer camp for 5th-7th graders in New York. They can track the grades of the pupils up to 2-3 years after the intervention. They find some effects of the camp on English language performance, but little effect on math performance. Matsudaira (2008) study an intensive learning camp for pupils in or above grade 3 in a large urban school district in the U.S. He finds positive effects on math and reading performance around one year after the camp. The RDD of Jacob and Lefgren (2004) is based on the Chicago Social Promotion Policy. They find positive effects on math and reading performance in the short and long run for grade 3 pupils but not for grade 6 pupils. Battistin and Schizzerotto (2019) exploit geographic variations in the implementation of mandatory summer courses for at-risk pupils in Italy. They find negative short term effects on academic performance (the marks given by teachers, final examination and a test based on the PISA tests) in vocational schools and no effects in academic schools. Using a difference-in-differences framework, Schueler et al. (2017) find positive effects of week-long vacation academies in math taken by 1,800 pupils.

The few studies that rely on randomization mostly have small sample sizes.[5] In a study with 573 observations, Somers et al. (2015) evaluate a summer camp for middle school pupils called the Building Educated Leaders for Life program. They find some positive effects on math performance, but little impact on reading. In a study with 263 observations, Lynch and Kim (2017) study a math summer camp for 3rd-9th graders from low-income households in the US. Only the condition where pupils participate in the camp and receive a laptop has an effect on academic performance. In a study with 435 observations, Gorard et al. (2015) study a summer camp for grade 5 and 6 pupils on English and math. They find a short-run effect in English, but no effect in math.

Schueler (2020) evaluates an RCT with 1,187 struggling sixth and seventh graders nominated for one week vacation academies in math offered as part of turnaround reforms in nine low-performing Massachusetts middle schools. She finds no significant improvement in standardized math test scores, but evidence of lower exposure

[5] In the following, we only review those studies that have at least 100 observations (which is still on the low side in terms of ensuring sufficient power).

to disciplinary actions. Compliance however was an issue with only 44 percent of the pupils assigned to treatment attending the academy and 18 percent of control pupils ending up getting treated.

Overall, there is weak evidence on the effect of intensive summer camps on academic outcomes in the transition from lower to upper secondary education. Evidence on school-based intensive learning camps and on non-academic outcomes is lacking in the literature.

School-based interventions. Dietrichson et al. (2020b) and Dietrichson et al. (2020a) provide systematic reviews and meta-analyses of the large literature on school-based interventions for low achieving pupils in grades k-6 and 7-12, respectively. Taken together, these two studies cover close to 300 interventions, however, none of which could be classified as an intensive learning camp.

Most closely related to our study are programs that increase instruction time in certain subjects, or that analyze the effects of teaching smaller groups. Regarding the latter, Dietrichson et al. (2020b) and Dietrichson et al. (2020a) find that peer-assisted instruction and small-group instruction (1-5 pupils per teacher) has the largest positive effects in comparison to other school-based interventions. Medium-group instruction (6-20 pupils per teacher) also shows significant positive effects, but this category only exists for the review of interventions aimed at grades k-6. Our intervention contributes here with evidence on the possible effects of such medium-group instructions for older pupils.

Regarding the effects of an increase in instruction time, Lavy and Schlosser (2005) exploit the gradual phasing in of schools to identify the causal effect of a program targeting underperforming pupils in Israel in grades 10-12. They find that participating pupils were more likely to receive matriculation certificates. Yet, the program was found to be less cost-effective than alternative interventions. Cortes et al. (2015) use an RDD to study the effects of doubling the instruction time in math for low-skilled 9th graders and find positive effects both in the short and long run.

Non-cognitive skills. Given the importance of non-cognitive skills for academic outcomes, a range of educational programs try to target these skills. Durlak et al. (2010) provide a meta-analysis of after-school programs that have the aim to enhance non-cognitive skills. They report positive effects not only on the non-cognitive skills targeted, but also on academic outcomes. Kautz et al. (2014) summarize the literature on interventions targeting cognitive and non-cognitive skills in children and adolescents. They emphasize the importance of not only considering cognitive skills (IQ and test scores) when evaluating interventions, but also non-cognitive skills. Further, they point out how interventions for younger children typically have a larger impact than interventions targeted at adolescents or young adults. While the former often impacts non-cognitive skills, the latter often treat problem behavior. They note

that the most successful interventions for adolescents are those that target (also) non-cognitive skills.

Unlike intensive learning camps, programs targeting non-cognitive skills typically run over extended periods of time and use or combine mentoring or training for parents and teachers (see, e.g., McCord, 1978; Tierney et al., 1995; Kemple and Willner, 2008; Durlak et al., 2010; Rodriguez-Planas, 2012; Holmlund and Silva, 2014; Martins, 2017; Kosse et al., 2020, for programs targeting a similar age group as our study). Most studies, with the exception of McCord (1978) and Rodriguez-Planas (2012), find positive effects on outcomes. Our study is most closely related to the studies by Alan and Ertac (2018) and Alan et al. (2019) discussed above, as these studies are also school-based and run over a comparable time frame.

1.2 Study design

1.2.1 Background

The camps are funded by a special grant by the Ministry of Children and Education to investigate intensive learning camps as tools for improving readiness for post-compulsory school education in grade 8. All pupils in Danish public schools undergo an Education Readiness Assessment (ERA henceforth) during the Fall term in grades 8 and 9, and from 2018/2019 also during the Spring term in grade 9. The evaluation is done by the pupil's school following detailed guidelines from the Ministry of Children and Education[6] and encompasses academic outcomes (grades) as well as personal and social skills. The purpose with the ERA is to identify pupils who are not ready for their desired upper secondary education and initiate discussion on suitable educational choice and implement interventions to make them ready. In a given year, between a quarter and a third of the pupils are assessed to lack the academic, social or personal skills required for a post-compulsory school education (Undervisningsministeriet, 2017a,b). The intensive learning camps in this study are targeted at such "non-ready"-pupils (NR-pupils henceforth).

1.2.2 Structure of the camp

We evaluate three rounds of intensive learning camps that took place in 2017-2019. There are two variants of the camp: *Camp+* and *Camp*, which vary in terms of whether part of the time is devoted to training non-cognitive skills (*Camp+*) or to additional training in Danish and math (*Camp*). The other key components of the camps are lower teacher-student ratio, teaching materials developed to low-performing pupils, school-based main camp, and 8-weeks follow-up program with one of the camp

[6] Guidelines (in Danish): https://www.uvm.dk/vejledning-og-stoettemuligheder/vejledning/ungdomsuddannelse/uddannelsesparathed/om-vurderingen .

teachers. Each camp is taught by the school's own teachers and pedagogues as a remedial education program.[7] The camps run during the regular school hours and replace the normal instruction that participating pupils otherwise would receive.

In each school, approximately 10-14 pupils participate in the camp. Two teachers (or a teacher and a pedagogue) teach these pupils. In case more than 15 pupils participate (a camp may take up to 20 pupils), 3 teachers are present. Prior to the camp, the teachers participate in a two-day work shop during which they learn about the teaching materials, key ideas behind the camp, and the rules and procedures of the camp. Materials include detailed teacher guides for both the camp and the 8-week follow-up.[8]

A camp lasts 2 weeks with 35 teaching hours per week (including around one hour of breaks each day). Each day is structured in the same way except for the first and last day of the camp. The consistent structure is supposed to help pupils create learning habits and routines, which in turn facilitate self-control (cf., e.g., the arguments in Galla and Duckworth, 2015). The first day has a longer introduction to the topics in math and Danish and includes brief tests in these subjects; the last day includes summaries for the different subjects and the camp in general. The daily structure is clearly communicated to participants, for example, by posting the timetable of the day on the wall.[9]

Each day starts with welcoming pupils and presenting the program of the day. Each day, pupils have two 2-hour blocks of math and Danish, one subject in the morning and one in the afternoon. In *Camp+* they have two 45-minute blocks covering non-cognitive skills around lunch (one block before lunch and one after). In variant *Camp*, the blocks on non-cognitive skills are substituted by additional blocks of math and Danish, so that pupils in *Camp* have 2.75 hours of math and 2.75 hours of Danish each day. The additional time in *Camp* is mainly used for practicing exercises. Some time also is devoted to explaining different strategies for solving specific problems in math and Danish.

After the two weeks of the learning camp, pupils in *Camp+* and *Camp* continue for 8 weeks in a follow-up program that builds on the material from the camp. Each week during this 8-week period, a teacher (typically one of the camp teachers) meets with 5-7 pupils - during regular school hours - for 1.5 hours to repeat, practice, and deepen the understanding of the material from the camp and help the pupils to apply the material in their regular school work. In treatment *Camp+*, pupils hence face material covering math, Danish, and non-cognitive skills; while in treatment *Camp*, they only cover math and Danish.

[7] The regular classroom would in the two weeks the teacher is conducting the camp have a substitute teacher.

[8] The materials (in Danish) can be accessed here: https://emu.dk/grundskole/overgange/uddannelsesparathed?b=t5-t28.

[9] We surveyed the schools after the camp and 91% of the schools state that the camp was conducted after the plan to a high degree and 8% in some degree.

1.2.3 Teaching materials

1.2.3.1 Math and Danish

Both camps train pupils in the most important areas in the math and Danish curricula to enable them to catch-up in both subjects and, in the longer run, pass the final exams in grade 9 and commence on an educational trajectory. Due to its intensive nature, the camp does not cover all topics of the curriculum in the respective subject. In math, geometry is not covered and statistics and probability is only covered cursorily. In Danish, decoding, a focus area of the national tests (see below), is not practiced.[10]

Specifically, in math, pupils work with decimal numbers, fractions, percentages, mathematical formulas, as well as general problem solving competencies. They revisit the concepts and perform calculations by hand and with a calculator. Some of the exercises are formulated in a similar way as the ones posed in the compulsory school leaving exam in grade 9.

In Danish, pupils work with topics that are tested in the school leaving exams: language comprehension, spelling (exercises targeted at expanding the vocabulary with the aim to foster a better understanding of texts and writing correctly), and text comprehension (including writing of texts and interpretation and discussion of literary texts). Pupils are confronted with different text genres. Next to literary texts, they read, for example, newspaper articles from different sections (such as news, opinion, science).

Pupils take short math and Danish tests at the first day of camp and at the end of camp. These tests were designed together with the teaching materials. The test results are available to the camp teachers as well as the pupils Danish and math teachers. They give the teachers an indication of the academic strengths and weaknesses of the pupil and inform conversations with the pupil on individual focus areas for the camp.[11]

1.2.3.2 Non-cognitive skills (variant *Camp+*)

Variant *Camp+* includes teaching modules on non-cognitive skills. The teaching materials contain various exercises and six short videos. Like Alan et al. (2019), we aim not only to teach the pupils certain concepts, but also to teach the teachers how to apply them. Accordingly, the teacher manual and the 2-day teacher training course emphasize, for example, how to create a growth mindset, how to build good working habits by practicing routines and providing structure, how to increase the self-control and attention of pupils, and how to give feedback.

[10] The decoding part of the national test, for example, asks pupils to separate three words, that were collapsed (like tabooallegiancetyphoon) into individual words.

[11] Since these tests were not administered to the control groups, and since their purpose explicitly was to inform the teachers about their pupils' progression during the camp, they are not used in the evaluations. Moreover, they also did not cover the 8-week follow-up period.

The teaching material on non-cognitive skills includes several components that other studies successfully implemented before. Self-regulation strategies are the first main focus area of the teaching material. Pupils learn about goal setting (Locke and Latham, 1990) and set goals for math and Danish.[12] As goal attainment is enhanced if used in combination with implementation intentions (Gollwitzer and Sheeran, 2006) and mental contrasting (Duckworth et al., 2011; Gollwitzer et al., 2011; Duckworth et al., 2013), the teaching material covers these methods and pupils apply them in concrete situations (such as doing homework). Under mental contrasting, pupils imagine the positive aspects of reaching a goal and then reflect on obstacles that could prevent them from reaching the goal. Implementation intentions are if-then plans, such as “if I try to do my homework, but am tempted to glimpse at my smartphone, I give the smartphone to my mother". Along these lines, pupils also learn about the self-control strategy of situation selection and modification (Duckworth et al., 2016b,a).

Mindsets (Dweck, 2006) are the second main focus of the teaching material. The aim of these modules is to influence the mindset of the pupils in different learning situations, and to help them develop a growth mindset. People with a growth mindset believe that ability is not fixed and initially given, but that effort can enhance ability. Such a view can help pupils not to attribute failure to a lack of ability and give up, but instead to persevere. Thus, pupils learn how to handle academic challenges and stay motivated to provide effort.

Identifying and evaluating own strengths and difficulties is the third main focus of the teaching material. Based on the VIA (Value in Action) Classification of Character Strengths (Seligman et al., 2004), pupils learn about the main categories, which are wisdom and knowledge (e.g., creativity and curiosity), courage (e.g., perseverance and honesty), humanity (e.g., kindness), justice (e.g., teamwork and fairness), temperance (e.g., forgiveness and prudence), and transcendence (e.g., hope). Pupils get inspiration and time to work on how to build their strengths, and they learn how to give feedback to other pupils about strengths in a constructive manner.

In addition, during the welcome session in the morning, pupils get an introduction to the “personal or social skill of the day". The teacher encourages the pupils to reflect upon how they want to work on enhancing this skill. During the day, smaller exercises address the skill. The list of the skills of the day correspond to the skills that are assessed in the ERA (motivation, independence, taking responsibilities, tolerance, reliability, respect, preparedness, working together).

Finally, the material includes smaller exercises on various topics, such as learning zones and healthy habits (sleeping enough and healthy eating). To provide small, productive breaks, pupils also work on some fun concentration tasks (for example, painting a mandala) and fun visual illusion tasks.

[12] The popular writing (for the education context, see, e.g. Conzemius and O'Neill, 2009) has translated the scientific insights on goal setting into the concept of SMART (Specific, Measurable, Achievable, Realistic, and Timely) goals that we also used in the teaching materials.

During the last four sessions, the pupils reflect upon the advice that they would give to another pupil based on what they have learned during the camp and they record this in a video.

1.2.4 Randomization procedure

Figure 1.1 provides an overview of the randomization procedure. The randomization was done at the school level and all public schools were eligible for participation in the trials.[13] In rounds 1 and 3, we randomized the schools that signed-up for participation into the treatment variant *Camp+* and a control group. In round 2, we randomized between the two treatment variants *Camp+* and *Camp*.[14] Schools were allowed to collaborate with each other in setting up a camp in order to have a sufficient number of participating pupils. Such schools were treated as one unit in the randomization.

We stratified schools to ensure that similar schools did not all end-up in the same group. First, schools were divided into strata based on how many camp-classes a school had registered for. Second, we stratified the schools on the basis of their share of NR-pupils in grade 8 in the previous year. After we ranked the schools, we divided them into clusters of six schools. From each cluster, half of the schools were randomly assigned to treatment *Camp+* and half to the control group (round 1 and 3) or *Camp* (round 2).[15] The six school clusters is included in the econometric models as dummies variables because the randomization is performed within those.

As the schools needed to plan in advance – e.g., reserving teachers' time for the camp – it was necessary to communicate the results of the randomization to the schools already in the fall, while the camp ran in the following spring term. The study involves a total of 15,559 pupils at 264 schools. In December 6,094 were assessed as NR-pupils at the ERA and in January 3,600 of the NR-pupils where selected to participate in the camp. In particular, schools in rounds 1 and 3 knew whether they were assigned to the treatment or control group when conducting the ERA and subsequently selecting pupils for the camp.[16] This is less problematic in round 2,

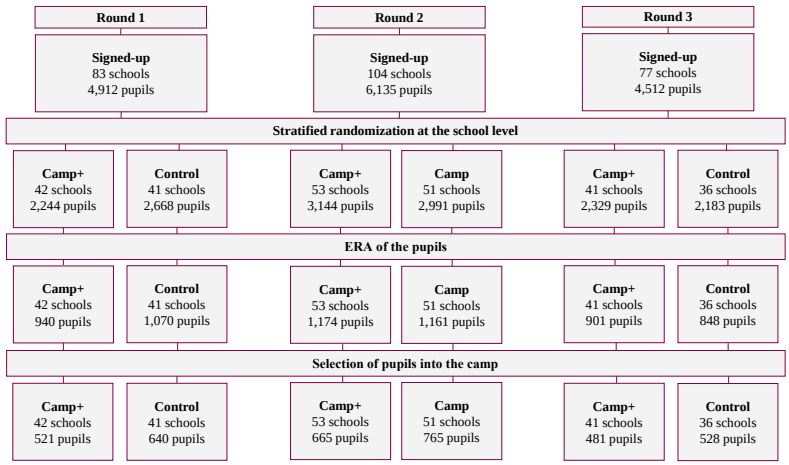
[13] Special needs schools were not eligible for participation because the intervention was design for regular pupils.

[14] This sequencing and design was a consequence of several factors. First, the design of *Camp* was not ready for round 1, and second, a power issue prevented us from having two treatment arms and a control arm in round 2. Moreover, since there could be differences in selection at sign-up, it would be difficult to compare two variants tested in different rounds against each other. After round 2 found no remarkable differences between the two camps, we decided to test *Camp+* again in round 3 to obtain as much power as possible.

[15] In some clusters, there were fewer than six schools, due to the randomization procedure. Thus, there is not necessarily an equal number of schools in the two groups.

[16] Both treatment and control schools designated which pupils would participate in the camp/function as the corresponding control group. In addition, participating pupils and their parents were asked for consent.

Figure 1.1: Overview of the randomization procedure



because all schools are treated (either with *Camp* or *Camp+*). We will return to this issue in section 1.3 when discussing balance tests and in section 1.4.2 when outlining the analysis groups.

1.2.5 Power

Power calculations indicated that 44 schools were needed in the treatment and 44 schools in the control group to detect a minimum effect size of 0.25 of a standard deviation for the main outcome variables. Pooling rounds 1 and 3 in the analysis yielded 83 treatment schools and 77 control schools. In round 2, we compared two different variants, *Camp* and *Camp+*, which share around 70 percent of the teaching materials. Ex-ante, the effect size difference was expected to be below 0.25. Thus, even though a total of 104 schools participated in round 2, the comparison of the two variants is under-powered, and the results should be interpreted with caution.

1.2.6 Division of responsibilities

The Ministry of Children and Education funded the intervention. A Steering group governed the entire project. The latter consisted of representatives from the ministry, the team developing the course materials (University College Copenhagen and VIA University College – two major Danish teacher training colleges), the implementation team (a private consultancy firm, Rambøll Management Consulting), and the research team (the authors of this paper).

The research team directed the design of the intervention. It proposed the intervention and evaluation design. The Steering group approved both without conflict. Randomization was conducted by one of the researchers using STATA's built-in runiform() random number generator. The development team developed the teaching materials for the camp and the follow-up program in consultation with the research team. The research team also provided the input for the material covering non-cognitive skills.[17] The implementation team governed the implementation of the intervention, except for the two-day courses for the teachers, which the development team held. Thus, the research team was only partially involved in the development of the intervention and not at all involved in its implementation. Nevertheless, the PI was responsible for the entire project vis-à-vis the Ministry of Children and Education.

1.3 Data

The analysis is based on registry data from Statistics Denmark, the Danish Agency for IT and Learning, as well as surveys that we carried out among all grade 8 pupils of participating schools. Table 1.1 provides an overview of the outcome variables, which we also describe in the following. The outcome variables are standardized[18], with the exception of post-compulsory education choices and the grade 9 ERA (Education Readiness Assessment, see Section 1.2.1).

Short-run academic performance. To measure the short-run academic effects of the camp, we rely on the national tests in math and Danish, which take place in the spring of grade 8 (for a description of the national tests, see Beuchert et al., 2018). As the name suggests, these are nationally administered, standardized and computerized tests. The tests are on average administered 36 days after the camp and hence measure the short-run effects of the camp. In math, the grade 8 national tests were only introduced from 2018 on, i.e., we cannot use this outcome measure for round 1 (which took place in the spring of 2017). With the follow-up program being 56 days the short-run academic test are measured midway though the follow-up period. There is a risk that this leads to conservative estimated effects if one expects the follow-up period to be important for academic performance.

In Danish, the tests measure the academic skills of the pupil within the following three areas: language comprehension, decoding and text comprehension. As mentioned above, decoding is not practiced during the camp, while language and text comprehension are. Thus, we expect effects of the camp on the targeted areas language and text comprehension (primary outcome variables). We also test whether

[17] We thank Kamilla Trille Gumede for her competent assistance in developing this material.

[18] Standardized to mean zero and standard derivation of one using the full population of pupils in Danish public schooling. Only exception is the self-collected survey data, which is standardized using all pupils included in this study.

Table 1.1: Outcome variables

Short-run academic performance		
National test in Danish in grade 8†	Language comprehension	(primary)
	Text comprehension	(primary)
	Decoding	(secondary)
National test in math in grade 8†,a	Numbers and algebra	(primary)
	Geometry	(secondary)
	Statistics and probability	(secondary)
Long-run academic performance		
Final exam grades in grade 9†,b	Danish	(primary)
	Math	(primary)
Post-compulsory education[c]		(primary)
Non-cognitive skills		
Social and personal skills evaluation in grade 9 (ERA)[d]		
Psychological scales administered in the post-survey†		

Notes. † *Standardized variable.* [a] *Only from 2018, i.e. not available for round 1.*
[b] *Not available for round 3, because final exams were not held due to the Covid-19 lockdown.*
[c] *Dummy =1 if no further education 2.5 years after the camp and =0 otherwise. Not yet available for round 3.*
[d] *Evaluated ready/not ready for a particular post-compulsory education.*

there is a spillover effect on decoding (secondary outcome variable). Such a spillover effect can occur, for example, because the pupil, as a result of the camp, gains better self-regulation skills or a growth mindset, because she/he is more motivated, or because the overall Danish skills that she/he acquired during the camp also help with decoding.

In math, the profile areas are numbers and algebra, geometry, and statistics and probability. The main focus of the camp is on numbers and algebra. Thus, we expect a positive treatment effect in this area (primary outcome variable). In addition, as for decoding in Danish, we test whether there are spillover effects on the non-targeted areas geometry and statistics and probability (secondary outcome variables).

Long-run academic performance. To measure the long-run effects of the camp, we use the final exam grades in math and Danish in grade 9. These grades are given slightly more than a year after the camp. The tests are either computerized or an external censor takes part in these tests. For round 3, final exams were not carried out because of the Covid-19 lockdown, and hence these measures are not available.

In Danish, the grades measure the competencies of the pupil in reading, spelling, writing, and speaking. In math, they measure how well the pupil can solve problems without aids and how well she/he can apply formulas with aids (such as computer programs, or calculators). The exams in Danish have both a written and an oral part. In math, all pupils take written tests, and a random draw determines whether

a pupil is orally examined in math or instead in another science subject. For math and Danish, we use the respective overall grade as the main outcome variable, but we also report separately results for the written and oral parts in the appendix.

Moreover, we observe whether pupils enroll in any post-compulsory education (e.g., vocational training, high school, taking the voluntary 10th grade) or take no further education. For round 3, this measure is not yet available. For round 1, we use the status 2.5 years after the camp as an outcome variable.

Non-cognitive skills. We use register data and self-administered surveys to evaluate the effects of the camp on non-cognitive skills. Each of the outcome variables described in the following has certain disadvantages. For this reason, we do not classify outcomes variables as primary or secondary outcome variables. Instead, rather than focusing on the statistical significance of any single measure, we will only conclude that the camp has an effect on non-cognitive skills if different measures point in the same direction.

From the register data, we use the evaluation of the social and personal skills of the pupil in the ERA in grade 9. A potential disadvantage with this measure is that the teachers who conduct the ERA know whether a pupil participated in the camp or not which could potentially influence the evaluation.

From self-administered pre- and post-surveys, we draw on several validated psychological scales (for a detailed description of the included items, see section A.1.1 in the appendix): the 8-item Grit Scale (Duckworth and Quinn, 2009), the Domain-Specific Impulsivity Scale for Children (Tsukayama et al., 2013), the Core Self-Evaluations Scale (Judge et al., 2003), and the Strengths and Difficulties Questionnaire for adolescents (Goodman et al., 1998). We measure beliefs about the malleability of abilities (mindset) with 4 items based on Dweck (2006). Further, we include the 1-item risk aversion question (Dohmen et al., 2011), the 1-item patience question (Vischer et al., 2013) and two questions on time-preferences from the GSOEP survey.[19] A disadvantage of the survey-based measures is the fact that they are designed by the research team in accordance with the teaching materials. Thus, pupils might give the "desired" answer. Further, the camp might make pupils (and teachers) more aware of certain skills rather than changing them – a difference the survey measures cannot capture.

The schools know when they are taking the pre-survey whether they are treatment or control schools. This might explain that we observe 12% more attrition from the control schools in the pre-survey. This number increases to 19% in the post-survey, which potentially affect our estimates.

[19] Rounds 2 and 3 included four additional, domain-specific risk questions about taking desirable risks (raising your hand or volunteering to present in class) and risky behavior (drinking, smoking, and illegal actions).

Control variables and balance tests. The register data allow us to include a range of school-related control variables, such as previous test scores, previous school absence, and previous responses in the national school well-being survey, as well as background characteristics of the pupils such as gender, ethnicity, family situation and variables related to the socioeconomic status of the parents. The balance test tables provide the complete list of control variables (cf. tables A.1-A.3 in the appendix). Section 1.4.1 describes how we use LASSO to exogenously select the covariates for each regression.

The balance tables demonstrate that in all three rounds there is balance between treatment and control over a range of observables when we look at all pupils in grade 8, as expected from the randomization process and selection of pupils into the camp (cf. section 1.2.4). When testing 130 covariates in a balances table, we would expect to find 7 significant variables at a 5% level simply due to chance. In table A.1 11 covariates are significant, which is close to our expectation. However, the important pre-math tests are significant, which would effect our treatment estimates and must be addressed in the econometric models. It appears that schools in *Camp+* more carefully classified pupils in the ERA than schools in the control group: In rounds 1 and 3, we observe from balance tests that the NR-pupils, as well as the pupils selected for the camp are not balanced between *Camp+* and control. Specifically, treatment schools select fewer, more poorly academic performing pupils with better personal competencies for the camp and to be NR in the ERA. In particular, we observe that *Camp+* NR-pupils are on average approximately 0.1 standard derivation below the control NR-pupils in reading and math pre-tests and less likely to be assessed academic ready at the grade 8 ERA. For pupils selected for camps the academic differences increases by additional 0.05 standard derivations. In general, the academic differences seems to be driven by pupils being dyslexics and not parental characteristics. Thus, we need to address this selection issue in the econometric models that we use. We turn to this issue in the next section.

1.4 Analysis

1.4.1 Estimation strategy

To test the null hypothesis that the camp has no effect on an outcome variable, we estimate the following equation:

$$y_i = \alpha + \delta D_i + \gamma S_i + \beta X_i + \epsilon_i, \tag{1.1}$$

where subscript i refers to each pupil. Thus, y_i is the outcome variable and ϵ_i is the error term. The coefficient of interest δ measures the effect of the treatment captured by dummy D_i, as explained below. S_i is a indicator variable for each pupil's school randomization-stratum, and X_i is a vector of covariates. Standard errors are clustered

at the school level. We apply the Double-Lasso (DL) variable selection method suggested by Belloni et al. (2014) to select covariates. We include covariates to account for the imbalance showed previously and to improve precision of the estimated intervention effects. The DL method is calibrated to not over-select spurious covariates, and Urminsky et al. (2016) concludes that this method is particularly useful for imperfect randomized experiments (and for under-powered analyses in a perfect randomized experiment, by increasing statistical power). It comprises three steps:

1. Predict the dependent variable using a LASSO regression with the complete list of control variables and save the variables with non-zero coefficients.

2. Predict the treatment indicator using a LASSO regression with the complete list of control variables and save the variables with non-zero coefficients. If the treatment is perfectly randomized, then no covariates should be selected in this step.

3. Combine the covariates selected in the first two steps and include them in equation (1).

In the LASSO regressions, we apply 10-fold cross-validation to estimate the optimal shrinkage parameter.

1.4.2 Analysis groups

To address the aforementioned imbalances, we report results for different sets of pupils as explained in the following.

***Camp+* vs. Control.** If we look at all grade 8 pupils, by design randomization should be perfect, which is confirmed by the aforementioned balance tests for rounds 1 and 3 where schools were randomized into *Camp+* or control. In our main specification, we therefore include all pupils who were in a treatment or control school in grade 8 (henceforth ITT group for Intention To Treat). We set D_i equal to one if in grade 8 a pupil was in a school that was selected for treatment and zero if she/he was in a control school. Thus, the estimate $\hat{\delta}$ corresponds to the ITT effect. That is, it measures the effect of the school being assigned to the treatment, but not necessarily of being treated.

The intention to treat estimate of the main specification likely constitutes an extreme lower bound on the true effect (see also section 1.5.5 for further caveats). We therefore also report two additional estimates relying on different groups of pupils. While these estimates help gauge the true effect size, they have to be interpreted with caution because schools carried out the ERA and selected pupils for the camp after learning their treatment status, as explained in section 1.2.4. This leads to some imbalances, as documented in table A.2.

The first additional estimate is the intention to treat estimate based only on the sample of pupils evaluated NR in grade 8, to which we refer as ITT-NR (Intention To Treat on the NR-pupils) in the following. The ITT-NR group is a subset of the ITT group.

With the second additional estimate, we estimate the effect of the camp on the pupils who actually participated in the camp with the local average treatment effect (LATE, Imbens and Angrist, 1994) for the ITT-NR group. An underlying assumption of the LATE estimate is that there are no spillover effects on non-treated pupils. Thus, if the schools use the teaching material beyond the camp, we risk having the LATE effect overestimating the effect of the camp. Thus, the LATE estimate likely constitutes an upper bound for the true effect.

***Camp+* vs. *Camp*.** In round 2, both groups receive an intervention and by design randomization should be perfect. Indeed, we observe balance between the two groups (cf. table A.3 in the appendix). We therefore restrict the sample to only those pupils who actually were selected for the camp in grade 8. The estimate $\hat{\delta}$ thus corresponds to the average treatment effect on the treated (ATT) of *Camp+* versus *Camp*. The ATT group is a subset of the ITT-NR group.

1.4.3 Short-run effects on academic outcomes

We first test for the short-run effects of *Camp+* vs. control using the national tests in math and Danish conducted in grade 8. We report the results in tables 1.2 and 1.3. Our preferred specifications include covariates, since we observe some imbalances despite the randomization, especially for the ITT-NR group (cf. section 1.4.2). But we also report specifications without covariates.

National test in Danish. For all estimates (ITT, ITT-NR, and LATE) we find positive and significant effects of *Camp+* on one of the primary outcomes, *text comprehension*. The effect sizes range from 0.07-0.18, with the LATE estimate being more than twice as large as the ITT estimate. Yet, contrary to our expectations we do not find a positive effect on the other primary outcome *language comprehension* as our main estimate, the ITT estimate, shows an insignificant negative effect. Given these inconclusive results and given that we perform multiple tests (cf. section 1.5), some caution should be taken in interpreting the positive effect on *text comprehension*.

Further, we do not find an effect on the secondary outcome variable *decoding*. If we aggregate the three outcomes (for comparison with longer term outcomes from grade 9 exams), there is no significant effect on the overall performance.

National test in math. This test was introduced in 2018 and is therefore not available for round 1. With our sample restricted to round 3, for all estimates, we find

Table 1.2: Short-run effects on the national grade 8 test in Danish, *Camp+* vs. control

	Language Comprehension		Decoding		Text Comprehension		Overall	
ITT	-0.075	-0.033	-0.016	0.017	0.047	0.065***	-0.018	0.017
	(0.052)	(0.043)	(0.037)	(0.024)	(0.035)	(0.021)	(0.041)	(0.028)
Mean outcome, Control	.024	.024	.000	.000	-.057	-.057	-.013	-.013
R-squared	.015	.186	.009	.510	.011	.520	.013	.549
Observations	8,953	8,953	8,953	8,953	8,953	8,953	8,953	8,953
ITT-NR	-0.113**	-0.046	-0.059	0.014	0.055	0.090***	-0.048	0.024
	(0.052)	(0.045)	(0.038)	(0.028)	(0.035)	(0.029)	(0.042)	(0.033)
LATE	-0.231**	-0.094	-0.121	0.028	0.113	0.184***	-0.098	0.050
	(0.107)	(0.092)	(0.078)	(0.058)	(0.073)	(0.059)	(0.087)	(0.067)
Mean outcome, Control	-.233	-.233	-.459	-.459	-.574	-.574	-.519	-.519
R-squared	.024	.213	.009	.431	.013	.342	.014	.442
Observations	3,418	3,418	3,418	3,418	3,418	3,418	3,418	3,418
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

evidence for positive effects of *Camp+* on the primary outcome *numbers and algebra*. Effect sizes range from 0.07-0.20, with the LATE estimate being almost three times as large as the ITT estimate.

The main estimate (ITT) does not show significant effects of the treatment on the secondary outcome variables, i.e., areas in math that were not targeted in the camp. Yet, for the ITT-NR and LATE estimates, we find some suggestive evidence for positive effects on *geometry* (effect sizes 0.09-0.19) and *statistics and probability* (effect sizes 0.12-0.24). If we aggregate the three areas, there is a positive effect on the overall performance (effect sizes 0.06-0.20). Comparing these effect sizes to the mean outcome in the control group, it is evident that the intervention closes between 10-30% of the gap in math ability that these pupils have accumulated during their time in school.

***Camp+* vs. *Camp*.** We do not find any significant differences between the two variants of the camp. We report the results of *Camp+* vs. *Camp* on the short-run academic outcomes in tables A.4 and A.5. We will return to these results when we discuss the impact of *Camp+* on non-cognitive skills in section 1.4.5.

Table 1.3: Short-run effects on the national grade 8 test in math, *Camp+* vs. control

	Numbers and Algebra		Geometry		Statistics and Probability		Overall	
ITT	0.025	0.068**	0.000	0.043	0.023	0.063	0.017	0.059*
	(0.060)	(0.032)	(0.063)	(0.035)	(0.065)	(0.039)	(0.065)	(0.034)
Mean outcome, Control	-.064	-.064	-.042	-.042	-.033	-.033	-.050	-.050
R-squared	.018	.640	.020	.607	.019	.601	.021	.706
Observations	4,283	4,283	4,283	4,283	4,283	4,283	4,283	4,283
ITT-NR	0.004	0.094**	-0.001	0.091*	0.017	0.117*	0.007	0.100**
	(0.051)	(0.047)	(0.051)	(0.046)	(0.069)	(0.062)	(0.057)	(0.050)
LATE	0.008	0.192**	-0.002	0.185*	0.035	0.240*	0.015	0.204**
	(0.104)	(0.096)	(0.103)	(0.093)	(0.141)	(0.128)	(0.116)	(0.102)
Mean outcome, Control	-.645	-.645	-.625	-.625	-.619	-.619	-.678	-.678
R-squared	.010	.452	.011	.452	.012	.460	.011	.550
Observations	1,581	1,581	1,581	1,581	1,581	1,581	1,581	1,581
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes	No	Yes

*Notes:***p<0.01; **p<0.05; *p<0.1. Outcome for math only exists for round 3. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

1.4.4 Long-run effects on academic outcomes

To measure the long-run effects of *Camp+* vs. control, we consider the final exam grades in grade 9 and whether pupils enrolled in any post-compulsory education 2.5 years after the camp, reported in table 1.4. Note that these outcomes are not available for round 3 yet (cf. table 1.1).

Final exams in Danish and math. We do not find a significant effect of *Camp+* on the final exams in Danish. We find suggestive evidence that the short-run effect of *Camp+* in math translates into a positive long-run effect on the performance in the math final tests in grade 9. Effect sizes for the ITT-NR and LATE range from 0.06-0.18, but neither are they significant at the 5% level, nor does the ITT estimate show a significant effect. Additional analyses suggest that the effect may stem from the better performance in the written part of the math exams (cf. table A.6 in the appendix).[20]

Post-compulsory education. There is some suggestive evidence that *Camp+* has a positive effect on pupils being enrolled in education 2.5 years after the camp. While our main estimate, the ITT, is not significant and also very close to zero, according to

[20] All pupils take the written tests. Oral examination occurs only in math or another science subject, depending on a random draw.

Table 1.4: Long-run effects of *Camp+* vs. control on grade 9 exams and on the fraction not enrolled in education 2.5 years after the camp

	Danish score		Math score		No education	
ITT	-0.055	-0.020	0.023	0.062	0.002	-0.004
	(0.053)	(0.034)	(0.061)	(0.048)	(0.012)	(0.009)
Mean outcome, Control	-.055	-.055	-.101	-.101	.095	.095
R-squared	.014	.684	.016	.660	.005	.152
Observations	4,761	4,761	4,744	4,744	4,820	4,820
ITT-NR	-0.042	-0.005	0.026	0.088*	-0.027	-0.037**
	(0.052)	(0.041)	(0.058)	(0.050)	(0.021)	(0.018)
LATE	-0.086	-0.011	0.053	0.180*	-0.055	-0.076**
	(0.106)	(0.085)	(0.118)	(0.103)	(0.043)	(0.038)
Mean outcome, Control	-.754	-.754	-.779	-.779	.193	.193
R-squared	.017	.520	.013	.477	.012	.163
Observations	1,880	1,880	1,864	1,864	1,931	1,931
Strata	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Outcome only exists for round 1. Educational enrollment is measured September* 31^{th} *three years after the camp. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

the ITT-NR estimate, NR-pupils in a school that offered the camp are more likely to be enrolled in some post-compulsory education than NR-pupils in the control group. The ITT-NR and LATE estimates show that uptake of post-compulsory education increases by 4-8 percentage points, which should be compared with the baseline of 19 percent of pupils not being enrolled in post-compulsory education in the control group. Additional analyses suggest that the effect may stem from treated pupils being more likely to take-up vocational training (cf. table A.7 in the appendix).

***Camp+* vs. *Camp*.** We report the long-run effects of *Camp+* vs. *Camp* in table A.8 in the appendix. As for the short-run academic outcomes, we find no significant differences between the two camp variants in the long-run exam grades or on the fraction enrolled in post-compulsory education.

1.4.5 Effects on non-cognitive skills

Overall, we find no evidence of an effect of *Camp+* vs. control on non-cognitive skills as measured by the ERA and the psychological scales in the post-survey (see tables

A.9a, A.9b, and A.10 in the appendix).[21]

Further, comparing non-cognitive skills between *Camp+* (which included around 16 hours of teaching on non-cognitive skills and revisiting them during the 8-week follow-up program) and *Camp* (which does not teach about non-cognitive skills), we find an inconsistent picture. Given that we did not classify any outcome measure as primary or secondary, we do not draw a conclusion in one or the other direction from this inconsistent picture, but only summarize the effects in the next paragraph.

There is a tendency that pupils who participated in *Camp+* are more likely to be assessed NR-pupils compared to *Camp*, which shows specifically in a worse evaluation of the personal skills (see tables A.14 and A.15 in the appendix).[22] Further, we do not find any difference in most of the non-cognitive skills targeted in *Camp+* (self-control, grit, and mindsets). Both the negative and the null result are against what we expected ex-ante. Yet, the 1-item question by Vischer et al. (2013) indicates an increase in patience relative to *Camp* (see table A.17a and A.17b in the appendix). *Camp+* also is associated with pupils being more likely to take risks – both positive risks (like saying something in class) and negative risks (like drinking alcohol or smoking).

Given the inconsistent picture, the comparison of academic skills can provide additional indication of whether *Camp+* enhanced non-cognitive skills. As discussed in the previous section, we find no difference in academic skills between *Camp* versus *Camp+*. On the one hand, *Camp* devotes more time to training in math and Danish. But on the other hand, if *Camp+* had a strong positive impact on non-cognitive skills and if these non-cognitive skills impacted academic performance, then pupils who participated in *Camp+* rather than *Camp* should perform better in school – at least in academic areas that were not directly targeted in the camp. Thus, taking all these results together suggests the interpretation that the time spent on non-cognitive skills in *Camp+* does not meaningfully affect non-cognitive skills.

1.4.6 Exploratory analysis: subgroups

While our study is not strictly powered to encompass subgroup analyses, it might nevertheless be enlightening to consider some exploratory results. Of course, given their exploratory nature, they should be interpreted with some caution. We focus here on the effects of *Camp+* vs. control in the national tests. The results are summarized in tables A.18-A.25 in the appendix and include also comparisons of *Camp+* vs. *Camp*.

[21] In the appendix, in tables A.11 and A.12, we further split-up the results of the ERA into the evaluation of the personal and social skills of pupil. No effects are found either. The ERA evaluates whether pupils are assessed to be ready for their preferred post-secondary educations. Table A.13 in the appendix rules out that the treatment affects what educations pupils prefer.

[22] The preferred education of the pupils is not affected by the treatments, as shown in table A.16 in the appendix.

Gender. As many education interventions have a stronger effect on girls than on boys, a natural question is whether the effects of the camp vary by gender. When we look at the national test in Danish, we observe for boys a positive effect on *text comprehension* (effect sizes 0.08-0.17). For girls, we do not see a corresponding ITT effect.[23] In contrast, for the national test in math, we observe for girls a positive effect on *probability and statistics* (effect sizes 0.09-0.35), as well as a positive ITT effect on *numbers and algebra* (effect size 0.09). These results suggest that the camp specifically pushes pupils in areas where – according to gender-stereotypical beliefs – they would expect not to do well.

Socioeconomic background. We use the education level of the mother as an indicator for the socioeconomic background of the pupil. Specifically, we classify pupils as having a low-socioeconomic background if their mothers have 12 years of education or less (i.e., they have at most completed high school). In the national tests in Danish, in the area of *text comprehension* we observe a positive significant treatment impact on pupils with low-socioeconomic background (effect sizes range from 0.14-0.28). Yet, in the national tests in math, we do not observe a positive ITT effect in the targeted area.[24]

Academic low performers. When we look at the worst performing pupils according to the previous national test in grade 6 (lowest 10% of all grade 8 pupils), we observe positive significant effects on *text comprehension* (effect sizes 0.09-0.20). Further, the camp has positive effects in math not only on the targeted area of *numbers and algebra* (effect sizes 0.12-0.25), but also on the non-targeted area *geometry* (effect sizes 0.14-0.28).[25] The effect sizes suggest that the treatment has a larger impact on the worst performing pupils compared to all pupils.

1.5 Discussion

1.5.1 Multiple hypotheses

We have reported a fairly large number of results for each treatment variant, even when limiting ourselves to the primary outcome variables. Thus, some of them might be significant due to statistical chance. We have tried to accommodate this problem by highlighting primarily results where all estimates point in the same direction. But

[23] There is some evidence of an effect from the ITT-NR and LATE estimates (effect sizes 0.10-0.21).

[24] There are some positive results for the ITT-NR and LATE estimates in the non-targeted area.

[25] There is even suggestive evidence for positive effects on the non-targeted area *probability and statistics* (0.15-0.31).

of course, we should further caution against interpreting results too strictly, especially when looking at subgroups and at secondary outcome variables.

1.5.2 Economic significance

There are only few studies relying on randomized experiments or quasi-experimental methods for assessing the impact of intensive learning camps and they provide a mixed picture (see the literature review in section 1.1). In addition, education studies with over 2,000 participants tend to have far smaller effect sizes than studies with fewer participants and RCTs tend to have much smaller effect sizes than quasi-experimental designs (Cheung and Slavin, 2016; Kraft, 2020).

Hence, it is interesting to place our results in a wider context. The short-run effect sizes that we find for overall performance in the math standardized test range from 0.06 to 0.2. Comparing them to the empirical distributions of effect sizes from randomized control trials of education interventions with standardized achievement outcomes in Kraft (2020), they lie roughly in the range of 50th to 80th percentile for interventions targeting math. For example, our results are comparable to the effect on math scores of increasing teacher quality in a term by one standard deviation (here effect sizes range from 0.11 to 0.13; cf. Aaronson et al., 2007; Rivkin et al., 2005).

We find weaker evidence of short-run effects in Danish, as only one out of the two targeted areas in the standardized test (namely, text comprehension) is positively affected. The effect sizes range from 0.07 to 0.18, which is comparable to the effect sizes found in other interventions in Danish primary schools conducted in earlier grades, such as mother tongue teaching (Andersen and Knoth Humlum, 2021) and the 'two-teacher in the classroom' program (Andersen et al., 2020). The former estimates, in the Danish national tests, an effect size of 0.15 for language comprehension and 0.17 for decoding. The latter study reports effects sizes on grades in Danish of 0.06 to 0.10.

When it comes to relevance for practice, effect sizes should be seen in relation to the costs and scalability of an intervention, which are key factors for policy makers (Harris, 2009; Cheung and Slavin, 2016; Kraft, 2020). The main costs of the camps stem from the one-time development of the teaching materials. Running the camp itself costs approximately DKK 5000 ($750) per pupil (mainly to pay for additional work hours of teaching staff). Camps are easy to set up because they take place at the school of the pupils during regular teaching hours, are of short duration (two weeks + 8 weeks of follow-up), and they draw on existing teaching staff who require little additional training due the detailed teaching materials and guides. In terms of cost-effectiveness ratio, our intervention thus can be classified as easy to scale with a medium effect size and moderate cost (Kraft, 2020).

Further, in contrast to, for example, summer camps, the camp is not in addition to the regular teaching, but substitutes for it. That is, while the pupils receive more intensive teaching in math and Danish than usual, they miss out on their regular

classes. Any effect thus indicates that the camp is more effective than regular classes.

Another important consideration is political feasibility of an intervention. Running camps during regular school hours has clear advantages over programs after school or during school vacation. For example, in several countries it was discussed whether learning losses due to Covid-19 lockdowns and restrictions in schools could be addressed by lengthening school days or shortening the school vacation. This met strong opposition from teachers and parents (e.g., Goldstein and Taylor, 2021; YouGov, 2021; SWR, 2021).

1.5.3 Results on social and personal skills

Contrary to our a priori expectations, *Camp+* has no effect on the non-cognitive skills of the pupils. These results are in contrast to the closely related studies by Alan and Ertac (2018) and Alan et al. (2019), who find large and persistent effects of their intervention on non-cognitive and academic skills.

There might be several reasons behind these different findings. One reason could lie in the composition of pupils and/or schools. Our pupils are older than the ones in the studies of Alan and Ertac (2018) and Alan et al. (2019), and it may be more difficult to change the non-cognitive skills of older pupils (cf. Kautz et al., 2014).

Further, our teaching material on non-cognitive skills do not focus on one particular non-cognitive skill (like grit or mindset), but cover a range of non-cognitive skills. The rationale for our approach stems from the evaluations that schools carry out in grade 8 (ERA), where both personal and social skills are evaluated in addition to academic skills. However, we do have some focus topics, such as self-regulation strategies and mindsets. Yet, the total hours on a particular focus topic (3-5 hours) is still less than the total number of hours spent on the specific topics in the intervention studied by Alan and Ertac (2018) and Alan et al. (2019). Their respective interventions focused on one specific topic at a time and used around 2 hours per week for 8-12 weeks. Thus, one reason for us not finding large effects on academic skills and essentially zero effects on non-cognitive skills could be that focusing on many topics is less effective than focusing on one topic.

Lastly, in Denmark, a lot of interventions take place in school (see subsection 1.5.2 and 1.5.5.2 for some examples), and there is generally a culture of implementing new teaching and learning concepts. In contrast, Alan et al. (2019) state that their program took place in under-resourced public schools in Turkey which are mainly attended by pupils from low-socioeconomic backgrounds. Denmark also scores higher in the PISA 2018 tests than Turkey (Schleicher, 2018). Thus, the small effects on academic skills and null results on non-cognitive skills that we find, in comparison to the large effects of Alan and Ertac (2018) and Alan et al. (2019), might be due to the fact that they stem from a setting with lower educational resources and fewer alternatives to the intervention than in our setting.

Thus, overall, caution should be taken in generalizing from our results to con-

clusions such as “non-cognitive skills cannot be changed in older children and thus the main focus should be on training academic skills". Our results are specific to, for example, the used teaching materials or to country specific characteristics. Future research may examine in more detail the role of age when trying to impact non-cognitive skills.

1.5.4 Organization of the camp

Our camp takes place during regular teaching hours in the school of the pupils. This form has some advantages over traditional summer camps. First, a school-based camp may remove potential barriers to participation. No special search effort or application is needed in order to participate in the camp. Further, the camp does not take away leisure time as the pupil has to attend school otherwise. Finally, the pupil knows the teachers, other participants and the location – decreasing potential psychological barriers. Second, a school-based camp permits to follow-up on the outcomes of the learning camp during the regular teaching. Either a teacher of the pupil or a close colleague (i.e., a teacher teaching the same cohort of pupils) is involved in the teaching of the camp. Colleagues within a school can easily exchange information and discuss how to follow-up on the outcomes of the camp.

Yet, as the camp takes places during regular teaching hours, pupils miss-out on the regular teaching. This might be less of a problem for Danish and math, where pupils receive more intensive training than normal. Yet, it might be a problem for other subjects (like natural sciences or foreign languages) that are not targeted in the camp. That is, while the camp, on the one hand, gives pupils something (more math and Danish lessons) it also takes away something from them (lessons in other subjects).

1.5.5 Caveats

1.5.5.1 Non-adherence in the control group

Of the 36 control schools in round 3, 12 had participated in the camp either in round 1 or 2. This creates scope for control schools in round 3 to use the teaching materials from either *Camp+* or *Camp*. In addition, some schools indicated beforehand that they planned to conduct a camp themselves should they not be selected as a treatment school. Thus, in round 3 the management of all schools was asked to sign a legally binding statement declaring whether or not they conducted in grade 8 a camp using our teaching materials. Four control schools indicated that they did use the materials from either *Camp+* or *Camp*. We take this into account in the calculation of the LATE estimate. Yet, for the ITT estimates, these schools will continue to be included in the analyses as control schools, supporting the point that the ITT is a conservative estimate of the true effect of the camp.

1.5.5.2 Treatment as usual

Control schools may provide other educational support for NR-pupils. In particular, if a pupil is assessed NR, the school is required by law to offer additional support. This could, for example, consist of enrolling the pupil in a camp that is offered by the relevant municipality (see the next paragraph), providing individual supervision, or offering activities outside the school. For this reason, we have to assume that all NR-pupils receive some kind of treatment, which probably also affects outcomes. In the statements from school management collected in round 3, we further asked whether the school implemented some type of camp in grade 8 (not using our teaching material). About half of the control schools indicated that they completed some type of camp themselves.

Further, there exist many opportunities for attending intensive learning camps in Denmark. Examples are a summer camp for boys in grade 8 ("Drengeakademiet" supported by Løkkefonden, cf. Andersen and Nissen, 2014; Andersen, 2015) or the camp "Plan T", which targets children in a specific municipality who are dyslexic. In addition, many municipalities offer similar types of intensive learning camps for grade 8 and 9 pupils with academic or social problems (e.g., the 1-week "MOVE" camp in the city of Aarhus).

We treat participation in all such intensive learning camps and initiatives as "treatment as usual". Yet, the presence of such camps and related initiatives implies that we are intervening at a high level of existing support for the NR-pupils. The small effects hence may be due to the fact that the "treatment as usual" has almost as much effect as the camp. Indeed, this might be an explanation for the effect sizes being larger in other countries.

1.6 Conclusion

Using a large randomized trial, we evaluate an intensive learning camp for grade 8 pupils that are assessed 'not ready' for further education. The main variant of the camp does not only train pupils in math and Danish, but also aims to strengthen their non-cognitive skills. In the short-run, we find positive effects on targeted-areas in math and some suggestive evidence for positive effects in Danish. In the long-run, there is some weak indication of the camp having a positive impact on math performance in the final exams in grade 9 and on being enrolled in post-compulsory education 2.5 years after the camp. We do not find any evidence that the camp impacts non-cognitive skills. Finally, when comparing the effects of the two types of camps it is evident that teaching non-cognitive skills is no worse than teaching Danish or math.

Acknowledgements

The design of the teaching material and evaluation of the camp was based on a grant from the Danish Ministry of Education. We are grateful to TrygFonden for funding through TrygFonden's Centre for Child Research. We would like to thank Kamilla Trille Gumede for her intellectual contribution to the project, and University College Copenhagen, VIA University College and Rambøll Management Consulting for their collaboration during the entire project. The study was pre-registered at the AEA RCT Registry under https://www.socialscienceregistry.org/trials/2391.

1.7 References

Aaronson, D., Barrow, L., Sander, W., 2007. Teachers and student achievement in the chicago public high schools. Journal of Labor Economics 25 (1), 95–135.

Alan, S., Boneva, T., Ertac, S., 2019. Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. The Quarterly Journal of Economics 134 (3), 1121–1162.

Alan, S., Ertac, S., 2018. Fostering patience in the classroom: Results from randomized educational intervention. Journal of Political Economy 126 (5), 1865–1911.

Almlund, M., Duckworth, A. L., Heckman, J., Kautz, T., 2011. Personality psychology and economics. In: Handbook of the Economics of Education. Vol. 4. Elsevier, Amsterdam, 1–181.

Andersen, F. Ø., 2015. Drengeakademiet: Drengeakademiets langtidseffekt: trivsel, læring og personlig udvikling i tiden der fulgte. 2013-2015. Dafolo.

Andersen, F. Ø., Nissen, P., 2014. Drengeakademiet: trivsel, læring og personlig udvikling for drenge på kanten 2013-2014. Dafolo.

Andersen, Simon Calmar, G. T., Knoth Humlum, M., 2021. How first-language instruction transfers to majority- language skills. Nature Human Behavior.

Andersen, S. C., Beuchert, L., Nielsen, H. S., Thomsen, M. K., 2020. The effect of teacher's aides in the classroom: Evidence from a randomized trial. Journal of the European Economic Association 18 (1), 469–505.

Battistin, E., Schizzerotto, A., 2019. Threat of grade retention, remedial education and student achievement: Evidence from upper secondary schools in italy. Empirical Economics 56 (2), 651–678.

Belloni, A., Chernozhukov, V., Hansen, C., 2014. High-dimensional methods and inference on structural and treatment effects. Journal of Economic Perspectives 28 (2), 29–50.

Beuchert, L. V., Nandrup, A. B., et al., 2018. The danish national tests at a glance. Nationaløkonomisk Tidsskrift, no. 1.

Cheung, A. C., Slavin, R. E., 2016. How methodological features affect effect sizes in education. Educational Researcher 45 (5), 283–292.

Conzemius, A., O'Neill, J., 2009. The power of SMART goals: Using goals to improve student learning. Solution Tree Press, Bloomington, IN.

Cooper, H., Charlton, K., Valentine, J. C., Muhlenbruck, L., Borman, G. D., 2000. Making the most of summer school: A meta-analytic and narrative review. Monographs of the Society for Research in Child Development, i–127.

Cortes, K. E., Goodman, J. S., Nomi, T., 2015. Intensive math instruction and educational attainment long-run impacts of double-dose algebra. Journal of Human Resources 50 (1), 108–158.

Di Pietro, G., Biagi, F., Costa, P., Karpiński, Z., Mazza, J., 2020. The likely impact of COVID-19 on education: reflections based on the existing literature and recent international datasets. Tech. rep., Publications Office of the European Union, LU.
URL https://data.europa.eu/doi/10.2760/126686

Dietrichson, J., Filges, T., Klokker, R. H., Nielsen, B. C. V., Bøg, M., Jensen, U. H., 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 (2), e1081.

Dietrichson, J., Filges, T., Seerup, J. K., Klokker, R. H., Viinholt, B. A., Bøg, M., Eiberg, M., 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), 1–60.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G. G., 2011. Individual risk attitudes: Measurement, determinants, and behavioral consequences. Journal of the European Economic Association 9 (3), 522–550.
URL http://dx.doi.org/10.1111/j.1542-4774.2011.01015.x

Duckworth, A. L., Gendler, T. S., Gross, J. J., 2016a. Situational strategies for self-control. Perspectives on Psychological Science 11 (1), 35–55.

Duckworth, A. L., Grant, H., Loew, B., Oettingen, G., Gollwitzer, P. M., 2011. Self-regulation strategies improve self-discipline in adolescents: Benefits of mental contrasting and implementation intentions. Educational Psychology 31 (1), 17–26.

Duckworth, A. L., Kirby, T. A., Gollwitzer, A., Oettingen, G., 2013. From fantasy to action: Mental contrasting with implementation intentions (mcii) improves academic performance in children. Social Psychological and Personality Science 4 (6), 745–753.

Duckworth, A. L., Peterson, C., Matthews, M. D., Kelly, D. R., 2007. Grit: perseverance and passion for long-term goals. Journal of Personality and Social Psychology 92 (6), 1087–1101.

Duckworth, A. L., Quinn, P. D., 2009. Development and validation of the short grit scale (grit–s). Journal of Personality Assessment 91 (2), 166–174.

Duckworth, A. L., Quinn, P. D., Tsukayama, E., 2012. What no child left behind leaves behind: The roles of iq and self-control in predicting standardized achievement test scores and report card grades. Journal of Educational Psychology 104 (2), 439–451.

Duckworth, A. L., Seligman, M. E., 2005. Self-discipline outdoes iq in predicting academic performance of adolescents. Psychological Science 16 (12), 939–944.

Duckworth, A. L., White, R. E., Matteucci, A. J., Shearer, A., Gross, J. J., 2016b. A stitch in time: Strategic self-control in high school and college students. Journal of Educational Psychology 108 (3), 329–341.

Durlak, J. A., Weissberg, R. P., Pachan, M., 2010. A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. American Journal of Community Psychology 45 (3-4), 294–309.

Dweck, C. S., 2006. Mindset: The new psychology of success. Random House, New York, NY.

Eskreis-Winkler, L., Duckworth, A. L., Shulman, E. P., Beal, S., 2014. The grit effect: Predicting retention in the military, the workplace, school and marriage. Frontiers in Psychology 5, 36.

Galla, B. M., Duckworth, A. L., 2015. More than resisting temptation: Beneficial habits mediate the relationship between self-control and positive life outcomes. Journal of Personality and Social Psychology 109 (3), 508.

Goldstein, D., Taylor, K., 2021. Summer school is a hot idea right now. could it work? New York Times, Feb. 5, https://www.nytimes.com/2021/02/05/us/summer-school-covid.html [accessed August 6, 2021].

Gollwitzer, A., Oettingen, G., Kirby, T. A., Duckworth, A. L., Mayer, D., 2011. Mental contrasting facilitates academic performance in school children. Motivation and Emotion 35 (4), 403–412.

Gollwitzer, P. M., Sheeran, P., 2006. Implementation intentions and goal achievement: A meta-analysis of effects and processes. Advances in Experimental Social Psychology 38, 69–119.

Goodman, R., Meltzer, H., Bailey, V., 1998. The strengths and difficulties questionnaire: A pilot study on the validity of the self-report version. European Child & Adolescent Psychiatry 7 (3), 125–130.

Gorard, S., Siddiqui, N., See, B. H., 2015. How effective is a summer school for catch-up attainment in english and maths? International Journal of Educational Research 73, 1–11.

Hanushek, E. A., 2011. The economic value of higher teacher quality. Economics of Education Review 30, 466–479.

Hanushek, E. A., Rivkin, S. G., 2006. Teacher quality. In: Handbook of the Economics of Education. Vol. 2. Elsevier, Amsterdam, 1051–1078.

Harris, D. N., 2009. Toward policy-relevant benchmarks for interpreting effect sizes: Combining effects with costs. Educational Evaluation and Policy Analysis 31 (1), 3–29.

Heckman, J. J., Stixrud, J., Urzua, S., 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. Journal of Labor Economics 24 (3), 411–482.

Holmlund, H., Silva, O., 2014. Targeting noncognitive skills to improve cognitive outcomes: Evidence from a remedial education intervention. Journal of Human Capital 8 (2), 126–160.

Imbens, G. W., Angrist, J. D., 1994. Identification and estimation of local average treatment effects. Econometrica 62 (2), 467–475.

Jacob, B. A., Lefgren, L., 2004. Remedial education and student achievement: A regression-discontinuity analysis. Review of Economics and Statistics 86 (1), 226–244.

Judge, T. A., Erez, A., Bono, J. E., Thoresen, C. J., 2003. The core self-evaluations scale: Development of a measure. Personnel Psychology 56 (2), 303–331.

Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., Borghans, L., 2014. Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. Working Paper No. 20749, National Bureau of Economic Research.

Kemple, J. J., Willner, C. J., 2008. Career academies: Long-term impacts on labor market outcomes, educational attainment, and transitions to adulthood. MDRC.

Kim, J. S., Quinn, D. M., 2013. The effects of summer reading on low-income children's literacy achievement from kindergarten to grade 8: A meta-analysis of classroom and home interventions. Review of Educational Research 83 (3), 386–431.

Kosse, F., Deckers, T., Pinger, P., Schildberg-Hörisch, H., Falk, A., 2020. The formation of prosociality: causal evidence on the role of social environment. Journal of Political Economy 128 (2), 434 – 467.

Kraft, M. A., 2020. Interpreting effect sizes of education interventions. Educational Researcher 49 (4), 241–253.

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., Martin-Glenn, M. L., 2006. Out-of-school-time programs: A meta-analysis of effects for at-risk students. Review of Educational Research 76 (2), 275–313.

Lavy, V., Schlosser, A., 2005. Targeted remedial education for underperforming teenagers: Costs and benefits. Journal of Labor Economics 23 (4), 839–874.

Lee, S. S., August, G. J., Bloomquist, M. L., Mathy, R., Realmuto, G. M., 2006. Implementing an evidence-based preventive intervention in neighborhood family centers: Examination of perceived barriers to program participation. Journal of Primary Prevention 27 (6), 573–597.

Locke, E. A., Latham, G. P., 1990. A theory of goal setting & task performance. Prentice-Hall, Inc, Englewood Cliffs, NJ.

Lynch, K., Kim, J. S., 2017. Effects of a summer mathematics intervention for low-income children: A randomized experiment. Educational Evaluation and Policy Analysis 39 (1), 31–53.

Mariano, L. T., Martorell, P., 2013. The academic effects of summer instruction and retention in new york city. Educational Evaluation and Policy Analysis 35 (1), 96–117.

Martins, P. S., 2017. Can non-cognitive skills programs improve achievement? quasi-experimental evidence from epis. Discussion Paper, Global Labor Organization (GLO).

Matsudaira, J. D., 2008. Mandatory summer school and student achievement. Journal of Econometrics 142 (2), 829 – 850.
URL http://www.sciencedirect.com/science/article/pii/S0304407607001194

McCord, J., 1978. A thirty-year follow-up of treatment effects. American Psychologist 33 (3), 284–289.

OECD, 2013. Pisa 2012 results: Ready to learn: Students' engagement, drive and self-beliefs (volume iii). OECD Publishing, Paris.

OECD, 2019. Pisa 2018 results (volume i): What students know and can do. OECD Publishing, Paris.

Rivkin, S. G., Hanushek, E. A., Kain, J. F., 2005. Teachers, schools, and academic achievement. Econometrica 73 (2), 417–458.

Rodriguez-Planas, N., 2012. Longer-term impacts of mentoring, educational services, and learning incentives: Evidence from a randomized trial in the united states. American Economic Journal: Applied Economics 4 (4), 121–39.

Schleicher, A., 2018. Pisa 2018: Insights and intepretations. OECD Publishing, Paris.

Schueler, B. E., 2020. Making the most of school vacation: A field experiment of small group math instruction. Education Finance and Policy 15 (2), 310–331.

Schueler, B. E., Goodman, J. S., Deming, D. J., 2017. Can states take over and turn around school districts? evidence from lawrence, massachusetts. Educational Evaluation and Policy Analysis 39 (2), 311–332.

Seligman, M. E., Park, N., Peterson, C., 2004. The values in action (via) classification of character strengths. Ricerche di Psicologia 27 (1), 63—-78.

Somers, M.-A., Welbeck, R., Grossman, J. B., Gooden, S., 2015. An analysis of the effects of an academic summer program for middle school students. New York: MDRC, March.

SWR, 2021. Heftige diskussion über kretschmann-vorschlag zu verkürzten schulferien [fierce discussion about kretschmann's proposal for shortened school holidays]. https://www.swr.de/swraktuell/baden-wuerttemberg/kurze-schulferien-wegen-corona-100.html [accessed August 6, 2021].

Tierney, J., Grossman, J., Resch, N., Ventures, P., of America, U. S., 1995. Making a difference: An impact study of big brothers/big sisters. Public/Private Ventures, Philadelphia, PA.

Tsukayama, E., Duckworth, A. L., Kim, B., 2013. Domain-specific impulsivity in school-age children. Developmental Science 16 (6), 879–893.

Undervisningsministeriet, 2017a. Uddannelsesparathed hos eleverne i 8. klasse – 2017. Styrelsen for IT og læring.

Undervisningsministeriet, 2017b. Uddannelsesparathed hos eleverne i 8. klasse – 2018. Styrelsen for IT og læring.

Urminsky, O., Hansen, C., Chernozhukov, V., 2016. Using double-lasso regression for principled variable selection. Available at SSRN 2733374.

Vischer, T., Dohmen, T., Falk, A., Huffman, D., Schupp, J., Sunde, U., Wagner, G. G., 2013. Validating an ultra-short survey measure of patience. Economics Letters 120 (2), 142–145.

YouGov, 2021. Teachers oppose extended school days and shorter summer holidays to help pupils catch up after covid. https://yougov.co.uk/topics/education/articles-reports/2021/03/30/teachers-oppose-extended-school-days-short-holiday [accessed August 6, 2021].

A.1 Appendix

A.1.1 Survey measures for non-cognitive skills

In our analysis, we draw on the following measures collected in the post-camp survey. We provide the English translations of the questions that were posed in Danish.

- **Grit: Grit Scale** (Duckworth and Quinn, 2009)

  1 = Strongly agree, 2 = Agree, 3 = Neither agree nor disagree, 4 = Disagree, 5 = Strongly disagree

  1. New ideas and projects sometimes distract me from previous ones.
  2. Setbacks don't discourage me.(R)
  3. I have been obsessed with a certain idea or project for a short time but later lost interest.
  4. I am a hard worker.(R)
  5. I often set a goal but later choose to pursue a different one.
  6. I have difficulty maintaining my focus on projects that take more than a few months to complete.
  7. I finish whatever I begin.(R)
  8. I am diligent.

- **Self-control: Domain-Specific Impulsivity Scale for Children (Tsukayama et al., 2013)**

  1 = At least once a day; 2 = Approx. once a week, 3 = Approx. 2-3 times a month, 4 = Approx. once a month, 5 = Almost never

  1. I forgot something I needed for school.
  2. I interrupted other pupils while they were talking.
  3. I said something rude.
  4. I couldn't find something because my table, closet, or bedroom was messy.
  5. I got really mad at home or at school.
  6. I couldn't remember what my teacher had asked me to do.
  7. I thought of something else while I should have listened.
  8. I talked back to my teacher or parent because I was angry or upset.

- **Self-concept: Core Self-Evaluations Scale** (Judge et al., 2003)

1 = Strongly agree, 2 = Agree, 3 = Neither agree nor disagree, 4 = Disagree, 5 = Strongly disagree

1. I am confident I get the success I deserve in life.(R)
2. Sometimes I feel depressed.
3. When I try, I generally succeed.(R)
4. Sometimes when I fail I feel worthless.
5. I complete tasks successfully.(R)
6. Sometimes, I do not feel in control of my work.
7. Overall, I am satisfied with myself.(R)
8. I am filled with doubts about my competence.
9. I determine what will happen in my life.(R)
10. I do not feel in control of my success in school.
11. I am capable of coping with most of my problems.(R)
12. There are times when things look pretty bleak and hopeless to me.

- **SDQ: Strengths and Difficulties Questionnaire for adolescents** (Goodman et al., 1998)

1=Not true, 2=Somewhat true, 3=Certainly true

1. I am restless, I cannot stay still for long.
2. I get a lot of headaches, stomach-aches or sickness.
3. I get very angry and often lose my temper.
4. I would rather be alone than with people of my age.
5. I usually do as I am told.(R)
6. I worry a lot.
7. I am constantly fidgeting or squirming.
8. I have one good friend or more.(R)
9. I fight a lot. I can make other people do what I want.
10. I am often unhappy, depressed or tearful.
11. Other people my age generally like me.(R)
12. I am easily distracted, I find it difficult to concentrate.
13. I am nervous in new situations. I easily lose confidence.
14. I am often accused of lying or cheating.

15. Other children or young people pick on me or bully me.
16. I think before I do things.(R)
17. I take things that are not mine from home, school or elsewhere.
18. I get along better with adults than with people my own age.
19. I have many fears, I am easily scared.
20. I finish the work I'm doing. My attention is good.(R)

    Prosocial scale (not included in the *Total difficulties score*):

21. I try to be nice to other people. I care about their feelings.
22. I usually share with others, for example CD's, games, food.
23. I am helpful if someone is hurt, upset or feeling ill.
24. I am kind to younger children.
25. I often offer to help others (parents, teachers, children).

- **Mindset: Malleability of ability to learn** (based on Dweck, 2006)

  1 = Strongly agree, 2 = Agree, 3 = Neither agree nor disagree, 4 = Disagree, 5 = Strongly disagree

  1. I can always improve my ability to learn no matter how old I am.(R)
  2. My ability to learn will never change.
  3. I am above the age, where it is possible to significantly improve my ability to learn.
  4. After a certain time during my childhood, I will no longer be able to improve my ability to learn.

- **Patience 1: 2 items from the GSOEP**

  1 = Strongly agree, 2 = Agree, 3 = Neither agree nor disagree, 4 = Disagree, 5 = Strongly disagree

  1. I do without today to be able to afford more tomorrow.(R)
  2. I prefer to have fun today and don't think about tomorrow.

- **Patience 2: 1-item patience question** (Vischer et al., 2013)

  Likert scale: 1= very impatient, 10= very patient

  Are you generally an impatient person, or someone who always shows great patience?

- **General risk preferences** (Dohmen et al., 2011)

Likert scale: 1=not at all willing to take risks, 10=very willing to take risks

Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?

- **Positive risk preferences**

  1 = Several times a day, 2 = Once a day, 3 = Several times a week, 4 = Once a week, 5 = Once a month, 6 = Never

  1. How often have you raised your hand in class during the last month?(R)
  2. How often have you volunteered to present something in class during the last month?(R)

- **Negative risk preferences**

  1 = Several times a day, 2 = Once a day, 3 = Several times a week, 4 = Once a week, 5 = Once a month, 6 = Never

  1. How often have you been drinking alcohol and / or smoking cigarettes during the last month?
  2. How often have you done something illegal (for example, jaywalked, stole something) during the last month?

A.2 Additional tables

Table A.1: Balance tests - all pupils in grade 8 (ITT)

	Round 1 and 3				Round 2			
	Camp+	Control	Diff.		Camp+	Camp	Diff.	
Language Comprehension								
- Grade 6	-0.004	0.028	-0.035		0.004	0.025	-0.028	
- Grade 4	-0.055	0.012	-0.063		0.038	0.040	-0.004	
- Grade 2	-0.001	-0.003	0.014		-0.014	0.007	-0.019	
Decoding								
- Grade 6	-0.015	0.032	-0.057		0.020	0.001	0.014	
- Grade 4	-0.026	0.018	-0.048		0.002	-0.025	0.026	
- Grade 2	-0.009	0.018	-0.018		-0.032	-0.002	-0.022	
Text Comprehension								
- Grade 6	-0.034	-0.011	-0.029		0.009	0.001	0.009	
- Grade 4	-0.046	-0.007	-0.039		-0.006	-0.015	0.013	
- Grade 2	-0.012	0.015	-0.021		-0.039	-0.003	-0.029	
Overall Reading								
- Grade 6	-0.020	0.019	-0.047		0.013	0.010	-0.002	
- Grade 4	-0.048	0.009	-0.057		0.013	0.000	0.013	
- Grade 2	-0.008	0.011	-0.009		-0.032	0.001	-0.027	
Numbers and Algebra								
- Grade 6	-0.077	0.033	-0.114	***	0.033	0.017	0.019	
- Grade 3	0.004	0.034	-0.027		0.022	-0.004	0.025	
Geometry								
- Grade 6	-0.051	0.028	-0.080	**	0.038	0.011	0.033	
- Grade 3	-0.007	-0.015	0.012		-0.023	0.021	-0.034	
Statistics and Probability								
- Grade 6	-0.042	0.020	-0.067		0.022	0.009	0.024	
- Grade 3	0.002	0.017	-0.008		0.003	0.009	-0.003	
Overall Math								
- Grade 6	-0.064	0.030	-0.098	**	0.035	0.014	0.029	
- Grade 3	0.000	0.014	-0.008		0.001	0.010	-0.004	
Educational interest at 8th grade assessment								
- 3-year High School	0.739	0.731	0.007		0.705	0.720	-0.012	
- Vocational training	0.526	0.523	-0.000		0.468	0.456	0.014	
- 2-year High School	0.229	0.193	0.024		0.102	0.088	0.014	
- Other	0.018	0.014	0.004		0.021	0.019	-0.001	
Personal ready at 8th grade assessment								
- 3-year High School	0.809	0.793	0.012		0.822	0.794	0.033	
- Vocational training	0.727	0.719	0.010		0.715	0.654	0.062	**
- 2-year High School	0.723	0.698	-0.015		0.717	0.674	0.047	
Social ready at 8th grade assessment								
- 3-year High School	0.863	0.862	0.003		0.882	0.867	0.018	

- Vocational training	0.805	0.804	0.011		0.799	0.765	0.029	
- 2-year High School	0.811	0.804	0.016		0.832	0.781	0.039	
Academic ready at 8th grade assessment								
- 3-year High School	0.587	0.607	-0.023		0.625	0.610	0.020	
- Vocational training	0.753	0.777	-0.027	*	0.786	0.755	0.037	**
- 2-year High School	0.754	0.778	-0.028	*	0.786	0.756	0.037	**
- GPA (std.)	-0.192	-0.168	-0.028		-0.137	-0.159	0.032	
Overall ready at 8th grade assessment								
- 3-year High School	0.655	0.661	-0.010		0.708	0.691	0.022	
- Vocational training	0.585	0.602	-0.017		0.582	0.535	0.049	*
- 2-year High School	0.639	0.637	-0.042		0.591	0.584	0.060	
Personality Traits								
- Grit	-0.002	-0.004	-0.009		0.009	-0.005	0.015	
- Self-control	0.014	-0.007	0.003		0.000	0.005	-0.006	
- Self-concept	0.000	0.011	0.002		0.012	-0.009	0.010	
- Mindset	0.004	0.007	0.003		-0.016	0.020	-0.036	
- General risk preferences	-0.020	0.021	-0.028		-0.009	0.014	-0.027	
- Positive risk preferences	-0.029	0.051	-0.067		0.001	-0.001	0.005	
- Negative risk preferences	0.040	-0.049	0.081		0.036	-0.028	0.065	
- Patience1	0.009	-0.015	0.019		-0.009	0.012	-0.014	
- Patience2	0.004	-0.005	0.005		0.009	-0.006	0.020	
- Self-control awareness	0.029	-0.043	0.055		-0.005	0.003	-0.003	
- Academic self-perception	-0.010	0.014	-0.026		0.033	-0.031	0.054	
- Academic self-concept	-0.014	0.020	-0.039		0.032	-0.027	0.053	
Strengths and Difficulties Questionnaire								
- Prosocial behaviour	-0.003	-0.001	-0.022		0.001	0.001	0.000	
- Emotional symptoms	0.013	-0.021	0.033		-0.007	0.002	0.000	
- Conduct problems	0.000	-0.004	0.025		0.016	-0.023	0.039	
- Hyperactivity/inattention	0.013	-0.014	0.046		-0.008	0.007	-0.016	
- Peer relationship problems	0.013	-0.030	0.055		-0.004	-0.008	0.006	
- Total difficulties score	0.015	-0.025	0.056		-0.003	-0.005	0.006	
Conscientiousness								
- Grade 7	-0.073	-0.026	-0.058	*	-0.001	-0.031	0.025	
- Grade 6	-0.005	0.013	-0.034		0.059	0.069	-0.023	
Agreeableness								
- Grade 7	-0.002	0.000	0.001		0.035	-0.014	0.055	
- Grade 6	0.011	0.055	-0.047	*	0.076	0.023	0.041	

Neuroticism								
- Grade 7	0.036	0.009	0.036		-0.073	0.029	-0.097	**
- Grade 6	-0.014	-0.056	0.054	**	-0.110	-0.069	-0.021	
Academic self-perception								
- Grade 7	-0.134	-0.103	-0.042		-0.108	-0.149	0.042	
- Grade 6	0.004	0.047	-0.067	**	0.044	0.019	0.019	
Academic well-being								
- Grade 7	-0.105	-0.058	-0.056	*	-0.037	-0.080	0.039	
- Grade 6	0.004	0.035	-0.053	*	0.062	0.057	-0.008	
Social well-being								
- Grade 7	-0.031	-0.009	-0.028		0.074	-0.011	0.079	*
- Grade 6	0.032	0.073	-0.056	*	0.122	0.085	0.018	
Order and quietness								
- Grade 7	-0.084	-0.061	-0.034		0.010	0.001	0.005	
- Grade 6	-0.092	-0.047	-0.051		0.023	0.027	-0.022	
Support and inspiration								
- Grade 7	-0.183	-0.217	0.027		-0.124	-0.210	0.078	
- Grade 6	-0.027	-0.033	-0.017		0.061	-0.062	0.107	**
Sick absence								
- Grade 7	3.132	3.518	-0.405	**	3.241	3.448	-0.179	
- Grade 6	3.180	3.126	0.057		3.118	3.231	-0.176	
Illegal absence								
- Grade 7	1.305	1.127	0.189		1.232	1.185	0.007	
- Grade 6	0.763	0.748	-0.002		0.526	0.685	-0.151	
Legal absence								
- Grade 7	1.644	1.501	0.103		1.680	1.618	0.051	
- Grade 6	1.308	1.250	0.053		1.269	1.296	-0.041	
Dyslexic information								
- Dyslexic	0.071	0.062	0.010		0.074	0.081	-0.006	
- Uncertain phonological	0.030	0.029	-0.001		0.025	0.026	-0.001	
- Not dyslexic	0.014	0.014	-0.001		0.012	0.013	-0.001	
- Not tested	0.885	0.895	-0.008		0.889	0.881	0.008	
- NOTA membership	0.092	0.083	0.007		0.095	0.114	-0.020	*
Schooling information								
- School starting age	6.213	6.203	0.009		6.220	6.197	0.016	
- Number of classes retaken	0.067	0.063	0.004		0.059	0.077	-0.011	
- Number of school changes	0.729	0.808	-0.069		0.732	0.740	0.005	
Child diagnosis								
- ADHD	0.013	0.011	0.001		0.003	0.002	0.001	
- Autisme	0.008	0.008	-0.001		0.000	0.001	-0.001	*
- OCD and anxiety	0.021	0.017	0.002		0.007	0.006	0.001	
- Other behavioral disorder	0.017	0.015	0.000		0.011	0.008	0.003	

Place of residence								
- Living with both parents	0.637	0.639	-0.003		0.629	0.636	-0.004	
- Living with one parents	0.351	0.349	0.002		0.360	0.352	0.005	
- Living with no parents	0.013	0.012	0.001		0.011	0.012	-0.002	
Etnicity								
- Danish	0.892	0.893	0.004		0.904	0.907	-0.003	
- Non-western	0.091	0.091	-0.003		0.080	0.077	0.002	
- Western	0.017	0.017	-0.000		0.016	0.015	0.001	
Age at birth								
- Mother	29.698	29.806	-0.114		29.784	30.059	-0.310	*
- Father	32.541	32.422	0.099		32.580	32.763	-0.222	
Income (1,000DKK)								
- Mother	278.126	290.945	-13.424	*	286.255	296.473	-10.506	
- Father	379.541	410.002	-31.230	***	400.040	407.164	-5.454	
Etnicity - Mother								
- Danish	0.863	0.851	0.017		0.873	0.878	-0.004	
- Non-western	0.109	0.117	-0.013		0.097	0.096	-0.000	
- Western	0.028	0.032	-0.003		0.030	0.025	0.004	
Etnicity - Father								
- Danish	0.872	0.864	0.011		0.881	0.890	-0.007	
- Non-western	0.101	0.107	-0.010		0.091	0.084	0.005	
- Western	0.028	0.029	-0.001		0.028	0.026	0.002	
Employment status - Mother								
- No benefits	0.721	0.750	-0.028	**	0.754	0.759	-0.004	
- ALMP	0.158	0.146	0.012		0.139	0.142	-0.004	
- SU	0.025	0.021	0.003		0.017	0.024	-0.007	*
- Pension/leave	0.096	0.083	0.013	**	0.090	0.075	0.015	*
Employment status - Father								
- No benefits	0.813	0.829	-0.015		0.825	0.823	0.005	
- ALMP	0.091	0.088	0.003		0.083	0.080	0.002	
- SU	0.005	0.006	-0.001		0.004	0.005	-0.002	
- Pension/leave	0.091	0.076	0.013	*	0.089	0.092	-0.005	
Education - Mother								
- No education	0.006	0.007	-0.002		0.005	0.010	-0.006	***
- Primary School	0.160	0.151	0.011		0.140	0.153	-0.011	
- High School	0.057	0.053	0.002		0.052	0.060	-0.006	
- Voccational	0.405	0.402	0.008		0.416	0.399	0.014	
- Short University Degree	0.059	0.053	0.002		0.048	0.047	0.001	
- Medium University Degree	0.258	0.253	0.004		0.269	0.259	0.013	
- Long University Degree	0.055	0.080	-0.026	**	0.069	0.073	-0.005	

Education - Father							
- No education	0.014	0.013	0.001		0.006	0.012	-0.006	**
- Primary School	0.202	0.188	0.013		0.191	0.199	-0.008	
- High School	0.037	0.043	-0.007	*	0.039	0.046	-0.008	
- Voccational	0.499	0.471	0.033	**	0.493	0.451	0.043	**
- Short University Degree	0.072	0.079	-0.009		0.084	0.079	0.005	
- Medium University Degree	0.107	0.116	-0.009		0.115	0.121	-0.005	
- Long University Degree	0.068	0.090	-0.022	**	0.072	0.093	-0.020	
Observations	4,573	4,851	9,424		3,144	2,991	6,135	

*Note: ***p<0.01; **p<0.05; *p<0.1. Column 1-2 and 4-5 report the mean for each treatment group. Column 3 and 6 report the differences between the two treatment groups by regressing each baseline covariate on the treatment indicator, including randomization strata fixed effect and clustering at the school level. Baseline data are obtained from register data collected before the beginning of the intervention. Pupils are allowed to have more than one educational interest at the 8th grade risk assessment. They are only evaluated on personal, social and overall educational readiness for educations in which they indicate interest. The table is based on non-missing data.*

Table A.2: Balance tests - all NR pupils in grade 8 (ITT-NR)

	Round 1 and 3				**Round 2**			
	Camp+	**Control**	**Diff.**		**Camp+**	**Camp**	**Diff.**	
Language Comprehension								
- Grade 6	-0.406	-0.335	-0.086		-0.404	-0.375	-0.033	
- Grade 4	-0.438	-0.352	-0.089	*	-0.366	-0.342	-0.015	
- Grade 2	-0.322	-0.309	-0.008		-0.345	-0.299	-0.054	
Decoding								
- Grade 6	-0.482	-0.393	-0.098	**	-0.444	-0.459	0.002	
- Grade 4	-0.469	-0.387	-0.080	**	-0.415	-0.469	0.047	
- Grade 2	-0.439	-0.412	-0.024		-0.487	-0.432	-0.064	
Text Comprehension								
- Grade 6	-0.514	-0.493	-0.039		-0.492	-0.531	0.029	
- Grade 4	-0.484	-0.481	-0.006		-0.458	-0.491	0.030	
- Grade 2	-0.435	-0.407	-0.018		-0.488	-0.436	-0.062	
Overall Reading								
- Grade 6	-0.544	-0.473	-0.087	*	-0.518	-0.528	-0.001	
- Grade 4	-0.526	-0.461	-0.066		-0.471	-0.496	0.024	
- Grade 2	-0.455	-0.429	-0.019		-0.505	-0.446	-0.069	

Numbers and Algebra								
- Grade 6	-0.532	-0.407	-0.131	***	-0.386	-0.445	0.061	
- Grade 3	-0.399	-0.322	-0.076	*	-0.344	-0.410	0.055	
Geometry								
- Grade 6	-0.474	-0.397	-0.086	**	-0.408	-0.489	0.085	*
- Grade 3	-0.390	-0.357	-0.025		-0.373	-0.369	0.002	
Statistics and Probability								
- Grade 6	-0.512	-0.439	-0.083	**	-0.423	-0.496	0.077	
- Grade 3	-0.445	-0.373	-0.066		-0.384	-0.430	0.033	
Overall Math								
- Grade 6	-0.569	-0.466	-0.113	***	-0.455	-0.534	0.083	*
- Grade 3	-0.464	-0.396	-0.063		-0.413	-0.453	0.034	
Educational interest at 8th grade assessment								
- 3-year High School	0.627	0.627	0.004		0.551	0.573	-0.014	
- Vocational training	0.636	0.640	-0.005		0.606	0.602	0.009	
- 2-year High School	0.266	0.229	0.040		0.131	0.107	0.019	
- Other	0.045	0.036	0.011		0.057	0.050	-0.000	
Personal ready at 8th grade assessment								
- 3-year High School	0.444	0.390	0.042		0.393	0.334	0.071	*
- Vocational training	0.439	0.420	0.016		0.412	0.325	0.084	**
- 2-year High School	0.391	0.347	-0.011		0.354	0.276	0.056	
Social ready at 8th grade assessment								
- 3-year High School	0.602	0.594	0.011		0.597	0.570	0.021	
- Vocational training	0.601	0.596	0.022		0.586	0.542	0.029	
- 2-year High School	0.584	0.577	0.040		0.616	0.514	0.054	
Academic ready at 8th grade assessment								
- 3-year High School	0.085	0.108	-0.031	**	0.120	0.096	0.019	
- Vocational training	0.382	0.428	-0.051	**	0.420	0.366	0.062	**
- 2-year High School	0.383	0.430	-0.052	**	0.421	0.367	0.061	**
- GPA (std.)	-1.062	-1.008	-0.061	**	-1.037	-1.100	0.068	
Overall ready at 8th grade assessment								
- 3-year High School	0.000	0.000	0.000		0.002	0.000	0.002	
- Vocational training	0.153	0.178	-0.024		0.138	0.093	0.041	*
- 2-year High School	0.200	0.211	-0.062		0.149	0.121	0.022	
Personality Traits								
- Grit	-0.411	-0.364	-0.061		-0.428	-0.445	0.009	
- Self-control	-0.161	-0.201	0.034		-0.237	-0.217	-0.041	
- Self-concept	-0.003	0.010	0.028		-0.283	-0.330	0.038	
- Mindset	0.001	0.005	0.002		-0.278	-0.220	-0.044	
- General risk preferences	-0.023	0.070	-0.095	**	-0.012	0.009	-0.024	
- Positive risk preferences	-0.450	-0.298	-0.146	**	-0.385	-0.373	-0.005	
- Negative risk preferences	-0.131	-0.290	0.149	*	-0.183	-0.264	0.073	

- Patience1	-0.006	-0.020	0.016		-0.175	-0.172	0.007	
- Patience2	-0.085	-0.148	0.084		-0.135	-0.169	0.036	
- Self-control awareness	0.233	0.145	0.069		0.197	0.202	0.009	
- Academic self-perception	-0.320	-0.347	0.047		-0.349	-0.409	0.035	
- Academic self-concept	-0.527	-0.545	0.035		-0.521	-0.605	0.071	
Strengths and Difficulties Questionnaire								
- Prosocial behaviour	-0.151	-0.225	0.031		-0.188	-0.164	-0.013	
- Emotional symptoms	0.104	0.115	-0.023		0.077	0.096	-0.007	
- Conduct problems	0.363	0.366	0.016		0.355	0.337	0.022	
- Hyperactivity/inattention	0.338	0.372	-0.044		0.361	0.391	-0.033	
- Peer relationship problems	0.318	0.238	0.096		0.294	0.259	0.043	
- Total difficulties score	0.381	0.377	0.006		0.365	0.370	0.001	
Conscientiousness								
- Grade 7	-0.442	-0.365	-0.087	**	-0.380	-0.435	0.046	
- Grade 6	-0.333	-0.344	-0.006		-0.285	-0.308	-0.004	
Agreeableness								
- Grade 7	-0.215	-0.243	0.034		-0.220	-0.299	0.086	*
- Grade 6	-0.169	-0.122	-0.044		-0.164	-0.251	0.081	*
Neuroticism								
- Grade 7	0.131	0.147	0.004		0.088	0.227	-0.135	**
- Grade 6	0.139	0.096	0.057		0.055	0.110	-0.033	
Academic self-perception								
- Grade 7	-0.550	-0.530	-0.040		-0.539	-0.573	0.016	
- Grade 6	-0.327	-0.308	-0.048		-0.328	-0.333	-0.008	
Academic well-being								
- Grade 7	-0.537	-0.483	-0.066		-0.497	-0.544	0.033	
- Grade 6	-0.377	-0.370	-0.031		-0.348	-0.366	-0.004	
Social well-being								
- Grade 7	-0.140	-0.191	0.036		-0.110	-0.233	0.120	**
- Grade 6	-0.120	-0.109	-0.034		-0.076	-0.120	0.027	
Order and quietness								
- Grade 7	-0.191	-0.190	-0.005		-0.114	-0.147	0.028	
- Grade 6	-0.163	-0.201	0.039		-0.078	-0.106	0.005	
Support and inspiration								
- Grade 7	-0.300	-0.386	0.077		-0.334	-0.383	0.037	
- Grade 6	-0.137	-0.184	0.041		-0.085	-0.201	0.097	
Sick absence								
- Grade 7	3.924	4.386	-0.476	*	4.178	4.402	-0.178	
- Grade 6	4.026	3.869	0.181		4.239	4.256	-0.076	
Illegal absence								

- Grade 7	2.260	1.845	0.373		2.193	2.071	0.054	
- Grade 6	1.281	1.170	0.051		0.963	1.178	-0.180	
Legal absence								
- Grade 7	1.826	1.648	0.114		1.834	1.729	0.071	
- Grade 6	1.398	1.309	0.094		1.495	1.315	0.137	
Dyslexic information								
- Dyslexic	0.133	0.106	0.028	**	0.133	0.144	-0.008	
- Uncertain phonological	0.048	0.045	0.000		0.034	0.046	-0.009	
- Not dyslexic	0.023	0.023	-0.001		0.020	0.025	-0.004	
- Not tested	0.796	0.826	-0.028		0.813	0.786	0.022	
- NOTA membership	0.167	0.141	0.022		0.171	0.205	-0.037	*
Schooling information								
- School starting age	6.250	6.253	-0.003		6.271	6.245	0.018	
- Number of classes retaken	0.113	0.101	0.004		0.112	0.116	-0.001	
- Number of school changes	0.902	0.965	-0.053		0.903	0.951	-0.009	
Child diagnosis								
- ADHD	0.023	0.022	-0.002		0.003	0.004	-0.001	
- Autisme	0.014	0.013	-0.001		0.000	0.001	-0.001	
- OCD and anxiety	0.026	0.021	0.003		0.012	0.006	0.007	*
- Other behavioral disorder	0.025	0.022	0.001		0.017	0.012	0.005	
Place of residence								
- Living with both parents	0.561	0.559	0.006		0.514	0.547	-0.022	
- Living with one parents	0.418	0.419	-0.006		0.468	0.433	0.024	
- Living with no parents	0.021	0.022	0.000		0.019	0.020	-0.002	
Etnicity								
- Danish	0.849	0.847	0.007		0.854	0.865	-0.013	
- Non-western	0.126	0.130	-0.008		0.124	0.111	0.016	
- Western	0.025	0.023	0.001		0.022	0.024	-0.002	
Age at birth								
- Mother	29.094	29.077	0.030		28.916	29.061	-0.162	
- Father	32.248	31.991	0.247		31.915	32.218	-0.348	
Income (1,000DKK)								
- Mother	231.962	243.491	-12.425		243.162	241.087	1.635	
- Father	331.748	346.997	-15.889		336.289	347.120	-10.610	
Etnicity - Mother								
- Danish	0.814	0.798	0.024		0.826	0.829	-0.005	
- Non-western	0.147	0.161	-0.021		0.140	0.136	0.008	
- Western	0.039	0.041	-0.002		0.034	0.034	-0.002	
Etnicity - Father								

- Danish	0.831	0.823	0.014		0.843	0.863	-0.025	
- Non-western	0.134	0.146	-0.018		0.131	0.111	0.024	
- Western	0.034	0.031	0.003		0.026	0.026	0.001	
Employment status - Mother								
- No benefits	0.642	0.661	-0.014		0.668	0.652	0.016	
- ALMP	0.216	0.203	0.008		0.194	0.208	-0.018	
- SU	0.031	0.027	0.003		0.024	0.035	-0.012	
- Pension/leave	0.111	0.109	0.003		0.115	0.104	0.014	
Employment status - Father								
- No benefits	0.750	0.767	-0.018		0.766	0.735	0.039	*
- ALMP	0.133	0.127	0.007		0.112	0.125	-0.015	
- SU	0.003	0.009	-0.006	**	0.005	0.006	-0.001	
- Pension/leave	0.113	0.097	0.017		0.117	0.134	-0.023	*
Education - Mother								
- No education	0.008	0.011	-0.002		0.009	0.014	-0.006	
- Primary School	0.245	0.232	0.015		0.221	0.252	-0.026	
- High School	0.057	0.050	0.006		0.049	0.054	-0.005	
- Voccational	0.433	0.449	-0.014		0.462	0.447	0.017	
- Short University Degree	0.045	0.037	0.009		0.043	0.036	0.004	
- Medium University Degree	0.181	0.185	-0.005		0.174	0.167	0.004	
- Long University Degree	0.030	0.037	-0.009		0.043	0.029	0.013	*
Education - Father								
- No education	0.017	0.016	0.000		0.008	0.014	-0.005	
- Primary School	0.287	0.268	0.015		0.275	0.312	-0.034	
- High School	0.032	0.043	-0.013	*	0.039	0.032	0.005	
- Voccational	0.511	0.510	0.012		0.522	0.461	0.060	***
- Short University Degree	0.051	0.052	-0.003		0.067	0.063	0.001	
- Medium University Degree	0.068	0.067	0.000		0.058	0.072	-0.012	
- Long University Degree	0.033	0.043	-0.012	*	0.031	0.046	-0.014	
Observations	1,841	1,918	3,759		1,174	1,161	2,335	

*Note: ***p<0.01; **p<0.05; *p<0.1. Column 1-2 and 4-5 report the mean for each treatment group. Column 3 and 6 report the differences between the two treatment groups by regressing each baseline covariate on the treatment indicator, including randomization strata fixed effect and clustering at the school level. Baseline data are obtained from register data collected before the beginning of the intervention. Pupils are allowed to have more than one educational interest at the 8th grade risk assessment. They are only evaluated on personal, social and overall educational readiness for educations in which they indicate interest. The table is based on non-missing data.*

Table A.3: Balance tests - camp selected pupils (ATT)

	Round 1 and 3				Round 2			
	Camp+	**Control**	**Diff.**		**Camp+**	**Camp**	**Diff.**	
Language Comprehension								
- Grade 6	-0.522	-0.391	-0.148	**	-0.530	-0.461	-0.064	
- Grade 4	-0.572	-0.408	-0.160	***	-0.516	-0.460	-0.044	
- Grade 2	-0.379	-0.349	-0.017		-0.446	-0.373	-0.079	
Decoding								
- Grade 6	-0.609	-0.474	-0.146	***	-0.595	-0.557	-0.047	
- Grade 4	-0.575	-0.463	-0.104	**	-0.593	-0.597	-0.000	
- Grade 2	-0.565	-0.460	-0.093	*	-0.647	-0.523	-0.136	**
Text Comprehension								
- Grade 6	-0.637	-0.554	-0.101	**	-0.640	-0.627	-0.018	
- Grade 4	-0.584	-0.561	-0.023		-0.588	-0.601	0.010	
- Grade 2	-0.553	-0.455	-0.094	*	-0.663	-0.531	-0.148	**
Overall Reading								
- Grade 6	-0.686	-0.550	-0.153	***	-0.682	-0.636	-0.050	
- Grade 4	-0.654	-0.541	-0.108	**	-0.646	-0.631	-0.013	
- Grade 2	-0.570	-0.481	-0.078		-0.671	-0.545	-0.139	**
Numbers and Algebra								
- Grade 6	-0.650	-0.461	-0.196	***	-0.486	-0.509	0.024	
- Grade 3	-0.495	-0.411	-0.083		-0.422	-0.531	0.095	*
Geometry								
- Grade 6	-0.557	-0.440	-0.121	***	-0.493	-0.522	0.032	
- Grade 3	-0.482	-0.420	-0.054		-0.453	-0.481	0.023	
Statistics and Probability								
- Grade 6	-0.634	-0.508	-0.133	***	-0.515	-0.572	0.058	
- Grade 3	-0.552	-0.452	-0.095	*	-0.508	-0.524	0.002	
Overall Math								
- Grade 6	-0.690	-0.528	-0.168	***	-0.558	-0.599	0.043	
- Grade 3	-0.575	-0.482	-0.087	*	-0.519	-0.576	0.045	
Educational interest at 8th grade assessment								
- High School	0.557	0.572	-0.019		0.488	0.548	-0.056	
- Vocational traning	0.690	0.675	0.015		0.694	0.652	0.043	
- 2-year High School	0.247	0.198	0.045		0.136	0.110	0.023	
- Other	0.026	0.012	0.016		0.017	0.021	-0.007	
Personal ready at 8th grade assessment								
- 3-year High School	0.323	0.203	0.109	***	0.286	0.224	0.055	
- Vocational traning	0.411	0.306	0.090	**	0.381	0.283	0.088	**
- 2-year High School	0.412	0.260	0.122	*	0.277	0.241	-0.028	
Social ready at 8th grade assessment								
- 2-year High School	0.525	0.480	0.034		0.542	0.517	0.016	

- Vocational training	0.595	0.540	0.060		0.568	0.512	0.039
- 2-year High School	0.596	0.510	0.134		0.607	0.497	0.051
Academic ready at 8th grade assessment							
- 3-year High School	0.048	0.099	-0.052	***	0.073	0.072	-0.001
- Vocational training	0.246	0.311	-0.075	**	0.298	0.295	0.009
- 2-year High School	0.246	0.312	-0.076	**	0.298	0.295	0.009
- GPA (std.)	-1.185	-1.109	-0.085	**	-1.153	-1.170	0.026
Overall ready at 8th grade assessment							
- 3-year High School	0.000	0.000	0.000		0.003	0.000	0.004
- Vocational training	0.083	0.066	0.005		0.073	0.068	0.002
- 2-year High School	0.177	0.115	0.047		0.058	0.110	-0.034
Personality Traits							
- Grit	-0.508	-0.432	-0.096	**	-0.477	-0.509	0.026
- Self-control	-0.151	-0.235	0.073		-0.325	-0.247	-0.075
- Self-concept	0.049	0.011	0.062		-0.312	-0.369	0.052
- Mindset	0.045	0.023	0.015		-0.316	-0.266	-0.039
- General risk preferences	-0.044	0.109	-0.142	***	0.030	-0.006	0.035
- Positive risk preferences	-0.520	-0.392	-0.133		-0.458	-0.400	-0.061
- Negative risk preferences	-0.046	-0.304	0.237	**	-0.228	-0.253	0.031
- Patience1	-0.009	-0.027	0.023		-0.200	-0.174	-0.010
- Patience2	-0.081	-0.113	0.030		-0.118	-0.210	0.092
- Self-control awareness	0.311	0.182	0.126	*	0.252	0.231	0.006
- Academic self-perception	-0.382	-0.313	-0.074		-0.384	-0.439	0.047
- Academic self-concept	-0.655	-0.597	-0.068		-0.602	-0.695	0.087
Strengths and Difficulties Questionnaire							
- Prosocial behaviour	-0.158	-0.246	0.073		-0.262	-0.193	-0.071
- Emotional symptoms	0.077	0.083	-0.018		0.055	0.100	-0.041
- Conduct problems	0.351	0.441	-0.077		0.418	0.389	0.024
- Hyperactivity/inattention	0.389	0.396	-0.000		0.426	0.452	-0.033
- Peer relationship problems	0.348	0.294	0.055		0.332	0.319	0.013
- Total difficulties score	0.396	0.411	-0.013		0.412	0.429	-0.020
Conscientiousness							
- Grade 7	-0.515	-0.387	-0.148	***	-0.434	-0.503	0.063
- Grade 6	-0.383	-0.383	-0.016		-0.344	-0.366	0.010
Agreeableness							
- Grade 7	-0.214	-0.320	0.112	**	-0.285	-0.343	0.071
- Grade 6	-0.153	-0.228	0.080		-0.211	-0.316	0.101 *

Neuroticism							
- Grade 7	0.149	0.174	-0.004		0.126	0.225	-0.090
- Grade 6	0.110	0.101	0.025		0.148	0.110	0.044
Academic self-perception							
- Grade 7	-0.696	-0.604	-0.119	**	-0.611	-0.658	0.039
- Grade 6	-0.402	-0.365	-0.068		-0.412	-0.437	0.026
Academic well-being							
- Grade 7	-0.652	-0.534	-0.139	***	-0.576	-0.617	0.036
- Grade 6	-0.436	-0.432	-0.023		-0.436	-0.440	-0.003
Social well-being							
- Grade 7	-0.161	-0.205	0.027		-0.107	-0.221	0.104
- Grade 6	-0.105	-0.116	-0.010		-0.139	-0.133	-0.011
Order and quietness							
- Grade 7	-0.179	-0.204	0.019		-0.135	-0.163	0.034
- Grade 6	-0.118	-0.229	0.086		-0.106	-0.096	-0.008
Support and inspiration							
- Grade 7	-0.269	-0.418	0.134	**	-0.319	-0.370	0.047
- Grade 6	-0.096	-0.215	0.107	*	-0.093	-0.195	0.107 *
Sick absence							
- Grade 7	3.740	4.285	-0.536	*	4.172	4.015	0.214
- Grade 6	3.790	3.926	-0.102		4.109	3.898	0.112
Illegal absence							
- Grade 7	2.182	2.087	0.040		1.716	1.791	-0.109
- Grade 6	1.354	1.204	0.120		0.826	1.060	-0.244
Legal absence							
- Grade 7	1.699	1.702	-0.032		1.698	1.643	0.103
- Grade 6	1.330	1.329	0.016		1.346	1.248	0.077
Dyslexic information							
- Dyslexic	0.176	0.113	0.063	***	0.165	0.168	-0.003
- Uncertain phonological	0.051	0.044	0.008		0.041	0.060	-0.015
- Not dyslexic	0.028	0.024	0.003		0.022	0.028	-0.006
- Not tested	0.745	0.819	-0.074	***	0.772	0.743	0.023
- NOTA membership	0.200	0.146	0.052	**	0.201	0.243	-0.042
Schooling information							
- School starting age	6.252	6.259	-0.008		6.273	6.257	0.007
- Number of classes retaken	0.113	0.103	0.005		0.113	0.129	-0.011
- Number of school changes	0.901	0.923	-0.008		0.921	0.925	0.006
Child diagnosis							
- ADHD	0.018	0.018	-0.002		0.000	0.005	-0.005 **
- Autisme	0.006	0.004	0.003		0.000	0.001	-0.001
- OCD and anxiety	0.012	0.018	-0.007		0.008	0.005	0.003
- Other behavioral disorder	0.017	0.025	-0.009		0.025	0.016	0.011 *

Place of residence								
- Living with both parents	0.553	0.527	0.026		0.504	0.542	-0.033	
- Living with one parents	0.426	0.447	-0.020		0.482	0.443	0.035	
- Living with no parents	0.021	0.026	-0.005		0.014	0.015	-0.002	
Etnicity								
- Danish	0.842	0.855	-0.002		0.868	0.861	0.007	
- Non-western	0.133	0.119	0.004		0.110	0.114	-0.003	
- Western	0.025	0.026	-0.002		0.022	0.025	-0.004	
Age at birth								
- Mother	28.949	28.904	0.008		28.823	28.962	-0.126	
- Father	31.910	31.843	0.029		31.779	31.965	-0.205	
Income (1,000DKK)								
- Mother	224.790	241.804	-12.816		235.731	234.917	0.339	
- Father	313.529	357.524	-42.531	***	332.146	337.463	-3.578	
Etnicity - Mother								
- Danish	0.809	0.804	0.019		0.842	0.825	0.016	
- Non-western	0.153	0.148	-0.008		0.124	0.142	-0.015	
- Western	0.038	0.048	-0.011		0.034	0.033	-0.001	
Etnicity - Father								
- Danish	0.821	0.840	-0.008		0.863	0.864	-0.001	
- Non-western	0.143	0.129	0.003		0.109	0.110	-0.001	
- Western	0.036	0.031	0.005		0.028	0.026	0.002	
Employment status - Mother								
- No benefits	0.638	0.650	-0.004		0.663	0.657	0.010	
- ALMP	0.221	0.213	-0.000		0.190	0.212	-0.028	
- SU	0.040	0.026	0.013	*	0.023	0.041	-0.019	**
- Pension/leave	0.102	0.111	-0.010		0.125	0.090	0.037	**
Employment status - Father								
- No benefits	0.724	0.776	-0.053	**	0.757	0.738	0.027	
- ALMP	0.152	0.118	0.034	*	0.114	0.123	-0.013	
- SU	0.004	0.008	-0.005		0.008	0.006	0.002	
- Pension/leave	0.120	0.097	0.024		0.120	0.134	-0.016	
Education - Mother								
- No education	0.010	0.012	-0.002		0.003	0.017	-0.013	**
- Primary School	0.266	0.223	0.041	**	0.228	0.245	-0.017	
- High School	0.050	0.052	-0.004		0.051	0.056	-0.006	
- Voccational	0.452	0.469	-0.010		0.500	0.466	0.035	
- Short University Degree	0.040	0.032	0.008		0.036	0.039	-0.003	
- Medium University Degree	0.155	0.182	-0.030	*	0.149	0.158	-0.011	
- Long University Degree	0.028	0.030	-0.002		0.033	0.019	0.014	*

Education - Father								
- No education	0.020	0.020	-0.001		0.007	0.013	-0.007	
- Primary School	0.304	0.262	0.036		0.294	0.295	-0.002	
- High School	0.032	0.037	-0.008		0.036	0.028	0.006	
- Voccational	0.504	0.518	-0.001		0.548	0.491	0.059	**
- Short University Degree	0.053	0.045	0.006		0.051	0.060	-0.010	
- Medium University Degree	0.060	0.075	-0.016		0.043	0.073	-0.027	**
- Long University Degree	0.028	0.042	-0.016	**	0.022	0.042	-0.019	*
Observations	955	1,103	2,058		631	748	1,379	

*Note: ***p<0.01; **p<0.05; *p<0.1. Column 1-2 and 4-5 report the mean for each treatment group. Column 3 and 6 report the differences between the two treatment groups by regressing each baseline covariate on the treatment indicator, including randomization strata fixed effect and clustering at the school level. Baseline data are obtained from register data collected before the beginning of the intervention. Pupils are allowed to have more than one educational interest at the 8th grade risk assessment. They are only evaluated on personal, social and overall educational readiness for educations in which they indicate interest. The table is based on non-missing data.*

Table A.4: Short-run effects of *Camp+* vs. *Camp*: performance in the national reading test in grade 8

	Language Comprehension		Decoding		Text Comprehension		Overall	
ATT	0.047	0.081	0.029	0.046	0.070	0.071	0.060	0.078
	(0.081)	(0.076)	(0.051)	(0.047)	(0.060)	(0.048)	(0.061)	(0.051)
Mean outcome, Camp	-.373	-.373	-.622	-.622	-.704	-.704	-.700	-.700
R-squared	.014	.211	.018	.426	.025	.376	.015	.432
Observations	1,316	1,316	1,316	1,316	1,316	1,316	1,316	1,316
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Baseline treatment category is Camp. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.5: Short-run effects of *Camp+* vs. *Camp*: performance in the national math test in grade 8

	Numbers and Algebra		Geometry		Statistics and Probability		Overall	
ATT	-0.019	-0.030	-0.020	-0.040	-0.012	-0.021	-0.018	-0.037
	(0.057)	(0.042)	(0.063)	(0.047)	(0.066)	(0.047)	(0.063)	(0.046)
Mean outcome, Camp	-.710	-.710	-.707	-.707	-.669	-.669	-.751	-.751
R-squared	.015	.437	.016	.414	.012	.428	.015	.512
Observations	1,321	1,321	1,321	1,321	1,321	1,321	1,321	1,321
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Baseline treatment category is Camp. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.6: Effects of the camp on the performance in the final exams in grade 9, split-up in the oral and written parts

	Danish				Math			
	Written		Oral		Written		Oral	
ITT	-0.043	-0.007	-0.070	-0.040	0.020	0.064	0.013	0.108
	(0.055)	(0.035)	(0.047)	(0.037)	(0.061)	(0.047)	(0.100)	(0.087)
Mean outcome, Control	-.059	-.059	-.038	-.038	-.102	-.102	-.100	-.100
R-squared	.015	.677	.011	.388	.015	.664	.053	.465
Observations	4,759	4,759	4,722	4,722	4,743	4,743	875	875
ITT-NR	-0.017	0.011	-0.090*	-0.052	0.028	0.096*	0.027	0.085
	(0.051)	(0.039)	(0.052)	(0.047)	(0.058)	(0.050)	(0.082)	(0.112)
LATE	-0.036	0.023	-0.183*	-0.105	0.057	0.197*	0.055	0.174
	(0.104)	(0.080)	(0.107)	(0.097)	(0.119)	(0.101)	(0.167)	(0.229)
Mean outcome, Control	-.732	-.732	-.581	-.581	-.778	-.778	-.673	-.673
R-squared	.019	.529	.018	.268	.013	.482	.056	.295
Observations	1,879	1,879	1,847	1,847	1,864	1,864	349	349
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Outcome only exists for round 1. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Oral math exam is an extract exam. Standard errors in parentheses are clustered at the school level.*

Table A.7: Effects of the camp on the education choice three years after the camp

	3-year High School		Vocational training		2-year High School		No educ.	
ITT	-0.036	-0.016	0.038**	0.027**	-0.006	-0.009	0.002	-0.004
	(0.028)	(0.016)	(0.017)	(0.013)	(0.010)	(0.009)	(0.012)	(0.009)
Mean outcome, Control	.607	.607	.201	.201	.091	.091	.095	.095
R-squared	.016	.420	.008	.238	.003	.063	.005	.152
Observations	4,820	4,820	4,820	4,820	4,820	4,820	4,820	4,820
ITT-NR	-0.023	0.001	0.062***	0.049***	-0.010	-0.008	-0.027	-0.037**
	(0.032)	(0.023)	(0.022)	(0.019)	(0.016)	(0.016)	(0.021)	(0.018)
LATE	-0.048	0.003	0.127***	0.101***	-0.020	-0.017	-0.055	-0.076**
	(0.066)	(0.047)	(0.046)	(0.038)	(0.032)	(0.034)	(0.043)	(0.038)
Mean outcome, Control	.342	.342	.320	.320	.130	.130	.193	.193
R-squared	.033	.333	.020	.185	.008	.057	.012	.163
Observations	1,931	1,931	1,931	1,931	1,931	1,931	1,931	1,931
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Outcome only exists for round 1. Educational enrollment is measured September* 31^{th} *three years after the camp. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.8: Long-run effects of *Camp+* vs. *Camp*: Performance in the final tests in grade 9 and likelihood of being enrolled in no education 1.5 years after the camp

	Danish score		Math score		No education	
ATT	0.008	0.021	-0.022	-0.055	-0.004	0.002
	(0.050)	(0.042)	(0.056)	(0.041)	(0.013)	(0.013)
Mean outcome, Camp	-.955	-.955	-.835	-.835	.066	.066
R-squared	.028	.489	.034	.483	.015	.141
Observations	1,317	1,317	1,310	1,310	1,344	1,344
Strata	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Baseline treatment category is Camp. Educational enrollment is measured September* 31^{th} *one year after the camp. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.9a: Effects of the camp on non-cognitive skills, measured by psychological scales

	Grit		Self-control		Self-concept		Mindset		Total difficult score	
ITT	-0.001	0.009	0.016	0.027	0.013	0.005	0.029	0.045	0.010	-0.010
	(0.034)	(0.021)	(0.030)	(0.022)	(0.030)	(0.020)	(0.044)	(0.046)	(0.037)	(0.030)
Mean outcome, Control	-.010	-.010	-.018	-.018	.001	.001	-.025	-.025	.006	.006
R-squared	.006	.490	.008	.414	.006	.487	.009	.221	.016	.403
Observations	6,993	6,993	6,976	6,976	6,957	6,957	3,533	3,533	6,876	6,876
ITT-NR	-0.017	0.005	0.012	0.007	0.009	-0.023	0.075	0.099	-0.002	-0.015
	(0.042)	(0.032)	(0.041)	(0.034)	(0.036)	(0.030)	(0.058)	(0.064)	(0.046)	(0.043)
LATE	-0.036	0.011	0.025	0.014	0.018	-0.047	0.154	0.202	-0.004	-0.032
	(0.086)	(0.065)	(0.084)	(0.070)	(0.073)	(0.062)	(0.119)	(0.131)	(0.095)	(0.087)
Mean outcome, Control	-.374	-.374	-.161	-.161	.040	.040	-.321	-.321	.346	.346
R-squared	.012	.410	.015	.411	.081	.485	.023	.187	.017	.350
Observations	2,550	2,550	2,539	2,539	2,529	2,529	1,249	1,249	2,481	2,481
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Mindset, Positive and Negative risk preference and Patience1 are only measured in round 3. Standard errors in parentheses are clustered at the school level.*

Table A.9b: Effects of the camp on non-cognitive skills, measured by psychological scales - cont'd

	General risk preferences		Positive risk preferences		Negative risk preferences		Patience2		Patience1	
ITT	0.033	0.030	-0.021	0.000	0.052	0.016	-0.012	-0.006	0.027	0.001
	(0.025)	(0.023)	(0.051)	(0.037)	(0.063)	(0.043)	(0.026)	(0.022)	(0.035)	(0.032)
Mean outcome, Control	-.018	-.018	.014	.014	-.057	-.057	.002	.002	-.018	-.018
R-squared	.003	.220	.004	.404	.009	.336	.003	.145	.008	.212
Observations	6,903	6,903	3,528	3,528	3,528	3,528	6,902	6,902	3,515	3,515
ITT-NR	0.007	0.038	-0.070	-0.062	0.074	0.029	-0.041	-0.045	0.039	0.017
	(0.040)	(0.039)	(0.074)	(0.060)	(0.073)	(0.063)	(0.042)	(0.040)	(0.054)	(0.053)
LATE	0.015	0.078	-0.144	-0.127	0.152	0.059	-0.084	-0.092	0.079	0.035
	(0.082)	(0.079)	(0.153)	(0.123)	(0.148)	(0.129)	(0.087)	(0.081)	(0.110)	(0.109)
Mean outcome, Control	-.019	-.019	-.308	-.308	-.212	-.212	-.103	-.103	-.176	-.176
R-squared	.013	.190	.009	.341	.011	.332	.010	.127	.008	.160
Observations	2,492	2,492	1,248	1,248	1,248	1,248	2,491	2,491	1,242	1,242
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Mindset, Positive and Negative risk preference and Patience1 are only measured in round 3. Standard errors in parentheses are clustered at the school level.*

Table A.10: Effects of the camp on the ERA

	December			
	High School		Vocational training	
ITT	-0.002	0.013	0.000	0.020
	(0.018)	(0.011)	(0.022)	(0.015)
Mean outcome, Control	.760	.760	.790	.790
R-squared	.011	.515	.008	.416
Observations	3,806	3,806	2,677	2,677
ITT-NR	-0.008	0.002	-0.012	0.036
	(0.030)	(0.028)	(0.034)	(0.028)
LATE	-0.017	0.004	-0.024	0.074
	(0.061)	(0.058)	(0.069)	(0.058)
Mean outcome, Control	.383	.383	.613	.613
R-squared	.010	.313	.008	.322
Observations	1,247	1,247	1,364	1,364
Strata	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Outcome only exists for round 1. Pupils are not able to apply separately for 2 and 3 years High School in 2017/2018 and the June assessment is introduced in 2018/2019. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.11: Effects of the camp on the personal skills assessed in the ERA

	December			
	High School		Vocational training	
ITT	0.017	0.017	0.001	0.021
	(0.014)	(0.012)	(0.023)	(0.016)
Mean outcome, Control	.861	.861	.813	.813
R-squared	.010	.463	.011	.401
Observations	3,710	3,710	2,606	2,606
ITT-NR	0.061*	0.043	-0.010	0.034
	(0.034)	(0.029)	(0.038)	(0.029)
LATE	0.124*	0.088	-0.019	0.069
	(0.069)	(0.060)	(0.077)	(0.059)
Mean outcome, Control	.607	.607	.655	.655
R-squared	.032	.397	.016	.329
Observations	1,197	1,197	1,337	1,337
Strata	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Outcome only exists for round 1. Pupils are not able to apply separately for 2 and 3 years High School in 2017/2018 and the June assessment is introduced in 2018/2019. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.12: Effects of the camp on the social skills assessed in the ERA

	December			
	High School		Vocational training	
ITT	0.005	0.001	-0.007	0.009
	(0.012)	(0.009)	(0.021)	(0.013)
Mean outcome, Control	.913	.913	.879	.879
R-squared	.008	.461	.015	.418
Observations	3,710	3,710	2,606	2,606
ITT-NR	0.024	-0.002	-0.017	0.010
	(0.032)	(0.024)	(0.036)	(0.027)
LATE	0.050	-0.005	-0.034	0.020
	(0.065)	(0.049)	(0.073)	(0.054)
Mean outcome, Control	.748	.748	.773	.773
R-squared	.033	.441	.024	.351
Observations	1,197	1,197	1,337	1,337
Strata	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Outcome only exists for round 1. Pupils are not able to apply separately for 2 and 3 years High School in 2017/2018 and the June assessment is introduced in 2018/2019. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.13: Effects of the camp on the preferred education choice in the ERA

	High School		December Vocational training		10th grade	
ITT	0.001	0.004	-0.027	-0.026	0.004	-0.001
	(0.023)	(0.014)	(0.042)	(0.021)	(0.005)	(0.005)
Mean outcome, Control	.785	.785	.572	.572	.009	.009
R-squared	.010	.493	.015	.515	.006	.200
Observations	4,874	4,874	4,874	4,874	4,874	4,874
ITT-NR	0.002	0.018	-0.044	-0.032	0.002	-0.008
	(0.034)	(0.026)	(0.036)	(0.023)	(0.008)	(0.009)
LATE	0.004	0.037	-0.090	-0.066	0.004	-0.016
	(0.070)	(0.054)	(0.074)	(0.047)	(0.016)	(0.018)
Mean outcome, Control	.634	.634	.723	.723	.023	.023
R-squared	.027	.425	.035	.429	.009	.225
Observations	1,977	1,977	1,977	1,977	1,977	1,977
Strata	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Outcome only exists for round 1. Pupils are not able to apply for 2-year High School in 2017/2018 and the June assessment is introduced in 2018/2019. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.14: Effects of *Camp+* vs. *Camp* on the ERA

	December						June					
	3-year High School		Vocational training		2-year High School		3-year High School		Vocational training		2-year High School	
ATT	-0.032	-0.057*	-0.016	-0.026	-0.053	-0.137***	-0.119***	-0.149***	0.008	-0.017	-0.109*	-0.084
	(0.041)	(0.033)	(0.038)	(0.031)	(0.066)	(0.050)	(0.040)	(0.033)	(0.037)	(0.035)	(0.056)	(0.067)
Mean outcome, Camp	.363	.363	.602	.602	.473	.473	.488	.488	.661	.661	.582	.582
R-squared	.041	.317	.020	.270	.077	.470	.062	.340	.017	.293	.101	.455
Observations	633	633	978	978	300	300	614	614	992	992	299	299
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Baseline treatment category is Camp. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.15: Effects of *Camp+* vs. *Camp* on the personal skills assessed in the ERA

	December						June					
	3-year High School		Vocational training		2-year High School		3-year High School		Vocational training		2-year High School	
ATT	-0.068	-0.091**	-0.029	-0.045	-0.142**	-0.151**	-0.087**	-0.102***	-0.002	-0.018	-0.135**	-0.105
	(0.042)	(0.035)	(0.037)	(0.032)	(0.068)	(0.063)	(0.038)	(0.033)	(0.035)	(0.032)	(0.052)	(0.070)
Mean outcome, Camp	.619	.619	.673	.673	.652	.652	.685	.685	.715	.715	.707	.707
R-squared	.023	.293	.018	.265	.079	.314	.031	.334	.016	.250	.120	.401
Observations	609	609	956	956	265	265	608	608	986	986	286	286
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Baseline treatment category is Camp. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.16: Effects of *Camp+* vs. *Camp* on the preferred education choice in the ERA

	December							
	3-year High School		Vocational training		10th grade		2-year High School	
ATT	-0.038	-0.009	0.022	0.009	-0.001	0.000	-0.018	-0.006
	(0.042)	(0.025)	(0.033)	(0.019)	(0.003)	(0.004)	(0.035)	(0.024)
Mean outcome, Camp	.497	.497	.732	.732	.006	.006	.238	.238
R-squared	.024	.455	.030	.476	.009	.062	.029	.328
Observations	1,321	1,321	1,321	1,321	1,321	1,321	1,321	1,321
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes	No	Yes
	June							
	3-year High School		Vocational training		10th grade		2-year High School	
ATT	-0.028	-0.008	0.056*	0.038*	-0.004	-0.002	0.014	0.028
	(0.041)	(0.026)	(0.033)	(0.022)	(0.005)	(0.006)	(0.037)	(0.029)
Mean outcome, Camp	.479	.479	.728	.728	.010	.010	.221	.221
R-squared	.025	.443	.021	.422	.014	.054	.034	.290
Observations	1,321	1,321	1,321	1,321	1,321	1,321	1,321	1,321
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Baseline treatment category is Camp. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.17a: Effects of *Camp+* vs. *Camp* on non-cognitive skills, measured by psychological scales

	Grit		Self-control		Self-concept		Mindset		Total difficult score	
ATT	-0.008	-0.030	-0.035	-0.040	-0.028	-0.044	-0.027	-0.038	0.066	0.048
	(0.059)	(0.045)	(0.076)	(0.053)	(0.066)	(0.047)	(0.059)	(0.047)	(0.064)	(0.046)
Mean outcome, Camp	-.474	-.474	-.198	-.198	-.331	-.331	-.290	-.290	.396	.396
R-squared	.014	.421	.015	.403	.020	.476	.011	.222	.022	.449
Observations	1,159	1,159	1,156	1,156	1,150	1,150	1,150	1,150	1,137	1,137
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Baseline treatment category is Camp. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.17b: Effects of *Camp+* vs. *Camp* on non-cognitive skills, measured by psychological scales - con'd

	General risk preferences		Positive risk preferences		Negative risk preferences		Patience2		Patience1	
ATT	0.067	0.073	0.086	0.109**	-0.145*	-0.150**	0.165***	0.191***	0.008	0.001
	(0.062)	(0.050)	(0.069)	(0.050)	(0.087)	(0.064)	(0.060)	(0.055)	(0.057)	(0.056)
Mean outcome, Camp	-.051	-.051	-.419	-.419	-.163	-.163	-.228	-.228	-.163	-.163
R-squared	.010	.248	.011	.396	.020	.298	.017	.182	.005	.162
Observations	1,148	1,148	1,150	1,150	1,150	1,150	1,146	1,146	1,144	1,144
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Baseline treatment category is Camp. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.18: Effects of camp by gender: Performance in the national Danish test in grade 8

	Language Comprehension			Decoding			Text Comprehension			Overall		
	All	Boy	Girl	All	Boy	Girl	All	Boy	Girl	All	Boy	Girl
ITT	-0.033	-0.046	-0.025	0.017	0.035	0.003	0.065***	0.094***	0.038	0.017	0.033	0.006
	(0.043)	(0.044)	(0.050)	(0.024)	(0.027)	(0.029)	(0.021)	(0.025)	(0.029)	(0.028)	(0.031)	(0.031)
Mean outcome, Control	.024	.044	.003	.000	-.095	.100	-.057	-.158	.048	-.013	-.086	.062
R-squared	.186	.211	.190	.510	.512	.509	.520	.493	.555	.549	.538	.568
Observations	8,953	4,594	4,359	8,953	4,594	4,359	8,953	4,594	4,359	8,953	4,594	4,359
ITT-NR	-0.046	-0.070	-0.023	0.014	0.028	-0.005	0.090***	0.084**	0.103**	0.024	0.022	0.037
	(0.045)	(0.054)	(0.063)	(0.028)	(0.034)	(0.045)	(0.029)	(0.037)	(0.039)	(0.033)	(0.041)	(0.040)
LATE	-0.094	-0.142	-0.046	0.028	0.057	-0.010	0.184***	0.172**	0.210**	0.050	0.046	0.075
	(0.092)	(0.110)	(0.129)	(0.058)	(0.070)	(0.092)	(0.059)	(0.077)	(0.081)	(0.067)	(0.084)	(0.082)
Mean outcome, Control	-.233	-.183	-.305	-.459	-.483	-.425	-.574	-.585	-.559	-.519	-.512	-.528
R-squared	.213	.246	.210	.431	.454	.427	.342	.347	.381	.442	.461	.460
Observations	3,418	2,028	1,390	3,418	2,028	1,390	3,418	2,028	1,390	3,418	2,028	1,390
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.19: Effects of camp by gender: Performance in the national math test in grade 8

	Numbers and Algebra			Geometry			Statistics and Probability			Overall		
	All	Boy	Girl	All	Boy	Girl	All	Boy	Girl	All	Boy	Girl
ITT	0.068**	0.037	0.089**	0.043	0.045	0.038	0.063	0.042	0.090**	0.059*	0.045	0.076*
	(0.032)	(0.036)	(0.040)	(0.035)	(0.043)	(0.040)	(0.039)	(0.047)	(0.042)	(0.034)	(0.040)	(0.039)
Mean outcome, Control	-.064	-.043	-.086	-.042	-.020	-.065	-.033	-.039	-.028	-.050	-.036	-.064
R-squared	.640	.669	.614	.607	.629	.587	.601	.620	.587	.706	.729	.691
Observations	4,283	2,197	2,086	4,283	2,197	2,086	4,283	2,197	2,086	4,283	2,197	2,086
ITT-NR	0.094**	0.087*	0.078	0.091*	0.109**	0.067	0.117*	0.087	0.173***	0.100**	0.106*	0.123**
	(0.047)	(0.051)	(0.065)	(0.046)	(0.054)	(0.057)	(0.062)	(0.068)	(0.064)	(0.050)	(0.057)	(0.060)
LATE	0.192**	0.178*	0.160	0.185*	0.223**	0.138	0.240*	0.178	0.354***	0.204**	0.216*	0.251**
	(0.096)	(0.104)	(0.133)	(0.093)	(0.111)	(0.118)	(0.128)	(0.138)	(0.130)	(0.102)	(0.117)	(0.122)
Mean outcome, Control	-.645	-.596	-.719	-.625	-.569	-.708	-.619	-.585	-.672	-.678	-.628	-.754
R-squared	.452	.496	.410	.452	.517	.357	.460	.485	.472	.550	.586	.479
Observations	1,581	952	629	1,581	952	629	1,581	952	629	1,581	952	629
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Notes:***p<0.01; **p<0.05; *p<0.1. Outcome for math only exists for round 3. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.20: Effects of camp by SES and academic performance: Performance in the national Danish test in grade 8

	Language Comprehension			Decoding			Text Comprehension			Overall		
	All	Low SES	Low perf.	All	Low SES	Low perf.	All	Low SES	Low perf.	All	Low SES	Low perf.
ITT	-0.033	-0.077	-0.080	0.017	0.006	-0.037	0.065***	0.132***	0.100**	0.017	0.016	-0.002
	(0.043)	(0.066)	(0.057)	(0.024)	(0.036)	(0.041)	(0.021)	(0.035)	(0.045)	(0.028)	(0.043)	(0.045)
Mean outcome, Control	.024	-.044	-.459	.000	-.169	-.720	-.057	-.354	-.840	-.013	-.232	-.827
R-squared	.186	.241	.228	.510	.487	.370	.520	.480	.253	.549	.522	.372
Observations	8,953	1,861	1,664	8,953	1,861	1,664	8,953	1,861	1,664	8,953	1,861	1,664
ITT-NR	-0.046	-0.061	-0.080	0.014	0.034	-0.037	0.090***	0.133**	0.100**	0.024	0.056	-0.002
	(0.045)	(0.077)	(0.057)	(0.028)	(0.048)	(0.041)	(0.029)	(0.052)	(0.045)	(0.033)	(0.054)	(0.045)
LATE	-0.094	-0.124	-0.164	0.028	0.071	-0.076	0.184***	0.273**	0.205**	0.050	0.114	-0.004
	(0.092)	(0.158)	(0.116)	(0.058)	(0.098)	(0.084)	(0.059)	(0.106)	(0.093)	(0.067)	(0.111)	(0.092)
Mean outcome, Control	-.233	-.286	-.459	-.459	-.544	-.720	-.574	-.697	-.840	-.519	-.625	-.827
R-squared	.213	.286	.228	.431	.460	.370	.342	.386	.253	.442	.471	.372
Observations	3,418	978	1,664	3,418	978	1,664	3,418	978	1,664	3,418	978	1,664
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.21: Effects of camp by SES and academic performance: Performance in the national math test in grade 8

	Numbers and Algebra			Geometry			Statistics and Probability			Overall		
	All	Low SES	Low perf.	All	Low SES	Low perf.	All	Low SES	Low perf.	All	Low SES	Low perf.
ITT	0.068**	0.058	0.125**	0.043	0.061	0.135**	0.063	0.094	0.154*	0.059*	0.081*	0.151**
	(0.032)	(0.047)	(0.055)	(0.035)	(0.047)	(0.059)	(0.039)	(0.060)	(0.082)	(0.034)	(0.048)	(0.064)
Mean outcome, Control	-.064	-.313	-.917	-.042	-.342	-.910	-.033	-.352	-.927	-.050	-.362	-.988
R-squared	.640	.572	.383	.607	.603	.364	.601	.616	.430	.706	.676	.466
Observations	4,283	787	818	4,283	787	818	4,283	787	818	4,283	787	818
ITT-NR	0.094**	0.136	0.125**	0.091*	0.195***	0.135**	0.117*	0.230**	0.154*	0.100**	0.197**	0.151**
	(0.047)	(0.082)	(0.055)	(0.046)	(0.066)	(0.059)	(0.062)	(0.088)	(0.082)	(0.050)	(0.077)	(0.064)
LATE	0.192**	0.278	0.256**	0.185*	0.399***	0.277**	0.240*	0.471**	0.316*	0.204**	0.403**	0.309**
	(0.096)	(0.168)	(0.113)	(0.093)	(0.136)	(0.121)	(0.128)	(0.180)	(0.167)	(0.102)	(0.158)	(0.130)
Mean outcome, Control	-.645	-.734	-.917	-.625	-.756	-.910	-.619	-.819	-.927	-.678	-.829	-.988
R-squared	.452	.471	.383	.452	.508	.364	.460	.583	.430	.550	.591	.466
Observations	1,581	399	818	1,581	399	818	1,581	399	818	1,581	399	818
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Notes:***p<0.01; **p<0.05; *p<0.1. Outcome for math only exists for round 3. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.22: Effects of *Camp+* vs. *Camp* by gender: Performance in the national Danish test in grade 8

	Language Comprehension			Decoding			Text Comprehension			Overall		
	All	Boy	Girl	All	Boy	Girl	All	Boy	Girl	All	Boy	Girl
ATT	0.081	0.120	-0.068	0.046	0.073	-0.060	0.071	0.012	0.071	0.078	0.088	-0.022
	(0.076)	(0.089)	(0.101)	(0.047)	(0.062)	(0.062)	(0.048)	(0.056)	(0.060)	(0.051)	(0.066)	(0.067)
Mean outcome, Camp	-.373	-.407	-.311	-.622	-.679	-.518	-.704	-.753	-.615	-.700	-.758	-.595
R-squared	.211	.253	.325	.426	.435	.449	.376	.450	.365	.432	.459	.498
Observations	1,316	851	465	1,316	851	465	1,316	851	465	1,316	851	465
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Baseline treatment category is Camp. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.23: Effects of *Camp+* vs. *Camp* by gender: Performance in the national math test in grade 8

	Numbers and Algebra			Geometry			Statistics and Probability			Overall		
	All	Boy	Girl	All	Boy	Girl	All	Boy	Girl	All	Boy	Girl
ATT	-0.030	-0.083	0.012	-0.040	-0.062	-0.046	-0.021	-0.080	0.050	-0.037	-0.087*	-0.003
	(0.042)	(0.053)	(0.062)	(0.047)	(0.053)	(0.066)	(0.047)	(0.052)	(0.079)	(0.046)	(0.052)	(0.059)
Mean outcome, Camp	-.710	-.619	-.876	-.707	-.603	-.897	-.669	-.593	-.806	-.751	-.653	-.928
R-squared	.437	.451	.501	.414	.409	.539	.428	.472	.432	.512	.512	.595
Observations	1,321	856	465	1,321	856	465	1,321	856	465	1,321	856	465
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Baseline treatment category is Camp. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.24: Effects of *Camp+* vs. *Camp* by SES and academic performance: Performance in the national Danish test in grade 8

	Language Comprehension			Decoding			Text Comprehension			Overall		
	All	Low SES	Low perf.	All	Low SES	Low perf.	All	Low SES	Low perf.	All	Low SES	Low perf.
ATT	0.081	0.031	-0.013	0.046	-0.065	0.038	0.071	0.028	0.081	0.078	-0.028	0.039
	(0.076)	(0.125)	(0.084)	(0.047)	(0.076)	(0.049)	(0.048)	(0.080)	(0.067)	(0.051)	(0.074)	(0.064)
Mean outcome, Camp	-.373	-.419	-.491	-.622	-.625	-.804	-.704	-.751	-.901	-.700	-.740	-.905
R-squared	.211	.280	.189	.426	.481	.420	.376	.379	.340	.432	.483	.385
Observations	1,316	374	834	1,316	374	834	1,316	374	834	1,316	374	834
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Baseline treatment category is Camp. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

Table A.25: Effects of *Camp+* vs. *Camp* by SES and academic performance: Performance in the national math test in grade 8

	Numbers and Algebra			Geometry			Statistics and Probability			Overall		
	All	Low SES	Low perf.	All	Low SES	Low perf.	All	Low SES	Low perf.	All	Low SES	Low perf.
ATT	-0.030	-0.055	-0.123**	-0.040	-0.127*	-0.135**	-0.021	-0.043	-0.090	-0.037	-0.079	-0.130**
	(0.042)	(0.065)	(0.050)	(0.047)	(0.065)	(0.054)	(0.047)	(0.071)	(0.054)	(0.046)	(0.060)	(0.051)
Mean outcome, Camp	-.710	-.841	-.889	-.707	-.836	-.835	-.669	-.814	-.843	-.751	-.896	-.923
R-squared	.437	.504	.349	.414	.500	.332	.428	.459	.356	.512	.597	.408
Observations	1,321	378	836	1,321	378	836	1,321	378	836	1,321	378	836
Strata	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Notes: ***p<0.01; **p<0.05; *p<0.1. Baseline treatment category is Camp. The covariates is imputed with the value zero and a missing-indicator equal to one is added to the conditioning set if data on the covariates is missing. Standard errors in parentheses are clustered at the school level.*

CHAPTER 2

BOYS LEFT BEHIND: THE EFFECTS OF SUMMER CAMP AND FOLLOW-UP STRATEGIES ON ACADEMIC, PERSONAL, AND SOCIAL COMPETENCIES

Søren Albeck Nielsen
Aarhus University and TrygFonden's Centre for Child Research

Abstract

I use Danish administrative data to investigate the consequences of summer camp participation for disadvantaged boys on academic, personal, and social competencies. My identification strategy relies on individual level panel data that enables me to observe outcomes before and after summer camp participation. Using a difference-in-differences strategy, I find overall positive effects on academic and personal competencies that reduce the gap to a matched group of boys with similar background characteristics by 40 to 80 percent. Further, I exploit a structural change in the follow-up program to evaluate how different mentoring strategies affect outcomes. In 2017, the follow-up program was changed from individual mentoring to group mentoring, which allows me to compare camp effects before and after the structural change. Using a triple differences strategy, I find that group mentoring in the follow-up program dramatically improves personal and social competencies, indicating that the camp itself affects academic abilities whereas the format of the follow-up program is crucial for effects on personal and social competencies.

Keywords: Intensive learning camps, Follow-up, Mentoring, Non-cognitive competencies, Difference-in-differences

JEL Codes: I21, I24, I28, C23

2.1 Introduction

This paper studies the consequences of summer camp participation and the structure of the follow-up program for disadvantaged boys in lower secondary education. During the last decades, there has been a rapid and substantial reversal of the gender gap in educational attainment in much of the developed world (Murnane, 2013). The OECD (2013) report shows that, for 28 of 34 OECD counties, females have surpassed males in higher education among adults aged 25 to 34. Autor et al. (2019) conclude that boys in disadvantaged household have higher rates of disciplinary problems, lower academic performance, and lower rates of high school completion, than girls with similar backgrounds. Similarly, boys from low socioeconomic background have been hit the hardest by the disruption to education caused by the Covid-19 pandemic (Di Pietro et al., 2020). Thus, there is a need for effective remedial education programs that can eliminate this learning gap for disadvantaged boys.

Intensive learning programs are widely used as a policy tool to increase educational attainment and often implemented using summer camps (Kim and Quinn, 2013; Lauer et al., 2006; Cooper et al., 2000). The three meta-analyses report positive short-run effects on math and reading, but with very small effect sizes. Despite the overwhelming use of summer camps in the society, there is little causal evidence. Only a handful of studies rely on experiments or quasi-experiments and often with low statistical power. The most convincing evidence on summer camps exploit regression discontinuity designs (henceforth RDD) and standardized tests. Mariano and Martorell (2013) find modest effects on language and little effect on math for low performing pupils. The RDD of Jacob and Lefgren (2004) finds positive effects on both reading and math but only for grade 3 pupils and not grade 6 pupils. Recent evidence from Battistin and Schizzerotto (2019) finds negative effects on academic performance of mandatory summer camp for at-risk pupils in Italy. Overall, there is weak evidence on summer camps effect for pupils in the transition from lower to upper secondary education and no evidence on summer camps effect on non-academic outcomes in general.

I use population-level Danish register data covering all grade 8 boys in the period 2015-2019 to study the effect of a two-week summer camp for academically disadvantaged boys with a one-year follow-up mentoring program. In particular, I exploit grade 8 and 9 individual level panel data to implement a difference-in-differences strategy. I ask, first, what are the effects of participating in a summer camp on academic performance and readiness[1] for upper secondary education? Secondly, what are the consequences of changing the follow-up program from individual mentoring with an adult to a group-mentoring program of 10 to 12 boys?

The analyses deliver a set of noteworthy answers to the research questions: First, I show that participating in a two-week summer camp during the holiday between

[1] Indicator variable taking the value one if ready and zero otherwise. See section 2.3 for the description of the readiness assessment.

grades 8 and 9 has positive impacts on academic performance and readiness for upper secondary education. I observe an increase in the average assessment mark of around 15% of a standard deviation, suggesting that overall academic abilities have improved on average. Additionally, participation in the summer camp increases readiness for upper secondary education by 18 percentage points at the educational readiness assessment in December grade 9. The overall readiness assessment consists of three sub-categories, in which the pupils must qualify in order to be ready for upper secondary education. Interestingly, academic readiness is the main reason for the overall effect with an increase of 22 percentage points, whereas the summer camp increases the personal readiness assessment by 10 percentage points, and the effect on social readiness is not statistically significant. Critically, this estimation strategy combined with the introduction of a group-mentoring follow-up program in 2017 allows me to identify the effect of moving from individual mentoring with one adult to a group-mentoring follow-up program. Using a triple differences strategy, I find that substituting individual mentoring to group mentoring increases both the personal and social readiness assessment by 15 and 16 percentage points and has no impact on academic readiness. This indicates that participating boys benefit both personal and socially from staying connected with equal peers and sharing experiences and solutions. Intuitively, it makes sense that mentoring these boys together should benefit their behavior but one might also fear that connecting challenged boys on a regular basic might increase poor behavior. Petrosino et al. (2013) show exactly this. They find that crime preventive camps for children at risk of becoming delinquent do not work and in fact increases poor behavior.

Thus, my analysis shows that at-risk boys improve their academic, personal, and social competencies from participating in a two-week intensive learning camp during the summer holiday between grade 8 and 9. Importantly, I find that the improvement in non-academic competencies is a result of shifting the follow-up program from individual mentoring to group mentoring. This finding is particular important, as it indicates that learning programs aimed at at-risk boys should focus on group mentoring in the follow-up program to boost the effects of the program.

I contribute to the literature on consequences of intensive learning camps. First, often the effect is measured shortly after the camp is completed. In this study, however, I exploit administrative data to investigate the effect up to one year after camp participation. Secondly, the educational readiness assessment allows me to supplement the estimated effects on academic outcomes with effects on measures of personal and social competencies directly related to future educational attainment. These outcomes are under-investigated in the previous literature, which is problematic when considering the importance of non-academic skills for future achievements (Almlund et al., 2011). Finally, the change in the follow-up strategy allows me to add novel evidence on how follow-up programs should be structured with the attempt to maximize the effect on all competencies that are considered relevant for future educational outcomes - and not focus exclusively on the academic competencies.

The paper is organized as follows: Section 2 provides background information on the structure of the summer camp. Section 3 describes the data and descriptive statistics and section 4 explains the empirical strategy. In section 5, I show the effect of the summer camp, section 6 investigates the effect of changing follow-up strategy, and finally, section 7 concludes.

2.2 The two weeks summer camp

The summer camp is a two-week intensive learning camp that takes place during the summer holiday between grades 8 and 9. It targets boys who are at risk of not becoming ready for upper secondary education at the end of compulsory school and, thus, not able to choose their desired education. This group comprises boys with a substantial academic backlog, low self-confidence, and low motivation for going to school. The main objective of the summer camp is to help these academically weak pupils to catch up such that they become prepared to continue in the education system. In order to achieve this object, the program focuses on academic abilities by training reading, writing, spelling, and math, but it also aims to develop socio emotional skills.

2.2.1 Content and organization

The recruitment into the summer camp occur by the boy in collaboration with their parent apply for the camp independently from their local school. The boys can apply from January and application deadline is medio March. The boys are notified of acceptance within a month from the deadline where all application is evaluated simultaneously. If the program is over-subscribed the boys are selected based on academic, personal and social problems.

The summer camp consists of an intro camp and a main camp. Approximately two weeks prior to the start of the main camp, pupils attend the intro camp during a weekend. The purpose of the intro camp is to introduce the teachers, structure, settings, and rules at the main camp as well as creating relationships to their peers in order to start the social connections and decrease some of the anxiety before the main camp. Additionally, there are individual conversations between the teachers and the boys for the teachers to get a greater insight into the boys' difficulties and potential. Finally, there is academic tests of abilities in order to organize the main camp learning plan according to the pupils' individual levels and needs (Andersen et al., 2019). The total cost for the summer camp, including the follow-up program, is approximately 5,000 USD per pupil.

The main camp takes place in the first two weeks of the summer holidays. The boys are divided into teams of approximately 15 pupils, where two teams receive instruction together. The two-week camp consists of a full schedule from morning

to evening with morning wake-up call at 6.30 AM and bedtime at 10.00 PM. This ensures a fixed structure that is repeated every day. The day consists of four 90-minute modules, which include the subjects Danish, mathematics, as well as modules focusing on socio emotional skills and plenty of physical activity. See Figure A.1 and A.2 for the schedules for the first and second week of the main camp. The camp is located at a boarding school with teaching facilities, kitchen, and sleeping rooms. Thus, there is no need to leave the camp during the two weeks. The teachers and peers are different from the participant's local learning environment, mobile phones are only allowed one hour per day, and the camp has a zero-sugar policy. Importantly, the boys do not miss any teaching at their home school, because the intro camp is held during a weekend and the main camp during the summer holiday.

The pedagogical aim of the summer camp is to offer boys teaching methods that, to a greater extent than at their home-school, are tailored the boys' individual needs and strengths. The intervention builds on inspiration from the literature on Visible Learning and Self-Determination Theory (Hattie and Yates, 2013; Ryan and Deci, 2000). The teaching itself is highly structured around Flipped classrooms, where blackboard teaching is replaced by e.g. short videos. The boys watch the videos individually to free up the teacher's time for more student-activating teaching. An additional advantage with this method is that this form of teaching also means that the boys can revisit the material after the camp. The teachers at the camp attempt to turn around the negative school experiences by testing the boys midway through the camp and at the final day. Thus, they visually illustrate the boys' academic progress to recreate a positive self-narrative as well as the motivation to learn. A clear pedagogical tool of the learning camp is to believe in the boys by praise and acknowledge them for their progress and to support them during the difficult periods.

2.2.2 The follow-up strategy

Receipt of educational activities in a new environment, away from the home-school peers and teachers, is a key element of the summer camp. The intention is to break bad habits and make new social connections, exploiting that all boys are on common ground by not knowing each other beforehand. This is an obvious strength of the learning program but at the same time, it also constitutes one of the biggest challenges. Andersen et al. (2019) show that many pupils find it difficult to maintain the good working habits, the positive academic development, and the joy of learning when they return to their ordinary classroom after the camp. The summer camp is perceived as being detached from everyday school life and, thus, returning to the local environment risks a setback when the boys discover that nothing has changed in how their peers and local teachers perceive them. Therefore, the summer camp has a great focus on follow-up after the camp. The primary follow-up program is a mandatory one-year mentoring scheme with a built-in parenting effort. Additionally, the camp teachers construct an "Exit package" for all boys, which is a written handover to the

boys' local schoolteachers. However, only 50% of the local teachers have heard about the exit package and fewer have used it (Rambøll, 2019). There is no additional formal contact between the camp teachers and the local teachers.

In 2015 and 2016, the mentoring scheme was an individual mentor-mentee program, where the boy and the adult mentor can train academic as well as non-academic skills during meetings twice a month. Additionally, the mentors can help the boys set new learning goals and maintain their positive development. In 2017, the mentoring scheme was changed to mentor groups of 10-12 boys, who meet twice a month, outside regular school hours, at mentor centers geographically spread across Denmark. At the centers, the boys receive academic and personal counseling and guidance from adults who have attended the summer camp and volunteering role models. Furthermore, the content of the summer is repeated at each meeting, they share experiences related to returning to the local learning environment, and they receive homework help and educational guidance.

2.3 Data and descriptive statistics

To investigate the effect of the summer camp and how changes in the follow-up strategies affect the outcomes of participants, I leverage Danish administrative register data available through Statistics Denmark covering the full population of pupils in the Danish school system. Focal to this study is the Danish Student Register comprising all educational choices in Denmark. This register is a unique longitudinal dataset that allows me to follow schooling information such as private vs public schooling, school and classroom movements, and special needs teaching from 2008/2009 to 2019/2020. Crucially, this data is informative about what grade a boy attends, enabling me to observe the boys' academic development though their average assessment mark and their readiness assessment for upper secondary education at grades 8 and 9 obtained from the Ministry of Education. I augment this data with information on dyslexia, psychiatric diagnoses, results from national tests in reading and math, school absence, personality traits, and school well-being. Furthermore, I exploit socio economic information describing demographics, employment, income and educational level of the parents. The sample consist of 158,231 boys of which 241 participated at the summer camp between 2015 and 2019[2]. I study two groups of outcomes that both characterize key information regarding future life-trajectory. Specifically, I include 1) the average assessment mark and 2) the readiness assessment for upper secondary education.

[2]The 241 boys are those who applied and showed up at the summer camp. The dropout rate is approximately 8%.

2.3.1 average assessment mark

The average assessment mark is an expression of the pupil's academic level averaged across all subjects the pupil attends.[3] The teachers grade the pupil in relation to the academic goals that apply for the subject. Thus, teachers base their assessment on the extent to which the pupil fulfills the subject's competence, skills and knowledge goals. Importantly, the pupil's work effort and/or behavior in the classroom is, generally, not included in the grading. Only if the subject has goals that include these abilities, are they included in the grading. The grading occur in December and June of grades 8 and 9.[4]

To measure the pupils' overall abilities, I standardize each continuous assessment mark within each subject and timing of the assessment to mean zero and standard deviation of one. Then, I calculate the average assessment mark and standardize within the timing of the assessment to mean zero and standard deviation of one. The latter standardization allows us to readily interpret regression coefficients in standard deviations units and, thus, render results comparable to effect sizes of other studies. The camp boys are on average 1.2 standard deviations below the non-camp boys at both grade 8 assessments, indicating that camp boys have a significant academic backlog.

2.3.2 Readiness assessment for upper secondary education

Assessing pupils' readiness in relation to choosing and completing upper secondary education is a process that starts in grade 8. The purpose of the assessment is to ensure that non-ready pupils receive school interventions and individual guidance in the process towards the end of grade 9 in order for them to make the best secondary education choice and be prepared for the education chosen. The readiness assessment includes all pupils in public and private schooling and takes place in December. The pupil's primary teachers assess the academic, personal, and social competencies of the pupil. All three criteria must be met for the pupils to be assessed as ready for upper secondary education. On average, 62% of the boys are assessed to be ready for upper secondary education, but for the camp participants in grade 8 only 12% are assessed to be ready for upper secondary education. This 50%-points difference clearly indicates that the camp participants need additional assistance in order to continue in the educational system.

[3] In Danish, there are four grades (Reading, Spelling, Written, and Oral). In math, there are three grades (Math with aids, Math without aids, and Oral). In foreign languages there are two grades (Oral and Written). For the following mandatory subjects there are only one grade (Physics/chemistry, Biology, Geography, History, Social studies, Religion, and Gym). Finally, the pupils are graded in one of the following four electives (Crafts and design, Food knowledge, Music, and Art).

[4] The June grade 8 grading is initiated in 2018 and do not exists for the previous years.

In order to be academically ready in grade 8, the pupil must have an average assessment mark of at least 4.0[5] for going to vocational training and two-year high school and 5.0 for a three-year high school. In grade 9, the requirements for vocational training drops to an average of 2.0 in Danish and math. In grade 8, 75% are academically ready for their desired educational choice, whereas this fraction is only 19% for the camp participants.

For the personal readiness assessment, the teachers assess whether the pupil has the necessary personal competencies to begin upper secondary education after grade 9. The teachers have five key areas to guide them through the assessment: 1) Motivation, 2) Independence, 3) Responsibility, 4) Meeting stability, and 5) Choice readiness. Motivation is about having a desire for education and learning, and contributing actively in teaching. Independence is about being able to act on your own, i.e. the pupil is able to take initiatives and to ask for help when needed. Responsibility is whether the pupil shows up prepared for classes and whether the pupil is able to keep appointments that he or she makes with teachers or fellow pupils. Meeting stability is whether the pupil attends school every day and does so timely. Finally, choice readiness is about whether the pupil can make decisions and whether the pupil is able to make a positive and active choice in the educational selection process. Often a pupil will not be able to make the educational choice in grade 8, but they must be able to reflect on it. In grade 8, 74% of all boys are assessed personally ready, while the fraction is 40% for the camp participants. The camp boys are thus still behind the average boy but the difference is not as great as for the academic readiness assessment.

The social readiness assessment focuses on whether the pupil has the social prerequisites needed to be able to start and complete upper secondary education. In order to assess this, the teachers have three focus points to guide them: 1) Collaboration ability, 2) Respect, and 3) Tolerance. Collaboration ability is about being able to solve tasks together with others, to keep common agreements and contribute positively to the community. Respect is about the pupil being able to show consideration for other pupils and teachers. Tolerance is the ability or willingness to accept what is unknown. Thus, tolerance is about being able to understand and accept other people's opinions, behavior, culture, religion, etc. On average, 82% of all boys are assessed to be socially ready, while 56% of the camp boys are socially ready.

Hence, clearly, the large difference in the overall readiness assessment is driven by academic performance. A large fraction of the camp participants does also have personal and social problems but not in the same magnitude as the academic backlog.

[5]Academic grading in Denmark is on a 7-point scale with the following grades from the best to the worst: 12(A), 10(B), 7(C), 4(D), 02(E), 00(FX), -3(F).

2.3.3 Characteristics of camp vs non-camp boys

Leveraging Danish administrative data allows me to characterize in detail the differences between camp and non-camp boys. In column 1 to 5 of table A.1, I present summary statistics for a rich set of covariates, recorded prior to camp participation. The descriptive statistics show significant differences between the two groups. There is an under-representation of boys with non-western ethnicity in the camp group. They are more likely to live in broken families, have additional relocations, and thus an increased number of school changes. The camp boys have received on average half a year more special needs teaching, they are over-represented in terms of psychiatric diagnoses, and 35% of them have been diagnosed with dyslexia, whereas only 10% of the non-camp group are dyslexics. They grow up in households with lower socioeconomic status, with both parents having approximately half a year less of education. Academically, the camp group perform significantly worse in the national test in reading and math across all grades and profile areas with gaps to non-camp boys between 0.5 and 1.0 standard derivations. The camp participants on average rate their school well-being lower, especially when assessing their learning self-efficacy. I observe similar findings across all three measured personality traits, the participants having lower levels with conscientiousness showing the largest difference of 0.5 standard derivation.

2.4 Empirical strategy

2.4.1 Identifying the consequences of summer camp participation

The first goal of this paper is to estimate the consequences of summer camp participation for all camp boys[6] on pupil-level outcomes. The key challenge in any program evaluation is to estimate the counterfactual, in the present case the outcomes in the absence of summer camp participation. A natural worry is that selection bias challenges the identification, i.e. summer camp participating boys comprise a different population compared to the remaining population of boys who do not participate in the summer camp.

I address this concern with a difference-in-differences strategy using individual level panel data similar to an individual fixed effect analysis. This strategy compares the change in participating pupil's outcomes from grade 8 to grade 9 to the similar change for non-participating pupils. This strategy implicitly controls for unobserved time-invariant individual school performance. However, it is likely that school performance develops differently based on the underlying distribution of covariates. To account for this, I combine the difference-in-differences strategy with Entropy Balancing along the lines of Freier et al. (2015). The basic idea is to reduce bias due

[6] I pool together all grade 8 boys in the school years 2014/2015 to 2018/2019 when estimating the overall consequences of summer camp participation.

to different distributions of covariates in the camp vs non-camp groups by matching the non-camp group such that it is identical to the camp group with respect to included characteristics (Heckman et al., 1997; Blundell et al., 2004; Abadie, 2005). Entropy Balancing is a data processing method introduced by Hainmueller (2012) that ensures perfect covariate balance with a binary treatment variable. See Hainmueller and Xu (2013) and Hainmueller (2012) for a detailed description of Entropy Balancing. Columns 6 to 8 in table A.1 show descriptive statistics of the observed covariates for the Entropy Balancing adjusted non-camp group. Column 8 shows that the means in the camp and non-camp groups are perfectly balanced across the full set of covariates.

I start the analyses with the following difference-in-differences equation:

$$y_{it} = u_i + \beta_0 + \beta_1 camp_i + \beta_2 Time_i + \delta_j (Camp_i \cdot Time_i) + \epsilon_{it} \tag{2.1}$$

where y_{it} is the outcome of interest, $camp_i$ is a binary variable indicating summer camp participation, and $Time_i$ is a time-period categorical variable containing two periods before and after camp participation for the average assessment mark and one for the readiness for upper secondary education assessment. δ_j are the effects of summer camp participation by time-periods relatively to December in grade 8 and are the parameters of interest, i.e. the average treatment effect of the treated (ATT henceforth). u_i is individual level fixed effect and ϵ_{it} is the error term. Standard errors are clustered at the individual level as outlined in Bertrand et al. (2004).

The key identifying assumption in the difference-in-differences strategy is that there can be no differential trends between the camp and non-camp groups in the absence of summer camp participation. To investigate the validity of this assumption, I first observe pre-camp trends for the average assessment mark and then test the robustness of my findings to different model specifications, alternative comparison groups, alternative balancing specifications, and by performing placebo analyses prior to camp participation and on a randomly selected, synthetic, "camp" group. Significant and large placebo effects would imply that the identification strategy does not capture systematically different trends in pre-camp school performance between camp and non-camp boys.

2.4.2 Detecting the consequences of transforming the follow-up program

The second goal of this paper is to estimate the impact of transforming the follow-up program - the 2017 structural change in the mentoring scheme - on pupil-level outcomes. Individual mentoring is substituted with group mentoring in the one-year mandatory follow-up program. The effect of this change can be estimated using a triple differences estimator[7]. This is equivalent to the difference between the 2015 to

[7] Individual fixed effect estimates is no longer possible because none of the boys have participated more than once in the summer camp.

2016 and 2017 to 2019 difference-in-differences estimates and is estimated using the following equation:

$$y_i = \beta_0 + \beta_1 Camp_i + \beta_2 Post_i + \beta_3 Group_i + \beta_4 Camp_i \cdot Post_i + \beta_5 Camp_i \cdot Group_i + \beta_6 Post_i \cdot Group_i + \beta_7 Camp_i \cdot Post_i \cdot Group_i + \beta_8 X_i + \epsilon_i \quad (2.2)$$

where y_i is the outcome of interest, $camp_i$ is a binary variable indicating summer camp participation, $Post_i$ is an indicator variable with the value one (zero) after (before) summer camp participation, $Group_i$ takes the value one for boys participating in the group mentoring scheme and zero for boys in the individual mentoring scheme, X_i is a matrix containing covariates measured prior to camp participation, and ϵ_i is the error term. β_7 is the effect of summer camp participation with the group mentoring follow-up program relatively to individual mentoring and is the parameter of interest.

Despite that the triple differences estimator can be computed as the difference between two difference-in-differences estimators, Gruber (1994) states that the identifying assumptions are weaker. Olden and Møen (2020) formally show this by proving that the triple differences estimator does not require two parallel trends assumptions in order to estimate causal effects. It requires only one parallel trends assumption to hold to provide causal interpretation. Thus, the difference between two biased difference-in-differences estimators will not be biased if the bias is the same in both estimators because the bias will be removed with triple differencing.

2.5 Results: Effect of camp participation on school performance outcomes

I start with a graphical analysis of the effects on the average assessment mark measured four times across grade 8 and 9. Here I use camp participation for all the years 2015-2019. Figure 2.1 shows the event study representation of summer camp participation on the average assessment mark with December grade 8 as reference point. The solid line presents the unadjusted event study (raw difference-in-differences estimates) and the dashed line presents the entropy balancing adjusted event study. Critically, there is no difference in effect between December and June in grade 8, consistent with parallel trend prior to summer camp participation. Additionally, the adjusted and unadjusted treatment effects are very similar across time, indicating that the different distribution of covariates does not affect the estimated effect of camp participation. Overall, I find effects of 15% of a standard deviation both 6 and 12 months after summer camp participation, suggesting that the effects on academic grades are persistent up to a year after the intervention. Participating in the two-week summer camp reduces the learning gap to the full population of boys by 14% and by 40% when compared to boys with similar characteristics.

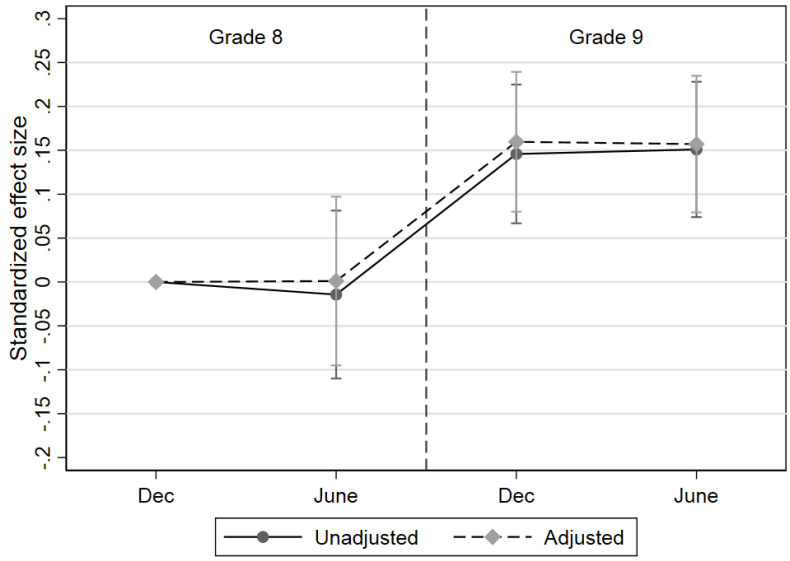
A worry might be that schoolteachers who are grading the boys in the average assessment mark might be influenced by the knowledge of who has participated in the summer camp. Therefore, I exploit data from the grade 9 school leaving exam in June that is graded by external censors to investigate if this changes the results. Figure A.3 shows the Entropy Balancing adjusted event study representation, where the average assessment mark in June of grade 9 is substituted with the exam performance. This worry does not seem to be valid since using exam performance does not change the findings. To investigate this further I use the written exams because these are completely independent of the teachers. Figure A.4 supports the conclusion that the teachers' believes of the camp do not influences the effects.

For simplicity and similarity in presentation with the readiness assessment for upper secondary education, which is only registered once in grade 8 and 9, I supplement my event study model with a simple difference-in-differences model that ignores the time to treatment aspect, i.e. it does not allow effects to vary with time distance to summer camp. I also conduct robustness checks using this simpler difference-in-differences model. This is not critical for my findings, conducting the robustness checks on the main event study model yields similar findings. Table A.2 shows in column 1 to 3 the unadjusted difference-in-differences estimations and in column 4 to 6 the Entropy Balancing adjusted estimations. I present the raw estimations without covariates in Columns 1 and 4, in columns 2 and 5 I include covariates, and in column 3 and 6 I exploit the panel structure of the data and conduct an individual fixed effects analysis. The table shows that the findings are robust to using a gradually richer specification.

Although the estimates are statistically significant it is important in a policy perspective to discuss if they are are economically significant. Kraft (2020) discusses how to interpret effect sizes of educational interventions using 750 randomized trials. He argues that effects of 15% of a standard deviation are of medium size when benchmarked against a large set of educational interventions conducted in social science. However, when taking into consideration the age of the boys in this study, the effect on average assessment mark is around the 70 percentile of the distribution of effect sizes included in Kraft (2020). In a Danish context, this summer camp is also very competitive when compared to other lower secondary school interventions (Rosholm et al., 2021).

Table 2.1 explores the impact of summer camp participation on the readiness assessment for upper secondary education. The results show that participation in the summer camp leads to a considerable increase in readiness for desired upper secondary education. Participating boys increase their readiness by 18 percentage points relatively to non-camp boys. Remarkably, the summer camp reduces the gap to the full population of boys by 35% and by 80% to the entropy balance adjusted boys with similar characteristics. Columns 1 to 3 and 4 to 6 show that the findings are robust to using a gradually richer specification and combination with entropy balancing.

Figure 2.1: Summer camp and average assessment mark



Notes: *This figure shows the main event study representation of the effect of summer camp participation on average assessment mark. The solid line presents the unadjusted event study and the dashed line presents the Entropy Balancing adjusted event study. Each point represents the assessment difference in average assessment mark between camp and non-camp boys with 95% confidence intervals. I use the assessment in December grade 8 as reference time-point.*

This result is consistent with the increase in the average assessment mark, which begs the question, if the increased readiness for upper secondary education is purely driven by an increase in academic abilities. Tables A.3 to A.5 show the effect of summer camp participation on the three criteria used in the readiness assessment. Clearly, the summer camp has the largest effect on academic competencies with an increase of 22 percentage points. For personal competencies the effect is borderline significant and show an increase by 11 percentage points in the teacher assessed personal readiness. Finally, the effect on social competencies is positive, however, not statistically significant. However, as described in section 3 the gap prior to summer camp is also largest for the academic competencies. Thus, there is more room for improvement and maybe a larger focus on academic problems at the summer camp.

An important worry for the educational readiness assessment is that camp boys are able to change their desired upper secondary education between grade 8 and 9, and that this may cause the positive effects. I.e. if the boys change from a 3-year high school to a vocational school, the requirements for being assessed ready are reduced. This does not seem to be a problem, as illustrated in table A.6; there are no effects for

Table 2.1: Effects of summer camp on overall readiness assessment

	Difference-in-difference			DiD & Entropy balance		
	(1)	(2)	(3)	(4)	(5)	(6)
Camp	**-0.504**	**-0.251**		**-0.220**	**-0.221**	
	(0.022)	(0.023)		(0.022)	(0.019)	
Post	**0.147**	**0.139**	**0.139**	**0.206**	**0.190**	**0.193**
	(0.001)	(0.001)	(0.002)	(0.004)	(0.005)	(0.006)
Camp X Post	**0.223**	**0.214**	**0.230**	**0.165**	**0.172**	**0.176**
	(0.038)	(0.038)	(0.060)	(0.038)	(0.037)	(0.060)
Observations	277,043	277,043	277,043	277,043	277,043	277,043
R-squared	0.026	0.352	0.820	0.119	0.359	0.763
Mean outcome, grade 9	0.776	0.776	0.776	0.522	0.522	0.522
Pupil background chars	No	Yes	-	No	Yes	-
Parental background chars	No	Yes	-	No	Yes	-
Reading abilities	No	Yes	-	No	Yes	-
Math abilities	No	Yes	-	No	Yes	-
Absence information	No	Yes	-	No	Yes	-
School well-being	No	Yes	-	No	Yes	-
Personality traits	No	Yes	-	No	Yes	-
Individual FE	No	No	Yes	No	No	Yes

Notes: This table presents the results for six separate difference-in-difference specifications comparing camp participants to non-participants. Columns 1 to 3 display the simple difference-in-difference specifications, and columns 4 to 6 show the results from the difference-in-difference combined with entropy balancing. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) numbers indicate significance at the 5% (10%) level.

the desired types of upper secondary education.

I also investigate the robustness of my findings to alternative comparison groups and balancing strategies. Using alternative comparison groups' works as further tests of the difference-in-differences combined with Entropy Balancing. I exploit that boys must apply for summer camp participation themselves and this is potentially related to local school interventions. Thus, if the local school does not have sufficient programs for the boys, they and their parents might be more inclined to apply for the summer camp in order to adjust for their local schools limitations. Therefore, I test the sensitivity of the results by using only boys from school that have camp participants enrolled, as well as schools that never sends pupils to the summer camp. The estimates in table A.7 indicate that the findings do not change. Comparing camp participants to boys from their local schools or from different schools yields estimates with the same sign and magnitude as in the main specification. Additionally, I perform robustness analyses of the re-weighting approach in table A.8 to investigate if the

findings are robust to alternative balancing methods. The first column of the table presents results from the main specification and balance on the first moment, column 2 balances on the first and second moments (mean and variance), and column 3 balances on three moments (mean, variance, and skewness). Columns 4 to 6 use logistic regression to perform propensity score matching with 3, 5, and 10 non-camp boys, respectively. Overall, the estimates have similar magnitudes and signs as the main specification, indicating that the analysis is robust to alternative balancing methods.

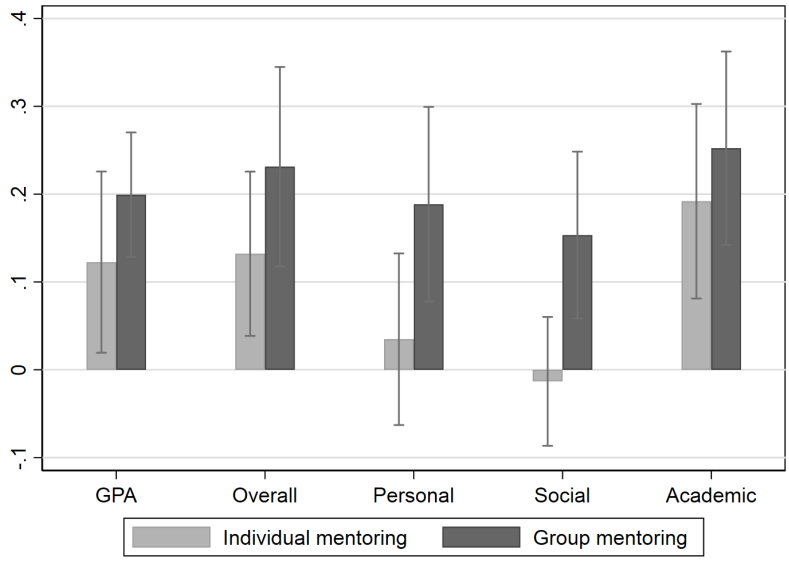
Finally, I investigate the parallel trends assumption using placebo tests. First, I perform a placebo test using a fake treatment group on the main outcomes. The fake treatment group is a random subset of boys not affected by the program. Thus, estimates different from zero indicate a violation of the identification strategy. Table A.9 shows that all estimates are essential zero. Another falsification test implements a fake camp date in the summer holiday between grades 5 and 6 for the boys participating in the camp between grades 8 and 9. To do so, I exploit the national tests in reading and math as additional outcomes. This fake camp date occurs prior to the real summer camp, and therefore the difference-in-differences strategy should find insignificant estimates close to zero in order to reject any underlying difference in trends between the camp and non-camp groups. Tables A.10 to A.12 show small and insignificant estimates using national tests in reading, grade 6 and 8, and math grade 6, indicating that the camp and non-camp group do not behave differently prior to the summer holiday between grade 8 and 9, at least not academically. An important worry is that the camp boys are more motivated to change their behavior than those who do not participate. Table A.11 show the effects on the national reading score in grade 8, which is measured after the boys have applied for the summer camp but prior to the summer camp. Thus, if motivation is driving the main effects we would expect to see significant results in this table. However, all estimates are insignificant and close to zero. To further investigate this change in behavior I construct similar placebo tests using absence, personality traits, and school well-being. Tables A.13-A.15 analysis if there is any differences from grade 7 to grade 8. This is particular interesting because behavior changes prior to summer camp could bias the ATT estimates. Absence data is measured though out the whole schoolyear whereas personality traits and school well-being is measured in the spring and similar to the national reading test this survey is conducted in the end or after the application deadline but prior to notification of enrollment at the summer camp. The tables show no evidence of increased motivation with the individual fixed effect estimates being statistically insignificant. This indicates no systematic change between the two groups leading up to the camp, which supports the main results.

2.6 Results: Individual vs group mentoring

In this section, I estimate the effect of the change from individual to a group mentoring strategy in the follow-up program. As a first indication, I present a visual illustration of the difference-in-differences results from the two school years with an individual mentoring against the later three years with group mentoring. Overall, figure 2.2 shows larger effects under the group mentoring follow-up program across all outcomes. The effect of camp participation on continuous assessment mark is positive and significant in both mentoring regimes. However, the effect nearly doubles from implementation of group mentoring. Interestingly, during the years of the individual mentoring program there are no effects on the personal and social readiness assessment whereas the introduction of group mentoring substantially increases these effects to 19 and 15 percentage points, respectively. For academic readiness, the difference-in-differences estimates are significantly different from zero under both the individual and group mentoring follow-up program with the latter being slightly larger. These differences in readiness assessments in different dimensions are also detectable in the overall readiness assessment, in which the difference-in-difference results increase from 12 percentage points to 22 percentage points from changing the individual mentoring program to a group mentoring program during the follow-up year.

The obvious question is whether these effects are statistically significant from each other. Therefore, I investigate the consequences of the follow-up mentoring strategy using the triple differences model described in equation (2.2). Table 2.2 presents the triple differences estimates, i.e. the effect of substituting the individual mentoring program with a group mentoring program. The first column shows that the effect of camp participation on continuous assessment mark is 8% of a standard deviation larger when the follow-up program consist of group mentoring. The difference is, however, not significantly different. For the educational readiness assessment, the table shows overall positive results with the effects on the personal and social readiness assessment being statistically significant. Substitution from the individual mentoring follow-up program to a group mentoring follow-up program increases the fraction of boys who are assessed personally ready by 15 percentage points and socially ready by 16 percentage points. These results imply that, in terms of personal and social readiness, camp boys reduce the gap to the average boy by up to 60%. This is an important finding because a major problem with intensive learning camps is the transition back into the local school environment. When returning to the local school and classroom, many boys perceive that "nothing has changed" and therefore, their school behavior is likely to revert to how it was before camp participation. However, the correct mentoring strategy in the follow-up program seems able to reduce this problem and avoid the complete fade-out in teacher assessed personal and social competencies. This is crucial if we want to hope for long-run effects of educational learning programs on lifetime success (Kautz et al., 2014).

Figure 2.2: Difference-in-difference estimates by follow-up program



Notes: *This figure presents the difference-in-difference effects of participating in the summer camp on outcomes separated by individual (2015-2016) and group (2017-2019) mentoring with 95% confidence intervals. GPA is the average assessment mark and is measured as standardized effect sizes. The remaining outcomes stems from the readiness assessment and can be interpreted as percentage points by multiplying with 100.*

The critical assumptions for the triple differences estimates to be causal is the parallel trends between the two difference-in-differences estimates. Entropy balancing ensures perfect balance across all covariates and the large set of robustness analyses conducted in the previous section show no indication of violation of the difference-in-differences parallel trend assumption. Data restrictions - educational readiness assessment only performed once in grade 8 and 9 and the continuous assessment mark for June in grade 8 is first registered in 2018 - makes visualization of pre-trends impossible. However, as in the above section, I perform placebo test with fake camp group and fake camp date. Table A.16 presents the triple differences results for the fake camp group, which are all insignificant. Tables A.17 and A.18 investigate academic pre-trends exploiting the national reading and math tests and a fake camp date in the summer holiday between grade 5 and 6. All triple differences estimates are insignificant, indicating similar pre-trends between the two difference-in-differences models. Additionally, the selection process to the summer camp has not changed as illustrated in table A.19. This all indicates that the changing effects on personal and social competencies are caused by the change in the follow-up program.

Table 2.2: Triple differences: Effects of group mentoring instead of individual mentoring

		Educational readiness assessment			
	GPA	Overall	Personal	Social	Academic
	(1)	(2)	(3)	(4)	(5)
DiDiD	0.082	0.097	**0.154**	**0.155**	0.062
	(0.065)	(0.076)	(0.076)	(0.063)	(0.081)
Observations	479,540	277,043	263,480	263,153	265,354
R-squared	0.565	0.362	0.268	0.261	0.418
Pupil background chars	Yes	Yes	Yes	Yes	Yes
Parental background chars	Yes	Yes	Yes	Yes	Yes
Reading abilities	Yes	Yes	Yes	Yes	Yes
Math abilities	Yes	Yes	Yes	Yes	Yes
Absence information	Yes	Yes	Yes	Yes	Yes
School well-being	Yes	Yes	Yes	Yes	Yes
Personality traits	Yes	Yes	Yes	Yes	Yes

Notes: This table shows triple difference (DiDiD) estimates by comparing camp participants attending the individual mentoring program with those attending group mentor centers in the year following the camp. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) numbers indicate significance at the 5% (10%) level.

2.7 Conclusion and discussion

I studied how summer camps targeted academically disadvantaged boys with lack of school motivation affect school performance and readiness for upper secondary education in Denmark. I did so by utilizing how boys' outcomes evolved from grade 8 to grade 9. Using a difference-in-differences strategy, I compared outcomes among those who participated in the two-week summer camp to those who did not. Additionally, I investigated – exploiting a 2017 structural change in the one-year follow-up program – how substituting individual mentoring with group mentoring affected outcomes. Using a triple differences strategy, I compared the difference-in-differences estimates before and after the change in the follow-up program.

I found large, positive effects of summer camp participation on the average assessment mark and on readiness for upper secondary education. Summer camp participation increased the average assessment mark by 15% of a standard deviation and overall readiness for upper secondary education by 18 percentage points. Thus, the summer camp reduced the gap to the average boy with up to 35% and to boys with similar characteristics with up to 80%. I found that the positive effect on the overall readiness assessment was mainly driven by increased academic competencies (22 percentage points), but the camp also increased personal competencies

(11 percentage points). Further, I saw an increase in the effects of summer camp participation on all outcomes when using group mentoring in the follow-up program instead of individual mentoring. In fact, substitution from the individual mentoring follow-up program to a group mentoring follow-up program significantly increased the fraction of boys who were assessed personally ready by 15 percentage points and socially ready by 16 percentage points. Thus, the change in follow-up strategy led to a dramatic increase in the effectiveness of the summer camp on non-academic competencies.

This study thus offers two key findings; 1) summer camps are effective in improving adolescent boys' school outcomes and 2) implementing a group mentoring scheme in the follow-up program increases the effects. Interestingly, the results suggest that the two weeks of camp primarily increases academic competencies while the follow-up program supports personal and social competencies. Intuitively, it makes sense that group mentoring, where such competencies automatically are in play, is more effective than individual mentoring in improving social and personal competencies. Additionally, group mentoring has the advantage of being cheaper because of the lower adult to boy ratio. These findings raise the question if a reduction in camp length and additional follow-up sessions with the mentor group, which would reduce the total cost of the intervention, is more cost-effective. A new similar one-week camp with similar follow-up program was, in the autumn holiday of 2021, commenced by the same organization that arrange the summer camp under study. Future research exploiting both the summer and autumn camp may be able to shed light on this question.

Kraft (2020) argues that, from a policy perspective, the effect sizes are important, but more important is the effect relative to program costs. Furthermore, he argues that learning programs are not as relevant if they cannot be taken to scale with high fidelity. This summer camp cost approximately 5,000USD per participant, which Kraft (2020) defines as being in the lower end of a high-cost intervention. However, when accounting for the degree of disadvantage facing this group of adolescent boys, which the literature has shown is difficult to affect, the effects of this summer camp appear quite impressive. There is naturally a limited number of participating boys at each camp, but with the comprehensive curriculum and fixed structure, it would seem fairly straightforward to scale up without affecting the gains by conducting several camps simultaneously.

Overall, the findings in this study are important for policymakers because they show that summer camp can be a relevant and effective tool for improving academic performance for disadvantaged boys who have fallen behind academically. Furthermore, with the large impact on readiness for upper secondary education, the summer camp has the potential to become a key instrument in making pupils ready for upper secondary education. Naturally, the cost of 5,000USD per participant has to be assessed against the benefits of camp participation. For example, the reduction in cost to society of boys' not attaining secondary education or delaying it. Finally, the

many disruptions to education caused by e.g. the Covid-19 pandemic flattens the learning curve and increases the socioeconomic learning gap. Di Pietro et al. (2020) show that these disruptions to a greater extent affect disadvantaged boys by stopping their learning process and in fact decrease their cognitive abilities. Using summer camps as a remedial educational program would have the potential to eliminate parts of this learning gap.

An important limitation with this study is the gender selection, i.e. the summer camp targets boys and the analysis provide no insight into its effect on girls. Thus, future research is needed on the effects of summer camp participation for disadvantaged girls. Another limitation to this study is methodological. The analysis builds on a difference-in-differences strategy using panel data. This requires parallel trends and removes the possibility to investigate long-run effects, such as educational attainment, future income, employment history, etc. Therefore, the next step is to study the summer camp using a randomized trial, which would enable long-term follow-up and require much less restrictive identifying assumptions.

Acknowledgements

I am grateful to TrygFonden for funding through TrygFonden's Centre for Child Research. I gratefully acknowledge helpful comments from Marianne Simonsen, Michael Rosholm, and seminar participants at Aarhus University. I would also like to thank Charlotte Hvidman and Søren Munkedal for their intellectual contribution to the project.

2.8 References

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

Almlund, M., Duckworth, A. L., Heckman, J., Kautz, T., 2011. Personality Psychology and Economics. Handbook of the Economics of Education, 1–181.
URL https://econpapers.repec.org/bookchap/eeeeduchp/4-1.htm

Andersen, F. , Mølgaard, L., Wåst, T., Nørgaad, L., 2019. Intensiv læring - relationer, mestringsoplevelser og transfer i det nye pædagogiske landskab. Dafolo A/S, Frederikshavn.
URL https://www.dafoloforlag.dk/dk/skole/indskoling/intensiv-laring-relationer-mestringsoplevelser-og-transfer-i-det-nye-padagogiske-landskab--7083

Autor, D., Figlio, D., Karbownik, K., Roth, J., Wasserman, M., Jul. 2019. Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes. American Economic Journal: Applied Economics 11 (3), 338–381.
URL https://pubs.aeaweb.org/doi/10.1257/app.20170571

Battistin, E., Schizzerotto, A., 2019. Threat of grade retention, remedial education and student achievement: Evidence from upper secondary schools in italy. Empirical Economics 56 (2), 651–678.

Bertrand, M., Duflo, E., Mullainathan, S., 2004. How much should we trust differences-in-differences estimates? Quarterly journal of economics, 27.

Blundell, R., Dias, M. C., Meghir, C., van Reenen, J., Jun. 2004. Evaluating the Employment Impact of a Mandatory Job Search Program. Journal of the European Economic Association 2 (4), 569–606.
URL https://academic.oup.com/jeea/article/2280805/Evaluating

Cooper, H., Charlton, K., Valentine, J. C., Muhlenbruck, L., Borman, G. D., 2000. Making the most of summer school: A meta-analytic and narrative review. Monographs of the Society for Research in Child Development, i–127.

Di Pietro, G., Biagi, F., Costa, P., Karpiński, Z., Mazza, J., 2020. The likely impact of COVID-19 on education: reflections based on the existing literature and recent international datasets. Tech. rep., Publications Office of the European Union, LU.
URL https://data.europa.eu/doi/10.2760/126686

Freier, R., Schumann, M., Siedler, T., Jun. 2015. The earnings returns to graduating with honors — Evidence from law graduates. Labour Economics 34, 39–50.
URL http://www.sciencedirect.com/science/article/pii/S0927537115000275

Gruber, J., 1994. The Incidence of Mandated Maternity Benefits. The American Economic Review vol. 84 (3), 622–641.
URL https://www.jstor.org/stable/2118071

Hainmueller, J., 2012. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. Political Analysis 20 (1), 25–46.
URL https://www.cambridge.org/core/product/identifier/S1047198700012997/type/journal_article

Hainmueller, J., Xu, Y., Sep. 2013. ebalance: A Stata Package for Entropy Balancing. Journal of Statistical Software 54 (1), 1–18.
URL https://www.jstatsoft.org/index.php/jss/article/view/v054i07

Hattie, J., Yates, G. C. R., Sep. 2013. Visible Learning and the Science of How We Learn. Routledge, London.

Heckman, J. J., Ichimura, H., Todd, P. E., Oct. 1997. Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. The Review of Economic Studies 64 (4), 605–654.
URL https://academic.oup.com/restud/article-lookup/doi/10.2307/2971733

Jacob, B. A., Lefgren, L., 2004. Remedial education and student achievement: A regression-discontinuity analysis. Review of economics and statistics 86 (1), 226–244, publisher: MIT Press.

Kautz, T., Heckman, J. J., Diris, R., ter Weel, B., Borghans, L., Dec. 2014. Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success. Working Paper 20749, National Bureau of Economic Research, series: Working Paper Series.
URL https://www.nber.org/papers/w20749

Kim, J. S., Quinn, D. M., 2013. The effects of summer reading on low-income children's literacy achievement from kindergarten to grade 8: A meta-analysis of classroom and home interventions. Review of Educational Research 83 (3), 386–431.

Kraft, M. A., May 2020. Interpreting Effect Sizes of Education Interventions. Educational Researcher 49 (4), 241–253.
URL http://journals.sagepub.com/doi/10.3102/0013189X20912798

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., Martin-Glenn, M. L., 2006. Out-of-school-time programs: A meta-analysis of effects for at-risk students. Review of Educational Research 76 (2), 275–313.

Mariano, L. T., Martorell, P., 2013. The academic effects of summer instruction and retention in New York City. Educational Evaluation and Policy Analysis 35 (1), 96–117, publisher: SAGE Publications Sage CA: Los Angeles, CA.

Murnane, R. J., Jun. 2013. U.S. High School Graduation Rates: Patterns and Explanations. Journal of Economic Literature 51 (2), 370–422.
URL https://pubs.aeaweb.org/doi/10.1257/jel.51.2.370

OECD, 2013. Education at a glance 2013: Oecd indicators. OECD Publishing, Paris.

Olden, A., Møen, J., 2020. The Triple Difference Estimator. NHH Discussion paper, 26.

Petrosino, A., Turpin-Petrosino, C., Hollis-Peel, M. E., Lavenberg, J. G., Apr. 2013. 'Scared Straight' and other juvenile awareness programs for preventing juvenile delinquency. The Cochrane Database of Systematic Reviews (4), CD002796.

Rambøll, 2019. Kvalitativ undersøgelse af Plan T, DrengeAkademiet og KøbenhavnerAkademiet. Tech. rep., Egmont Fonden and Bikubenfonden.
URL https://emu.dk/sites/default/files/2021-01/Ramb%C3%B8ll%202019%20-%20Kvalitativ%20unders%C3%B8gelse%20af%20Plan%20Tmm.pdf

Rosholm, M., Paul, A., Bleses, D., Højen, A., S. Dale, P., Jensen, P., M. Justice, L., Svarer, M., Calmar Andersen, S., Feb. 2021. Are impacts of early interventions in the scandinavian welfare state consistent with a heckman curve? a meta-analysus. Journal of Economic Surveys 35 (1), 106–140.
URL https://onlinelibrary.wiley.com/doi/10.1111/joes.12400

Ryan, R. M., Deci, E. L., 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 55 (1), 68–78, place: US Publisher: American Psychological Association.

A.1 Appendix

Table A.1: Entropy Balancing Statistics

	Camp		Non-Camp					
			Unadjusted			Adjusted		
	Mean	SD	Mean	SD	Diff.	Mean	SD	Diff.
Pupil:								
- Age (years)	15.299	(0.459)	15.227	(0.419)	**0.072**	15.299	(0.458)	0.000
- Non-western etnicity (1/0)	0.046	(0.210)	0.090	(0.287)	**-0.045**	0.046	(0.209)	0.000
- Living with both parents (1/0)	0.556	(0.498)	0.646	(0.478)	**-0.090**	0.556	(0.497)	0.000
- Number of relocations	2.095	(1.723)	1.865	(1.616)	**0.230**	2.095	(1.760)	0.000
- Years in private schooling	0.817	(1.906)	1.073	(2.253)	*-0.255*	0.817	(1.969)	0.000
- Number of school changes	0.834	(0.835)	0.630	(0.773)	**0.204**	0.834	(0.885)	0.000
- Years in special needs teaching	0.734	(1.253)	0.293	(0.929)	**0.441**	0.734	(1.281)	0.000
- Total hours of special needs teaching	6.421	(18.835)	4.563	(21.237)	1.859	6.422	(21.010)	0.000
- Dyslexics (1/0)	0.353	(0.479)	0.101	(0.301)	**0.252**	0.353	(0.478)	0.000
- OCD (1/0)	0.071	(0.257)	0.025	(0.155)	**0.046**	0.071	(0.256)	0.000
- ADHD (1/0)	0.095	(0.294)	0.038	(0.191)	**0.058**	0.095	(0.294)	0.000
- Any psychological diagnosis (1/0)	0.207	(0.406)	0.100	(0.299)	**0.108**	0.208	(0.406)	0.000
Mother:								
- Age at birth (years)	29.737	(4.932)	30.306	(4.764)	*-0.570*	29.737	(4.917)	0.000
- Wage income (log)	5.607	(1.078)	5.682	(1.017)	-0.075	5.607	(0.987)	0.000
- Years of education	14.125	(2.668)	14.512	(2.732)	**-0.388**	14.125	(2.376)	0.000
- Highschool or less (1/0)	0.191	(0.394)	0.200	(0.400)	-0.009	0.191	(0.393)	0.000
- Employed in November (1/0)	0.722	(0.449)	0.763	(0.425)	-0.041	0.722	(0.448)	0.000
Father:								
- Age at birth (years)	32.456	(5.774)	32.952	(5.620)	-0.497	32.456	(5.662)	0.000
- Wage income (log)	5.821	(1.089)	5.958	(1.150)	*-0.137*	5.821	(1.165)	0.000
- Years of education	13.859	(2.437)	14.414	(2.615)	**-0.555**	13.859	(2.457)	0.000
- Highschool or less (1/0)	0.266	(0.443)	0.249	(0.433)	0.016	0.266	(0.442)	0.000
- Employed in November (1/0)	0.787	(0.410)	0.827	(0.378)	-0.039	0.788	(0.409)	0.000
Language comprehension (std.):								
- Grade 2	-0.509	(0.933)	-0.026	(0.990)	**-0.483**	-0.509	(1.069)	0.000
- Grade 4	-0.680	(1.008)	0.001	(0.995)	**-0.680**	-0.680	(1.207)	0.000
- Grade 6	-0.639	(0.894)	0.002	(0.991)	**-0.640**	-0.639	(1.083)	0.000
- Grade 8	-0.546	(1.113)	0.036	(1.028)	**-0.583**	-0.546	(1.186)	0.000
Decoding (std.):								
- Grade 2	-0.982	(0.829)	-0.064	(1.006)	**-0.918**	-0.982	(0.990)	0.000

- Grade 4	-0.973	(1.059)	-0.026	(1.008)	**-0.947**	-0.973	(1.182)	0.000
- Grade 6	-0.956	(0.841)	-0.068	(0.987)	**-0.888**	-0.956	(1.139)	0.000
- Grade 8	-1.095	(0.914)	-0.071	(1.010)	**-1.024**	-1.095	(1.306)	0.000
Text comprehension (std.):								
- Grade 2	-0.942	(0.957)	-0.090	(1.020)	**-0.851**	-0.942	(1.080)	0.000
- Grade 4	-0.906	(0.975)	-0.065	(0.996)	**-0.841**	-0.906	(1.116)	0.000
- Grade 6	-0.821	(0.847)	-0.049	(0.992)	**-0.772**	-0.821	(1.066)	0.000
- Grade 8	-0.927	(0.982)	-0.061	(0.998)	**-0.866**	-0.927	(1.158)	0.000
Numbers and algrebra (std.):								
- Grade 3	-0.699	(0.962)	0.035	(0.999)	**-0.734**	-0.699	(1.116)	0.000
- Grade 6	-0.667	(0.911)	0.058	(0.985)	**-0.725**	-0.667	(1.032)	0.000
- Grade 8	-0.833	(0.766)	0.051	(1.001)	**-0.884**	-0.833	(1.021)	0.000
Geometry (std.):								
- Grade 3	-0.606	(0.779)	0.023	(1.028)	**-0.629**	-0.606	(1.053)	0.000
- Grade 6	-0.625	(0.826)	0.011	(0.995)	**-0.636**	-0.625	(0.976)	0.000
- Grade 8	-0.834	(0.667)	0.063	(0.984)	**-0.897**	-0.834	(0.935)	0.000
Statistics and probability (std.):								
- Grade 3	-0.724	(0.954)	0.080	(1.008)	**-0.805**	-0.724	(1.166)	0.000
- Grade 6	-0.733	(0.904)	0.081	(1.008)	**-0.814**	-0.733	(1.075)	0.000
- Grade 8	-0.814	(0.777)	0.035	(0.987)	**-0.849**	-0.814	(1.072)	0.000
Sick absence (percent):								
- Grade 2	2.301	(2.244)	2.804	(3.183)	-0.503	2.301	(2.417)	0.000
- Grade 3	2.733	(2.867)	2.822	(3.298)	-0.089	2.733	(3.112)	0.000
- Grade 4	3.172	(3.461)	3.020	(3.591)	0.152	3.172	(3.830)	0.000
- Grade 5	3.270	(3.761)	3.196	(3.921)	0.074	3.270	(4.031)	0.000
- Grade 6	3.230	(3.878)	3.288	(4.163)	-0.058	3.229	(3.834)	0.000
- Grade 7	3.374	(4.424)	3.355	(4.394)	0.018	3.374	(4.283)	0.000
- Grade 8	3.172	(4.098)	3.413	(4.745)	-0.241	3.172	(3.733)	0.000
Illegal absence (percent):								
- Grade 2	0.587	(1.536)	0.432	(1.821)	0.155	0.587	(3.173)	0.000
- Grade 3	0.317	(1.176)	0.460	(1.908)	-0.143	0.317	(1.108)	0.000
- Grade 4	0.383	(0.936)	0.532	(2.011)	-0.148	0.383	(1.192)	0.000
- Grade 5	0.629	(2.340)	0.622	(2.279)	0.007	0.629	(2.301)	0.000
- Grade 6	0.754	(2.052)	0.790	(2.755)	-0.036	0.754	(2.293)	0.000
- Grade 7	1.616	(3.936)	1.280	(3.580)	0.336	1.616	(4.563)	0.000
- Grade 8	2.332	(5.236)	1.753	(4.973)	*0.578*	2.332	(5.946)	0.000
Legal absence (percent):								
- Grade 2	2.057	(2.757)	1.546	(2.384)	*0.510*	2.056	(5.982)	0.000
- Grade 3	1.476	(2.179)	1.484	(2.316)	-0.008	1.476	(2.346)	0.000
- Grade 4	1.625	(2.630)	1.420	(2.316)	0.205	1.625	(3.136)	0.000
- Grade 5	1.359	(1.905)	1.409	(2.439)	-0.050	1.359	(2.386)	0.000
- Grade 6	1.268	(1.702)	1.383	(2.475)	-0.115	1.268	(2.060)	0.000
- Grade 7	1.701	(2.264)	1.462	(2.502)	0.239	1.701	(3.613)	0.000
- Grade 8	2.443	(7.064)	1.724	(2.941)	**0.719**	2.443	(5.367)	0.000
Conscientiousness (std.):								

- Grade 4	-0.621	(0.964)	0.040	(1.038)	**-0.661**	-0.621	(1.115)	0.000
- Grade 5	-0.591	(0.901)	0.074	(0.999)	**-0.666**	-0.591	(1.088)	0.000
- Grade 6	-0.420	(0.839)	0.092	(0.956)	**-0.512**	-0.420	(1.045)	0.000
- Grade 7	-0.433	(0.890)	0.054	(0.943)	**-0.487**	-0.433	(1.035)	0.000
- Grade 8	-0.616	(0.932)	0.031	(0.947)	**-0.647**	-0.616	(1.082)	0.000
Agreeableness (std.):								
- Grade 4	-0.061	(0.983)	-0.028	(1.025)	-0.033	-0.061	(1.084)	0.000
- Grade 5	-0.575	(1.168)	-0.088	(1.007)	**-0.487**	-0.575	(1.168)	0.000
- Grade 6	-0.378	(0.892)	-0.136	(0.987)	**-0.242**	-0.378	(1.081)	0.000
- Grade 7	-0.435	(0.941)	-0.184	(0.996)	**-0.252**	-0.435	(1.085)	0.000
- Grade 8	-0.527	(1.051)	-0.172	(0.998)	**-0.355**	-0.527	(1.128)	0.000
Emotional stability (std.):								
- Grade 4	0.097	(0.827)	0.131	(0.981)	-0.034	0.097	(1.023)	0.000
- Grade 5	-0.276	(1.198)	0.179	(0.953)	**-0.455**	-0.277	(1.217)	0.000
- Grade 6	-0.042	(1.012)	0.198	(0.933)	**-0.240**	-0.042	(1.075)	0.000
- Grade 7	-0.008	(0.982)	0.165	(0.910)	**-0.173**	-0.008	(1.020)	0.000
- Grade 8	-0.103	(0.980)	0.153	(0.904)	**-0.255**	-0.103	(1.053)	0.000
School connectedness (std.):								
- Grade 4	0.158	(0.869)	0.173	(0.959)	-0.015	0.158	(0.964)	0.000
- Grade 5	-0.310	(1.181)	0.186	(0.942)	**-0.496**	-0.310	(1.199)	0.000
- Grade 6	-0.096	(0.975)	0.158	(0.938)	**-0.254**	-0.096	(1.082)	0.000
- Grade 7	-0.048	(0.930)	0.094	(0.916)	*-0.142*	-0.048	(0.971)	0.000
- Grade 8	-0.209	(0.933)	0.046	(0.919)	**-0.255**	-0.209	(1.051)	0.000
Learning self-efficacy (std.):								
- Grade 4	-0.378	(0.782)	0.120	(1.014)	**-0.498**	-0.378	(1.051)	0.000
- Grade 5	-0.615	(0.978)	0.120	(0.988)	**-0.735**	-0.615	(1.110)	0.000
- Grade 6	-0.563	(0.778)	0.100	(0.966)	**-0.664**	-0.563	(1.063)	0.000
- Grade 7	-0.620	(0.935)	0.036	(0.965)	**-0.656**	-0.620	(1.048)	0.000
- Grade 8	-0.962	(0.992)	-0.029	(0.982)	**-0.932**	-0.962	(1.157)	0.000
Learning environment (std.):								
- Grade 4	0.370	(0.840)	0.374	(0.997)	-0.003	0.370	(1.080)	0.000
- Grade 5	-0.038	(1.045)	0.189	(0.986)	*-0.227*	-0.038	(1.114)	0.000
- Grade 6	0.044	(0.913)	0.002	(0.984)	0.042	0.044	(1.000)	0.000
- Grade 7	-0.240	(1.020)	-0.177	(0.980)	-0.063	-0.240	(1.004)	0.000
- Grade 8	-0.420	(0.978)	-0.298	(0.980)	-0.122	-0.420	(1.020)	0.000
Classroom management (std.):								
- Grade 4	0.018	(0.959)	0.023	(1.020)	-0.004	0.018	(1.093)	0.000
- Grade 5	-0.306	(1.133)	0.021	(0.993)	**-0.327**	-0.307	(1.128)	0.000
- Grade 6	0.006	(0.981)	0.036	(0.989)	-0.029	0.006	(1.026)	0.000
- Grade 7	-0.226	(1.029)	0.014	(0.995)	**-0.240**	-0.226	(1.077)	0.000
- Grade 8	-0.132	(1.158)	0.059	(0.997)	**-0.191**	-0.132	(1.093)	0.000
Cohort:								
- 2015	0.274	(0.447)	0.201	(0.400)	**0.073**	0.274	(0.446)	0.000
- 2016	0.195	(0.397)	0.198	(0.399)	-0.003	0.195	(0.396)	0.000
- 2017	0.203	(0.403)	0.194	(0.395)	0.009	0.203	(0.402)	0.000

- 2018	0.141	(0.349)	0.203	(0.402)	**-0.062**	0.141	(0.348)	0.000
- 2019	0.187	(0.390)	0.204	(0.403)	-0.017	0.187	(0.390)	0.000
Number of pupils	241		157,990					

Notes: The table shows descriptive statistics of covariates used in the analysis. The mean, standard derivation and difference in means are reported for the camp group and for the non-camp group before and after adjustment. The non-camp group are reweighted using entropy balancing such that it mimics the camp group. Pupil and parent background characteristics are measured at grade 8. The table is based on non-missing data. Bold (italic) indicates significance at the 5% (10%) level.

Figure A.1: Example of the main camp schedule for the first week

Figure A.2: Example of the main camp schedule for the second week

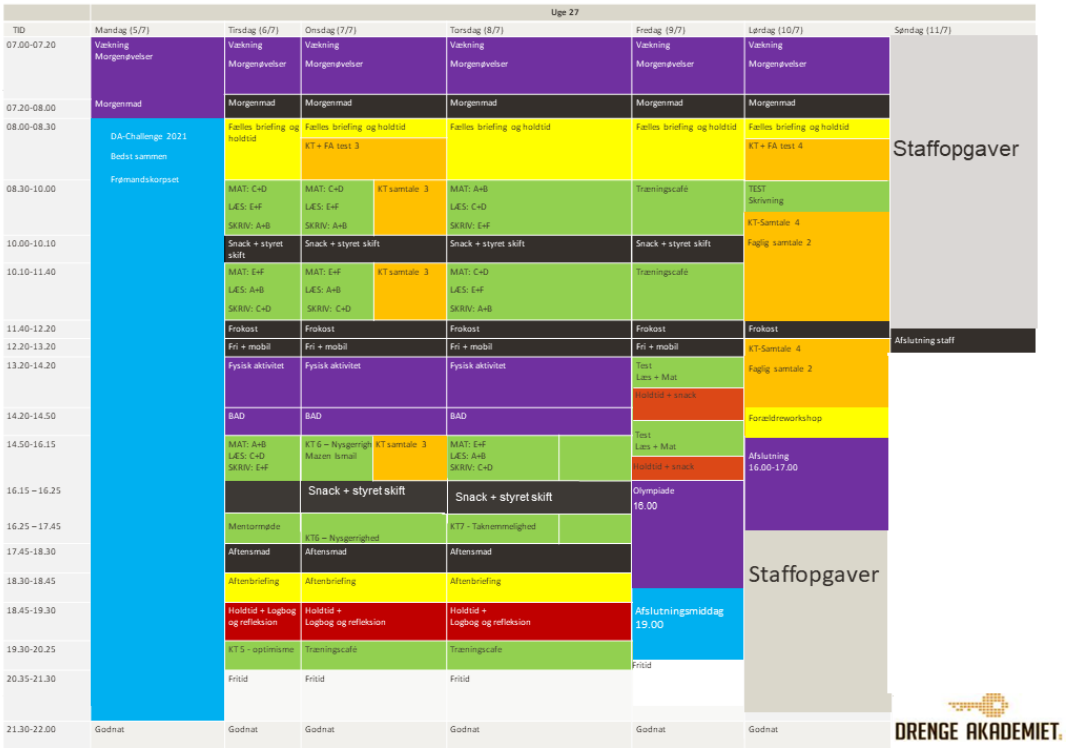
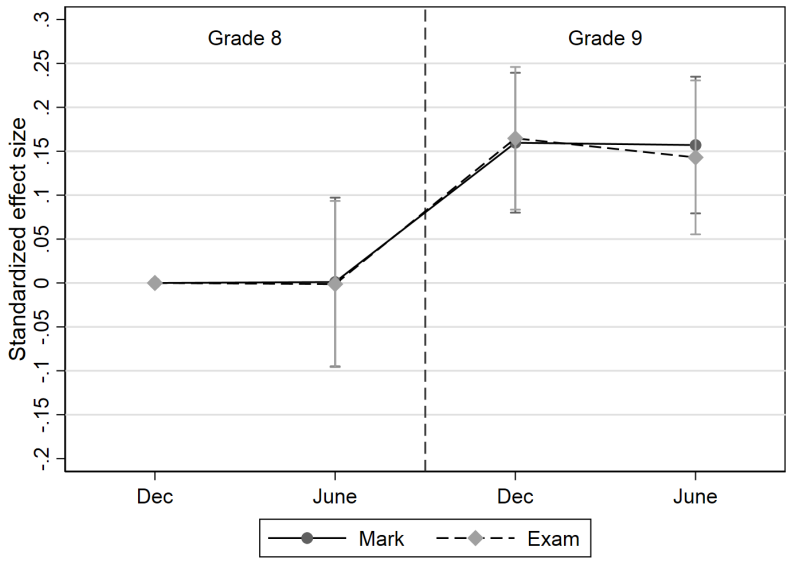
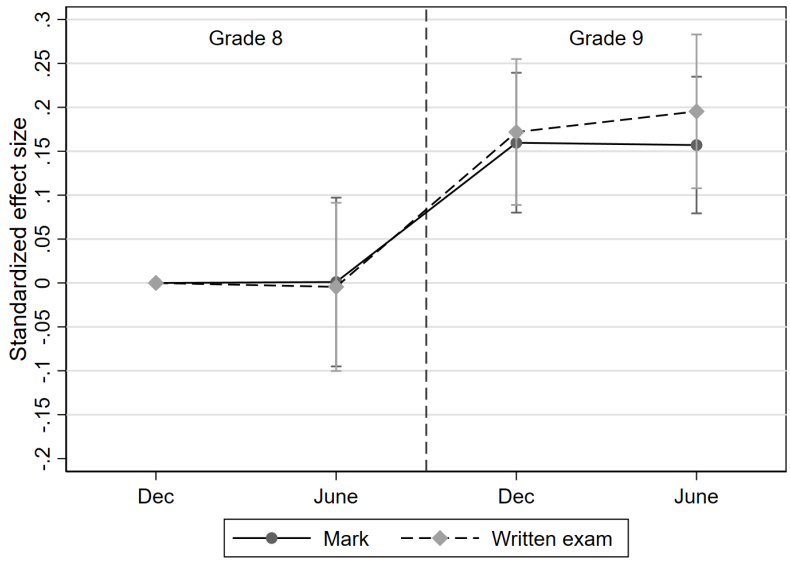


Figure A.3: Summer camp and Exam



Notes: *This figure shows the entropy balancing adjusted event study representation of the effect of summer camp participation. The average assessment mark in June grade 9 is substituted with the exam performance. The solid line presents the effects on the average assessment mark and the dashed line presents the effects in which the June grade 9 grading have been substituted with the exam performance. Each point represents the assessment difference in outcome between camp and non-camp boys with 95% confidence intervals. I use the assessment in December grade 8 as reference time-point.*

Figure A.4: Summer camp and written exam



Notes: *This figure shows the entropy balancing adjusted event study representation of the effect of summer camp participation. The average assessment mark in June grade 9 is substituted with the written exam performance. The solid line presents the effects on the average assessment mark and the dashed line presents the effects in which the June grade 9 grading have been substituted with the exam performance. Each point represents the assessment difference in outcome between camp and non-camp boys with 95% confidence intervals. I use the assessment in December grade 8 as reference time-point.*

Table A.2: Effects of summer camp on average assessment mark

	Difference-in-difference			DiD & Entropy balance		
	(1)	(2)	(3)	(4)	(5)	(6)
Camp	**-1.179**	**-0.314**		**-0.390**	**-0.392**	
	(0.038)	(0.041)		(0.039)	(0.028)	
Post	**0.025**	**0.018**	**0.020**	**0.073**	**0.020**	**0.014**
	(0.002)	(0.001)	(0.001)	(0.007)	(0.007)	(0.005)
Camp X Post	**0.167**	**0.125**	**0.152**	**0.120**	**0.154**	**0.158**
	(0.035)	(0.036)	(0.036)	(0.036)	(0.031)	(0.037)
Observations	479,540	479,540	479,540	479,540	479,540	479,540
R-squared	0.002	0.624	0.944	0.049	0.564	0.910
Mean outcome, grade 9	-0.088	-0.088	-0.088	-0.966	-0.966	-0.966
Pupil background chars	No	Yes	-	No	Yes	-
Parental background chars	No	Yes	-	No	Yes	-
Reading abilities	No	Yes	-	No	Yes	-
Math abilities	No	Yes	-	No	Yes	-
Absence information	No	Yes	-	No	Yes	-
School well-being	No	Yes	-	No	Yes	-
Personality traits	No	Yes	-	No	Yes	-
Individual FE	No	No	Yes	No	No	Yes

Notes: This table presents the results for six separate difference-in-difference specifications comparing camp participants to non-participants. Columns 1 to 3 display the simple difference-in-difference specifications, and columns 4 to 6 show the results from the difference-in-difference combined with entropy balancing. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. The mean outcome in grade 9 is below zero since the standardization of outcome also include girls. Bold (italic) numbers indicate significance at the 5% (10%) level.

Table A.3: Effects of summer camp on personal readiness assessment

	Difference-in-difference			DiD & Entropy balance		
	(1)	(2)	(3)	(4)	(5)	(6)
Camp	**-0.348**	**-0.141**		**-0.132**	**-0.126**	
	(0.033)	(0.035)		(0.034)	(0.027)	
Post	**0.067**	**0.069**	**0.072**	**0.092**	**0.097**	**0.090**
	(0.001)	(0.001)	(0.002)	(0.004)	(0.007)	(0.005)
Camp X Post	**0.121**	**0.114**	**0.137**	**0.097**	**0.096**	*0.119*
	(0.041)	(0.040)	(0.065)	(0.041)	(0.038)	(0.065)
Observations	263,480	263,480	263,480	263,480	263,480	263,480
R-squared	0.007	0.252	0.834	0.029	0.259	0.802
Mean outcome, grade 9	0.807	0.807	0.807	0.598	0.598	0.598
Pupil background chars	No	Yes	-	No	Yes	-
Parental background chars	No	Yes	-	No	Yes	-
Reading abilities	No	Yes	-	No	Yes	-
Math abilities	No	Yes	-	No	Yes	-
Absence information	No	Yes	-	No	Yes	-
School well-being	No	Yes	-	No	Yes	-
Personality traits	No	Yes	-	No	Yes	-
Individual FE	No	No	Yes	No	No	Yes

Notes: This table presents the results for six separate difference-in-difference specifications comparing camp participants to non-participants. Columns 1 to 3 display the simple difference-in-difference specifications, and columns 4 to 6 show the results from the difference-in-difference combined with entropy balancing. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) numbers indicate significance at the 5% (10%) level.

Table A.4: Effects of summer camp on social readiness assessment

	Difference-in-difference			DiD & Entropy balance		
	(1)	(2)	(3)	(4)	(5)	(6)
Camp	**-0.267**	**-0.120**		**-0.109**	**-0.106**	
	(0.034)	(0.034)		(0.034)	(0.027)	
Post	**0.048**	**0.048**	**0.050**	**0.070**	**0.078**	**0.065**
	(0.001)	(0.001)	(0.001)	(0.004)	(0.007)	(0.005)
Camp X Post	**0.072**	**0.071**	*0.097*	0.050	0.049	0.083
	(0.035)	(0.034)	(0.052)	(0.035)	(0.033)	(0.052)
Observations	263,153	263,153	263,153	263,153	263,153	263,153
R-squared	0.005	0.203	0.837	0.019	0.251	0.846
Mean outcome, grade 9	0.871	0.871	0.871	0.705	0.705	0.705
Pupil background chars	No	Yes	-	No	Yes	-
Parental background chars	No	Yes	-	No	Yes	-
Reading abilities	No	Yes	-	No	Yes	-
Math abilities	No	Yes	-	No	Yes	-
Absence information	No	Yes	-	No	Yes	-
School well-being	No	Yes	-	No	Yes	-
Personality traits	No	Yes	-	No	Yes	-
Individual FE	No	No	Yes	No	No	Yes

Notes: This table presents the results for six separate difference-in-difference specifications comparing camp participants to non-participants. Columns 1 to 3 display the simple difference-in-difference specifications, and columns 4 to 6 show the results from the difference-in-difference combined with entropy balancing. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) numbers indicate significance at the 5% (10%) level.

Table A.5: Effects of summer camp on academic readiness assessment

	Difference-in-difference			DiD & Entropy balance		
	(1)	(2)	(3)	(4)	(5)	(6)
Camp	**-0.567**	**-0.321**		**-0.269**	**-0.256**	
	(0.027)	(0.026)		(0.027)	(0.023)	
Post	**0.080**	**0.065**	**0.063**	**0.212**	**0.186**	**0.183**
	(0.001)	(0.001)	(0.002)	(0.005)	(0.006)	(0.006)
Camp X Post	**0.355**	**0.322**	**0.335**	**0.223**	**0.218**	**0.216**
	(0.040)	(0.040)	(0.067)	(0.040)	(0.039)	(0.067)
Observations	265,354	265,354	265,354	265,354	265,354	265,354
R-squared	0.011	0.353	0.804	0.145	0.417	0.767
Mean outcome, grade 9	0.834	0.834	0.834	0.645	0.645	0.645
Pupil background chars	No	Yes	-	No	Yes	-
Parental background chars	No	Yes	-	No	Yes	-
Reading abilities	No	Yes	-	No	Yes	-
Math abilities	No	Yes	-	No	Yes	-
Absence information	No	Yes	-	No	Yes	-
School well-being	No	Yes	-	No	Yes	-
Personality traits	No	Yes	-	No	Yes	-
Individual FE	No	No	Yes	No	No	Yes

Notes: This table presents the results for six separate difference-in-difference specifications comparing camp participants to non-participants. Columns 1 to 3 display the simple difference-in-difference specifications, and columns 4 to 6 show the results from the difference-in-difference combined with entropy balancing. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) numbers indicate significance at the 5% (10%) level.

Table A.6: Effects of summer camp on desire for upper secondary education

	3-year HS	Voc. training	2-year HS	Other
	(1)	(2)	(3)	(4)
Camp				
Post	**0.022**	**0.058**	**0.014**	**-0.028**
	(0.005)	(0.005)	(0.002)	(0.004)
Camp X Post	-0.048	0.033	0.002	0.001
	(0.056)	(0.048)	(0.024)	(0.031)
Observations	280,656	280,656	280,656	280,656
R-squared	0.819	0.814	0.775	0.766
Mean outcome, grade 9	0.456	0.759	0.034	0.063
Pupil background chars	-	-	-	-
Parental background chars	-	-	-	-
Reading abilities	-	-	-	-
Math abilities	-	-	-	-
Absence information	-	-	-	-
School well-being	-	-	-	-
Personality traits	-	-	-	-
Individual FE	Yes	Yes	Yes	Yes

Notes: This table shows difference-in-difference combined with entropy balancing results for secondary educational wishes at the ERA. It is possible for a pupil to wish more than one education and thus be assessed for all of them. Column 1 shows the results for 3-year high school, column 2 for vocational training, column 3 for 2-year high school, and column 4 for other types of education. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) numbers indicate significance at the 5% (10%) level.

Table A.7: Effects of summer camp participation - Alternative comparison groups

	Overall ERA			GPA		
	All	Camp schools	Non-camp schools	All	Camp schools	Non-camp schools
	(1)	(2)	(3)	(4)	(5)	(6)
Post	**0.193**	**0.192**	**0.193**	**0.014**	**0.028**	**0.010**
	(0.006)	(0.015)	(0.006)	(0.005)	(0.013)	(0.005)
Camp X Post	**0.176**	**0.177**	**0.176**	**0.158**	**0.144**	**0.162**
	(0.060)	(0.062)	(0.060)	(0.037)	(0.039)	(0.037)
Observations	277,043	54,313	223,155	479,540	95,353	384,901
R-squared	0.763	0.720	0.758	0.910	0.868	0.908
Mean outcome, grade 9	0.522	0.494	0.363	-0.966	-1.076	-1.061
Pupil background chars	-	-	-	-	-	-
Parental background chars	-	-	-	-	-	-
Reading abilities	-	-	-	-	-	-
Math abilities	-	-	-	-	-	-
Absence information	-	-	-	-	-	-
School well-being	-	-	-	-	-	-
Personality traits	-	-	-	-	-	-
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows difference-in-difference combined with entropy balancing specifications for overall ERA (Columns 1 to 3) and GPA (columns 4 to 6). Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) numbers indicate significance at the 5% (10%) level.

Table A.8: Effects of summer camp participation - Robustness of balancing specification

	Entropy-I	Entropy-II	Entropy-III	NN-3	NN-5	NN-10
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Overall ERA						
Post	**0.193**	**0.210**	**0.210**	**0.195**	**0.202**	**0.210**
	(0.006)	(0.007)	(0.007)	(0.033)	(0.025)	(0.018)
Camp X Post	**0.176**	**0.160**	**0.160**	**0.174**	**0.167**	**0.159**
	(0.060)	(0.060)	(0.061)	(0.069)	(0.065)	(0.063)
Observations	277,043	277,043	277,042	1,679	2,512	4,507
R-squared	0.763	0.753	0.755	0.752	0.754	0.752
Panel B: GPA						
Post	**0.014**	*0.011*	0.009	0.029	**0.032**	0.013
	(0.005)	(0.006)	(0.007)	(0.021)	(0.016)	(0.012)
Camp X Post	**0.158**	**0.162**	**0.163**	**0.143**	**0.140**	**0.160**
	(0.037)	(0.037)	(0.037)	(0.043)	(0.040)	(0.038)
Observations	479,540	479,540	479,537	2,765	4,122	7,389
R-squared	0.910	0.896	0.894	0.905	0.904	0.904
Pupil background	-	-	-	-	-	-
Parental background	-	-	-	-	-	-
Reading abilities	-	-	-	-	-	-
Math abilities	-	-	-	-	-	-
Absence information	-	-	-	-	-	-
School well-being	-	-	-	-	-	-
Personality traits	-	-	-	-	-	-
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows difference-in-difference combined with different balancing specifications for overall ERA (panel A) and GPA (panel B). Column 1 presents the main specification, column 2 (3) exploits entropy balance to ensure balance up to the second (third) moment. Column 4 to 6 exploits propensity score matching with respectively 3,5, and 10 nearest neighbour. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) numbers indicate significance at the 5% (10%) level.

Table A.9: Effect of being in placebo treatment group on main outcomes

	Difference-in-difference			DiD & Entropy balance		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Overall ERA						
Camp	-0.001	-0.001	0.000	0.001	-0.001	0.000
	(0.002)	(0.002)	(.)	(0.002)	(0.002)	(.)
Post	**0.148**	**0.140**	**0.140**	**0.149**	**0.140**	**0.140**
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Camp X Post	-0.003	-0.002	-0.001	-0.003	-0.002	-0.001
	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.004)
Observations	276,618	276,618	276,618	276,618	276,618	276,618
R-squared	0.025	0.352	0.820	0.025	0.352	0.820
Panel B: GPA						
Camp	-0.002	-0.000	0.000	0.003	-0.000	0.000
	(0.005)	(0.003)	(.)	(0.005)	(0.003)	(.)
Post	**0.027**	**0.019**	**0.020**	**0.028**	**0.019**	**0.020**
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Camp X Post	-0.004	-0.002	0.001	-0.004	-0.002	0.001
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
Observations	478,826	478,826	478,826	478,826	478,826	478,826
R-squared	0.000	0.623	0.944	0.000	0.623	0.944
Pupil background chars	No	Yes	-	No	Yes	-
Parental background chars	No	Yes	-	No	Yes	-
Reading abilities	No	Yes	-	No	Yes	-
Math abilities	No	Yes	-	No	Yes	-
Absence information	No	Yes	-	No	Yes	-
School well-being	No	Yes	-	No	Yes	-
Personality traits	No	Yes	-	No	Yes	-
Individual FE	No	No	Yes	No	No	Yes

Notes: This table shows the main results for a randomly selected placebo treatment group. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) numbers indicate significance at the 5% (10%) level.

Table A.10: Placebo test - Effect of "camp" on national reading test in grade 6

	Language Comprehension			Decoding			Text Comprehension		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				Difference-in-difference					
Camp	**-0.680**	**-0.472**		**-0.947**	**-0.616**		**-0.841**	**-0.596**	
	(0.070)	(0.072)		(0.074)	(0.070)		(0.068)	(0.072)	
Post	0.001	**0.011**	**0.017**	**-0.042**	**-0.031**	**-0.024**	**0.016**	**0.027**	**0.033**
	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
Camp X Post	0.040	0.028	0.052	0.059	0.043	0.071	0.069	0.056	0.078
	(0.076)	(0.075)	(0.111)	(0.072)	(0.071)	(0.101)	(0.080)	(0.080)	(0.116)
Observations	255,424	255,424	255,424	255,424	255,424	255,424	255,424	255,424	255,424
R-squared	0.001	0.164	0.779	0.002	0.221	0.865	0.001	0.186	0.833
				Difference-in-difference & Entropy balance					
Camp	**-0.473**	**-0.463**		**-0.642**	**-0.621**		**-0.626**	**-0.614**	
	(0.071)	(0.066)		(0.074)	(0.065)		(0.068)	(0.064)	
Post	**0.040**	**0.049**	**0.051**	**-0.021**	*-0.010*	-0.004	**0.017**	**0.024**	**0.029**
	(0.006)	(0.007)	(0.009)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.008)
Camp X Post	0.000	-0.002	0.018	0.037	0.025	0.051	0.068	0.066	0.083
	(0.076)	(0.075)	(0.111)	(0.072)	(0.070)	(0.101)	(0.081)	(0.079)	(0.116)
Observations	255,424	255,424	255,424	255,424	255,424	255,424	255,424	255,424	255,424
R-squared	0.053	0.152	0.761	0.084	0.272	0.841	0.085	0.191	0.769
Covariates	No	Yes	-	No	Yes	-	No	Yes	-
Individual FE	No	No	Yes	No	No	Yes	No	No	Yes

Notes: This table shows placebo test on national tests in reading using camp vs non-camp pupils and the placebo camp date to be in the summer between grade 5 and 6. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) numbers indicate significance at the 5% (10%) level.

Table A.11: Placebo test - Effect of "camp" on national reading test in grade 8

	Language Comprehension			Decoding			Text Comprehension		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Difference-in-difference								
Camp	**-0.680**	**-0.469**		**-0.947**	**-0.614**		**-0.841**	**-0.594**	
	(0.070)	(0.071)		(0.074)	(0.070)		(0.068)	(0.072)	
Post	**0.035**	**0.054**	**0.057**	**-0.046**	**-0.021**	**-0.017**	0.004	**0.031**	**0.035**
	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
Camp X Post	0.098	0.088	0.146	-0.077	-0.084	-0.030	-0.025	-0.035	0.017
	(0.098)	(0.100)	(0.147)	(0.076)	(0.076)	(0.101)	(0.089)	(0.090)	(0.132)
Observations	250,235	250,235	250,235	250,235	250,235	250,235	250,235	250,235	250,235
R-squared	0.001	0.114	0.703	0.002	0.213	0.838	0.001	0.189	0.819
	Difference-in-difference & Entropy balance								
Camp	**-0.473**	**-0.465**		**-0.642**	**-0.626**		**-0.626**	**-0.619**	
	(0.071)	(0.067)		(0.074)	(0.066)		(0.068)	(0.064)	
Post	**0.064**	**0.076**	**0.077**	**-0.029**	-0.006	-0.008	-0.004	**0.020**	0.013
	(0.007)	(0.010)	(0.010)	(0.006)	(0.008)	(0.008)	(0.007)	(0.008)	(0.009)
Camp X Post	0.069	0.078	0.125	-0.093	-0.081	-0.039	-0.017	0.007	0.039
	(0.099)	(0.097)	(0.147)	(0.077)	(0.072)	(0.102)	(0.089)	(0.086)	(0.132)
Observations	250,235	250,235	250,235	250,235	250,235	250,235	250,235	250,235	250,235
R-squared	0.043	0.116	0.703	0.097	0.294	0.849	0.089	0.195	0.765
Covariates	No	Yes	-	No	Yes	-	No	Yes	-
Individual FE	No	No	Yes	No	No	Yes	No	No	Yes

Notes: This table shows placebo test on national tests in reading using camp vs non-camp pupils and the placebo camp date to be in the summer between grade 5 and 6. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) numbers indicate significance at the 5% (10%) level.

Table A.12: Placebo test - Effect of "camp" on national math test in grade 6

	Numbers and algebra			Geometry			Statistics and probability		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
					Difference-in-difference				
Camp	**-0.734**	**-0.533**		**-0.629**	**-0.449**		**-0.805**	**-0.585**	
	(0.069)	(0.070)		(0.056)	(0.058)		(0.068)	(0.070)	
Post	**0.024**	**0.032**	**0.040**	**-0.012**	-0.003	0.005	0.001	**0.011**	**0.018**
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
Camp X Post	0.009	0.011	0.018	-0.007	-0.007	-0.019	-0.009	-0.009	0.017
	(0.081)	(0.082)	(0.119)	(0.071)	(0.073)	(0.104)	(0.080)	(0.082)	(0.115)
Observations	253,000	253,000	253,000	253,000	253,000	253,000	253,000	253,000	253,000
R-squared	0.001	0.144	0.771	0.001	0.136	0.758	0.001	0.160	0.787
					Difference-in-difference & Entropy balance				
Camp	**-0.558**	**-0.543**		**-0.473**	**-0.466**		**-0.604**	**-0.592**	
	(0.069)	(0.065)		(0.056)	(0.054)		(0.068)	(0.064)	
Post	-0.001	0.010	0.011	**-0.033**	**-0.022**	**-0.022**	**-0.017**	-0.002	-0.004
	(0.006)	(0.008)	(0.009)	(0.006)	(0.007)	(0.008)	(0.006)	(0.008)	(0.009)
Camp X Post	0.034	0.025	0.046	0.014	0.010	0.008	0.009	-0.001	0.039
	(0.081)	(0.081)	(0.119)	(0.071)	(0.072)	(0.104)	(0.080)	(0.080)	(0.115)
Observations	253,000	253,000	253,000	253,000	253,000	253,000	253,000	253,000	253,000
R-squared	0.071	0.181	0.748	0.061	0.155	0.737	0.083	0.194	0.769
Covariates	No	Yes	-	No	Yes	-	No	Yes	-
Individual FE	No	No	Yes	No	No	Yes	No	No	Yes

Notes: This table shows placebo test on national tests in math using camp vs non-camp pupils and the placebo camp date to be in the summer between grade 5 and 6 with test conducted in grade 3 and 6. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) numbers indicate significance at the 5% (10%) level.

Table A.13: Placebo test - Effect of "camp" on absence between grade 7 and 8

	Sick			Illegal			Legal		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				Difference-in-difference					
Camp	0.018	-0.176		0.336	0.260		0.239	0.163	
	(0.310)	(0.278)		(0.276)	(0.253)		(0.159)	(0.157)	
Post	**0.058**	**0.052**	**0.069**	**0.473**	**0.459**	**0.470**	**0.261**	**0.260**	**0.281**
	(0.013)	(0.013)	(0.018)	(0.013)	(0.012)	(0.017)	(0.010)	(0.010)	(0.014)
Camp X Post	-0.259	-0.278	-0.251	0.243	0.236	0.231	0.481	0.482	0.310
	(0.268)	(0.267)	(0.385)	(0.353)	(0.354)	(0.523)	(0.509)	(0.510)	(0.726)
Observations	251,655	251,655	251,655	251,655	251,655	251,655	251,655	251,655	251,655
R-squared	0.000	0.156	0.786	0.003	0.132	0.776	0.002	0.044	0.650 [b]
				Difference-in-difference & Entropy balance					
Camp	-0.164	-0.162		0.310	0.375		0.183	0.194	
	(0.311)	(0.253)		(0.276)	(0.228)		(0.159)	(0.164)	
Post	**0.098**	**0.094**	**0.117**	**0.582**	**0.578**	**0.580**	**0.345**	**0.342**	**0.357**
	(0.025)	(0.025)	(0.036)	(0.030)	(0.030)	(0.043)	(0.021)	(0.021)	(0.029)
Camp X Post	-0.299	-0.315	-0.300	0.133	0.118	0.120	0.397	0.371	0.234
	(0.269)	(0.271)	(0.386)	(0.354)	(0.352)	(0.524)	(0.510)	(0.493)	(0.727)
Observations	251,655	251,655	251,655	251,655	251,655	251,655	251,655	251,655	251,655
R-squared	0.002	0.209	0.792	0.007	0.186	0.737	0.007	0.058	0.587
Pupil background chars	No	Yes	-	No	Yes	-	No	Yes	-
Parental background chars	No	Yes	-	No	Yes	-	No	Yes	-
Reading abilities	No	Yes	-	No	Yes	-	No	Yes	-
Math abilities	No	Yes	-	No	Yes	-	No	Yes	-
Absence information	No	Yes	-	No	Yes	-	No	Yes	-
School well-being	No	Yes	-	No	Yes	-	No	Yes	-
Personality traits	No	Yes	-	No	Yes	-	No	Yes	-
Individual FE	No	No	Yes	No	No	Yes	No	No	Yes

Notes: This table tests for motivation effect by investigating behavior change leading up to camp participation. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) indicates significance at the 5% (10%) level.

Table A.14: Placebo test - Effect of "camp" on personality traits between grade 7 and 8

	Conscientiousness			Agreeableness			Emotional Stability		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				Difference-in-difference					
Camp	**-0.487**	**-0.281**		**-0.252**	-0.098		**-0.173**	-0.069	
	(0.076)	(0.072)		(0.081)	(0.078)		(0.085)	(0.085)	
Post	**-0.023**	**-0.019**	**-0.024**	**0.012**	**0.036**	**0.043**	**-0.013**	**-0.009**	**-0.021**
	(0.003)	(0.003)	(0.005)	(0.004)	(0.004)	(0.006)	(0.003)	(0.003)	(0.005)
Camp X Post	*-0.160*	**-0.194**	-0.141	-0.104	-0.122	-0.230	-0.082	-0.105	-0.128
	(0.084)	(0.082)	(0.133)	(0.096)	(0.094)	(0.154)	(0.100)	(0.100)	(0.173)
Observations	194,129	194,129	194,129	194,344	194,344	194,344	190,467	190,467	190,467
R-squared	0.001	0.114	0.829	0.000	0.069	0.792	0.000	0.071	0.820
				Difference-in-difference & Entropy balance					
Camp	**-0.286**	**-0.288**		-0.094	-0.099		-0.061	-0.099	
	(0.076)	(0.064)		(0.081)	(0.070)		(0.085)	(0.081)	
Post	0.009	0.004	-0.010	**0.042**	0.000	**0.054**	0.012	0.006	-0.012
	(0.007)	(0.018)	(0.010)	(0.008)	(0.020)	(0.012)	(0.007)	(0.021)	(0.011)
Camp X Post	**-0.191**	**-0.186**	-0.155	-0.134	-0.125	-0.240	-0.106	-0.102	-0.137
	(0.084)	(0.078)	(0.133)	(0.096)	(0.091)	(0.155)	(0.100)	(0.100)	(0.174)
Observations	194,129	194,129	194,129	194,344	194,344	194,344	190,467	190,467	190,467
R-squared	0.044	0.191	0.852	0.008	0.128	0.826	0.005	0.103	0.818
Pupil background chars	No	Yes	-	No	Yes	-	No	Yes	-
Parental background chars	No	Yes	-	No	Yes	-	No	Yes	-
Reading abilities	No	Yes	-	No	Yes	-	No	Yes	-
Math abilities	No	Yes	-	No	Yes	-	No	Yes	-
Absence information	No	Yes	-	No	Yes	-	No	Yes	-
School well-being	No	Yes	-	No	Yes	-	No	Yes	-
Personality traits	No	Yes	-	No	Yes	-	No	Yes	-
Individual FE	No	No	Yes	No	No	Yes	No	No	Yes

Notes: This table tests for motivation effect by investigating behavior change leading up to camp participation. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) indicates significance at the 5% (10%) level.

Table A.15: Placebo test - Effect of "camp" on well-being between grade 7 and 8

	School Connectedness			Learning Self-efficacy			Learning Environment			Classroom Management		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
						Difference-in-difference						
Camp	*-0.142*	-0.032		**-0.656**	**-0.411**		-0.063	0.005		**-0.240**	*-0.155*	
	(0.083)	(0.080)		(0.085)	(0.082)		(0.091)	(0.087)		(0.088)	(0.086)	
Post	**-0.047**	**-0.045**	**-0.058**	**-0.066**	**-0.057**	**-0.056**	**-0.121**	**-0.127**	**-0.142**	**0.046**	**0.048**	**0.041**
	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)	(0.005)	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)	(0.006)
Camp X Post	-0.113	-0.136	-0.174	**-0.276**	**-0.295**	-0.194	-0.059	-0.091	-0.051	0.049	0.026	-0.011
	(0.096)	(0.095)	(0.154)	(0.099)	(0.098)	(0.166)	(0.098)	(0.097)	(0.131)	(0.109)	(0.109)	(0.184)
Observations	186,750	186,750	186,750	177,224	177,224	177,224	185,973	185,973	185,973	194,204	194,204	194,204
R-squared	0.001	0.077	0.837	0.002	0.149	0.874	0.004	0.072	0.832	0.001	0.044	0.785
						Difference-in-difference & Entropy balance						
Camp	-0.026	-0.050		**-0.423**	**-0.439**		-0.006	0.001		*-0.153*	**-0.160**	
	(0.083)	(0.075)		(0.086)	(0.073)		(0.091)	(0.074)		(0.088)	(0.078)	
Post	**-0.024**	-0.027	**-0.050**	**-0.036**	-0.016	**-0.046**	**-0.105**	**-0.095**	**-0.133**	**0.063**	0.032	**0.052**
	(0.007)	(0.021)	(0.010)	(0.007)	(0.022)	(0.010)	(0.007)	(0.020)	(0.011)	(0.008)	(0.023)	(0.012)
Camp X Post	-0.136	-0.151	-0.182	**-0.306**	**-0.287**	-0.204	-0.075	-0.103	-0.060	0.031	0.035	-0.022
	(0.096)	(0.093)	(0.155)	(0.099)	(0.092)	(0.167)	(0.098)	(0.090)	(0.131)	(0.109)	(0.104)	(0.184)
Observations	186,750	186,750	186,750	177,224	177,224	177,224	185,973	185,973	185,973	194,204	194,204	194,204
R-squared	0.006	0.113	0.846	0.092	0.233	0.891	0.006	0.128	0.874	0.006	0.090	0.804
Pupil background chars	No	Yes	-	No	Yes	-	No	Yes	-	No	Yes	-
Parental background chars	No	Yes	-	No	Yes	-	No	Yes	-	No	Yes	-
Reading abilities	No	Yes	-	No	Yes	-	No	Yes	-	No	Yes	-
Math abilities	No	Yes	-	No	Yes	-	No	Yes	-	No	Yes	-
Absence information	No	Yes	-	No	Yes	-	No	Yes	-	No	Yes	-
School well-being	No	Yes	-	No	Yes	-	No	Yes	-	No	Yes	-
Personality traits	No	Yes	-	No	Yes	-	No	Yes	-	No	Yes	-
Individual FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes

Notes: This table tests for motivation effect by investigating behavior change leading up to camp participation. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) indicates significance at the 5% (10%) level.

Table A.16: Triple differences on main outcomes - Placebo treatment group

		Educational readiness assessment			
	GPA	Overall	Personal	Social	Academic
	(1)	(2)	(3)	(4)	(5)
DiDiD	0.004	-0.003	-0.002	-0.004	0.002
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Observations	478,826	276,618	263,087	262,760	264,960
R-squared	0.623	0.353	0.252	0.203	0.354
Pupil background chars	Yes	Yes	Yes	Yes	Yes
Parental background chars	Yes	Yes	Yes	Yes	Yes
Reading abilities	Yes	Yes	Yes	Yes	Yes
Math abilities	Yes	Yes	Yes	Yes	Yes
Absence information	Yes	Yes	Yes	Yes	Yes
School well-being	Yes	Yes	Yes	Yes	Yes
Personality traits	Yes	Yes	Yes	Yes	Yes

Notes: This table shows triple difference (DiDiD) estimates by comparing placebo camp participants attending the individual mentoring program with those attending group mentor centers in the year following the camp. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) numbers indicate significance at the 5% (10%) level.

Table A.17: Placebo test - Triple differences estimates of "camp" between grade 5 and 6 on national reading tests

	Language comprehension		Decoding		Text comprehension	
	Grade 6	Grade 8	Grade 6	Grade 8	Grade 6	Grade 8
	(1)	(2)	(3)	(4)	(5)	(6)
DiDiD	0.112	0.056	0.082	-0.196	0.017	-0.074
	(0.151)	(0.198)	(0.140)	(0.149)	(0.160)	(0.175)
Observations	255,424	250,235	255,424	250,235	255,424	250,235
R-squared	0.152	0.116	0.274	0.295	0.193	0.196
Pupil background chars	Yes	Yes	Yes	Yes	Yes	Yes
Parental background chars	Yes	Yes	Yes	Yes	Yes	Yes
Reading abilities	Yes	Yes	Yes	Yes	Yes	Yes
Math abilities	Yes	Yes	Yes	Yes	Yes	Yes
Absence information	Yes	Yes	Yes	Yes	Yes	Yes
School well-being	Yes	Yes	Yes	Yes	Yes	Yes
Personality traits	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows triple difference (DiDiD) estimates by comparing camp participants attending the individual mentoring program with those attending group mentor centers using a fake camp date between grade 5 and 6. I use the national test in reading grade 4 as pre-test and the national test in reading grade 6 and 8 as post-tests. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) numbers indicate significance at the 5% (10%) level.

Table A.18: Placebo test - Triple differences estimates of "camp" between grade 5 and 6 on national math tests

	Numbers and algebra	Geometry	Statistics and probability
	(1)	(2)	(3)
DiDiD	0.039	-0.181	-0.070
	(0.166)	(0.144)	(0.162)
Observations	253,000	253,000	253,000
R-squared	0.181	0.156	0.194
Pupil background chars	Yes	Yes	Yes
Parental background chars	Yes	Yes	Yes
Reading abilities	Yes	Yes	Yes
Math abilities	Yes	Yes	Yes
Absence information	Yes	Yes	Yes
School well-being	Yes	Yes	Yes
Personality traits	Yes	Yes	Yes

Notes: This table shows triple difference (DiDiD) estimates by comparing camp participants attending the individual mentoring program with those attending group mentor centers using a fake camp date between grade 5 and 6. I use the national test in math grade 3 as pre-test and the national test in reading grade 6 as post-test. Standard errors are clustered on the individual level. Missing values are imputed with the value zero and a binary indicator is added to the conditioning set. Bold (italic) numbers indicate significance at the 5% (10%) level.

Table A.19: Descriptive statistics of camp boys by follow-up strategy

	Camp		
	2015-2016	2017-2019	Diff.
	(1)	(2)	(3)
Pupil:			
- Age (years)	15.204	15.383	**0.179**
- Non-western ethnicity (1/0)	0.054	0.039	-0.015
- Living with both parents (1/0)	0.584	0.531	-0.053
- Number of relocations	2.000	2.180	0.180
- Years in private schooling	0.832	0.805	-0.027
- Number of school changes	0.858	0.813	-0.046
- Years in special needs teaching	1.159	0.359	**-0.800**
- Total hours of special needs teaching	7.938	5.082	-2.856
- Dyslexics (1/0)	0.327	0.375	0.048
- OCD (1/0)	0.080	0.063	-0.017
- ADHD (1/0)	0.062	0.125	*0.063*
- Any psychological diagnosis (1/0)	0.186	0.227	0.041
Mother:			
- Age at birth (years)	29.633	29.828	0.195
- Wage income (log)	5.615	5.600	-0.015
- Years of education	14.030	14.208	0.178
- High school or less (1/0)	0.204	0.180	-0.024
- Employed in November (1/0)	0.735	0.711	-0.024
Father:			
- Age at birth (years)	32.308	32.586	0.278
- Wage income (log)	5.877	5.773	-0.104
- Years of education	13.800	13.913	0.113
- High school or less (1/0)	0.257	0.273	0.017
- Employed in November (1/0)	0.805	0.772	-0.034
Language comprehension (std.):			
- Grade 2	-0.522	-0.504	0.018
- Grade 4	-0.654	-0.702	-0.048
- Grade 6	-0.656	-0.624	0.032
- Grade 8	-0.579	-0.518	0.061
Decoding (std.):			
- Grade 2	-1.011	-0.973	0.038
- Grade 4	-1.034	-0.920	0.114
- Grade 6	-1.073	-0.857	*0.216*
- Grade 8	-1.072	-1.115	-0.043
Text comprehension (std.):			
- Grade 2	-1.026	-0.914	0.112
- Grade 4	-0.982	-0.840	0.141
- Grade 6	-0.902	-0.752	0.150
- Grade 8	-0.955	-0.902	0.054
Numbers and algebra (std.):			

- Grade 3	-0.664	-0.727	-0.063
- Grade 6	-0.675	-0.660	0.015
Geometry (std.):			
- Grade 3	-0.637	-0.582	0.055
- Grade 6	-0.563	-0.677	-0.114
Statistics and probability (std.):			
- Grade 3	-0.711	-0.735	-0.024
- Grade 6	-0.722	-0.742	-0.020
Sick absence (percent):			
- Grade 4	2.771	3.328	0.557
- Grade 5	3.459	3.099	-0.359
- Grade 6	3.263	3.201	-0.062
- Grade 7	3.209	3.516	0.307
- Grade 8	3.236	3.116	-0.120
Illegal absence (percent):			
- Grade 4	0.259	0.432	0.173
- Grade 5	0.787	0.486	-0.300
- Grade 6	1.002	0.539	-0.463
- Grade 7	1.422	1.783	0.361
- Grade 8	2.197	2.450	0.253
Legal absence (percent):			
- Grade 4	1.757	1.573	-0.184
- Grade 5	1.253	1.454	0.201
- Grade 6	1.277	1.260	-0.016
- Grade 7	1.493	1.881	0.388
- Grade 8	1.816	2.997	1.181
Conscientiousness (std.):			
- Grade 7	-0.364	-0.457	-0.093
- Grade 8	-0.563	-0.661	-0.098
Agreeableness (std.):			
- Grade 7	-0.496	-0.414	0.083
- Grade 8	-0.313	-0.696	**-0.383**
Emotional stability (std.):			
- Grade 7	0.104	-0.048	-0.153
- Grade 8	-0.024	-0.172	-0.147
School connectedness (std.):			
- Grade 7	0.064	-0.090	-0.154
- Grade 8	-0.110	-0.292	-0.182
Learning self-efficacy (std.):			
- Grade 7	-0.558	-0.641	-0.083
- Grade 8	-0.991	-0.934	0.058
Learning environment (std.):			
- Grade 7	-0.197	-0.256	-0.058
- Grade 8	-0.478	-0.369	0.108
Classroom management (std.):			

- Grade 7	-0.188	-0.239	-0.051
- Grade 8	-0.032	-0.219	-0.187
Number of pupils	113	128	241

Notes: The table shows descriptive statistics of covariates used in the analysis. The mean and difference in means are reported for before and after the introduction of group mentoring (in 2017) for the Camp boys. The table is based on non-missing data. Bold (italic) numbers indicate significance at the 5% (10%) level.

CHAPTER 3

HOW TO COPE WITH DYSLEXIA: THE EFFECTS OF SPECIAL EDUCATION ON ACADEMIC PERFORMANCE, PERSONALITY TRAITS, AND WELL-BEING

Søren Albeck Nielsen
Aarhus University and TrygFonden's Centre for Child Research

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 by 33 percent and the well-being gap by 80 percent.

Keywords: Dyslexia, Special education, Assistive technology, Personality traits, Difference-in-differences

JEL Codes: C21, H52, I21, I28

3.1 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 learning program - Reading Competency Center for Dyslexics - is a specialized dyslexia program for pupils with severe dyslexia in public schools grade 4 to 8. The objective is to enhance academic abilities by focusing on qualifications in general use of assistive technology[1] and training conventional reading and writing. The 10 consecutive weeks camp consist of three key components 1) Small group instruction, 2) Non-cognitive skills training, and 3) Training the use of assistive technology. The one and a half years follow up program consists of one-to-one meetings with the pupil, parents, teachers and the management team at the local school to ensure continued progression.

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 21% to 24% of a standard deviation. 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 one out of the three personality traits investigated in this study with effect size of 15% of a standard deviation for conscientiousness. Moreover, the intervention increases the participants' school well-being by 14% to 17% of a standard deviation (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 by 80%. Secondly, I show

[1] Assistive technology is any device, piece of equipment, or software program that increase, maintain, or improve the functional capabilities of persons with reading disabilities. E.g. text-to-speech software for smartphones or computers.

that the effects on reading scores are persistent over the four post-treatment school years I am able to observe. For learning self-efficacy, the effect increases throughout the post-intervention time-periods with effect sizes of 40% of a standard deviation.

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. Ballis and Heath (2021) exploits a sharp reduction in special education from a state policy. This reduction generated significant reduction in educational attainment, suggesting long-run benefits from special education. 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). Toffalini et al. (2021) finds in a recent meta-study that causal evidence on program specifically targeting pupils with dyslexia is limited. They find 40 randomized control trails related to dyslexia and reading disorder. However, all studies suffers from small sample bias with on average only 20 treated children per study. Additionally, the previous studies only investigate the impact on spelling and not general reading abilities and well-being.

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 3.2 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 3.3 and the empirical strategy in section 3.4. Section 3.5 presents the results and robustness tests. Section 3.6 provides a discussion and interpretations of the findings. The last section concludes.

3.2 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.

3.2.1 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. Thirdly, 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[2] 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

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

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[3] 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.

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).

[3] 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).

3.2.2 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 3.6, I discuss the implication of *treatment as usual* in further details.

3.2.3 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.

3.3 Data and descriptive statistics

The sample consists of 46,348 NOTA pupils in public schooling across all 98 Danish municipalities, of which 513 pupils were enrolled in RCCD at some point in between 2010/2011 and 2018/2019. The low treatment ratio occur because only pupils in the municipality of Aarhus are subjects for treatment. 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 3.1 provides an illustration of the timeline for the outcomes. Finally, I characterize the RCCD pupils against NOTA members and non-dyslexics.

Table 3.1: 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.

3.3.1 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 3.2 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 3.2: 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
National reading test								
- Language Comprehension (std.)	-0.823	(1.175)	-0.697	(1.125)	0.091	(0.935)	**-0.126**	**-0.914**
- Text Comprehension (std.)	-0.862	(1.053)	-0.770	(0.984)	0.101	(0.944)	**-0.093**	**-0.963**
- Decoding (std.)	-1.475	(0.894)	-1.127	(0.917)	0.143	(0.902)	**-0.348**	**-1.618**
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.

3.3.2 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.

Some of the 40 items from the national well-being survey are used in both the personality trait and the school well-being scales. 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 3.2). Especially, for *Learning self-efficacy* and *Conscientiousness* in which RCCD participants scores between 0.45 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 deviations. This indicates that dyslexics do not thrive in school.

3.3.3 Covariates

Table 3.2 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.

3.4 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 municipality centrally handles visitation to RCCD based on a recommendation from the local school and the municipalities reading consultants in agreement with the pupil's parents. An obvious worry is that municipality select the pupils most severely affected by dyslexia conditional on other special education activities having failed. Exploiting only dyslexics within the municipality of Aarhus as comparison

group generates a selection problem leading to a downward bias of the treatment effects. Therefore, I utilize information for all of the Danish municipalities to account for this selection problem within the municipality of Aarhus. Access to administrative data on, especially, school performance, legal absences and special needs teaching allows me to identify struggling untreated pupils in other municipalities with similar unresponsive behavior to general special education. Table 3.2 show that parents influence the likelihood of being treated, i.e. pupils of high educated parents is more likely to participated. Falling to account for parents abilities positively correlates with RCCD participation would likely create a upward bias of the treatment effects.

The analyses begin with the following event study equation:

$$y_{it} = u_i + \sum_{j=-5, j\neq 0}^{4} \delta_j \cdot 1(j = t)_{it} \cdot 1(RCCD = 1)_i + \sum_{j=-5, j\neq 0}^{4} \gamma_j \cdot 1(j = t)_{it} + \epsilon_{it} \quad (3.1)$$

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 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 as outlined in Bertrand et al. (2004).

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 3.5 below, I also report the results when using alternative comparison groups and discuss the parallel trend assumption.

3.4.1 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. The RCCD participation grade is likely to affect the selection process and thus entropy balancing is conducted within each grade. 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.[4] In section 3.5.3 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.

[4] 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.

3.5 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.

3.5.1 Event study representation

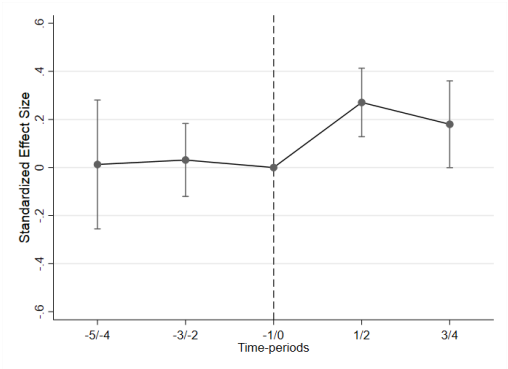
I start by analyzing the effects of RCCD using a event study representation. Figure 3.1 presents the ATT at each time-period, with the time-periods leading up to RCCD (t=0) as baseline for the three profile areas of the national reading test. The national reading test takes place every second school year starting at grade 2 and ending at grade 8 as illustrated in table 3.1. 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-treatment trends, i.e. the differences are not significantly different from zero. Indicating similar development in reading performance prior to RCCD participation across the two groups. Both the language comprehension (a) and text comprehension (b) graphs show large immediate treatment effects of 24% and 20% of a standard deviation 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 (c), 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 and not focusing on their decoding issues caused by their diagnose. Therefore, decoding works as a placebo test.

Figure 3.2 and 3.3 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 3.3 and table 3.1. 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.

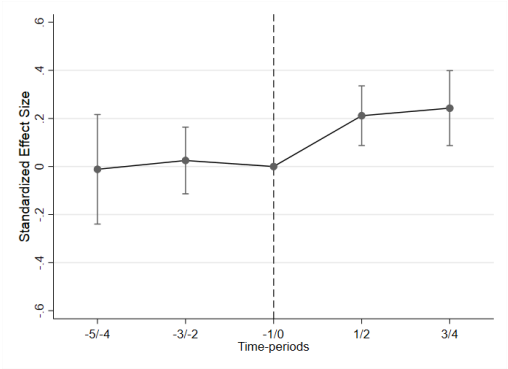
In figure 3.2 (a) the RCCD pupils have a higher level of conscientiousness prior to RCCD. However, the difference is not significantly different from zero and their

Figure 3.1: Event study on Reading performance

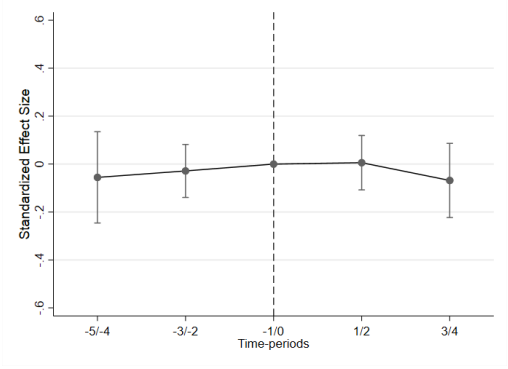
(a) Language Comprehension



(b) Text Comprehension



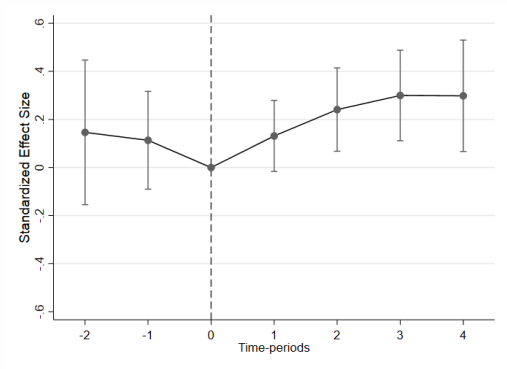
(c) Decoding



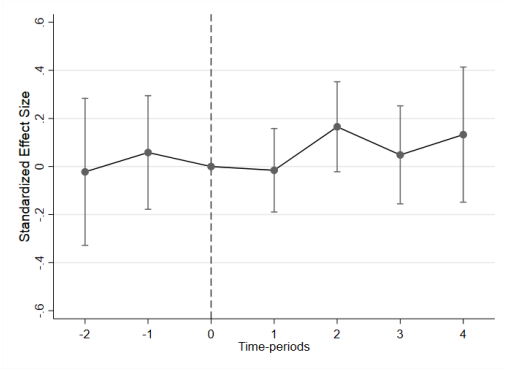
Notes: The figure presents time-varying treatment effects from the main pupils fixed effect model combined with entropy balancing. Time-period, t=1, is the school year of RCCD participation. The vertical lines show the 95% confidence interval.

Figure 3.2: Event study on Personality traits

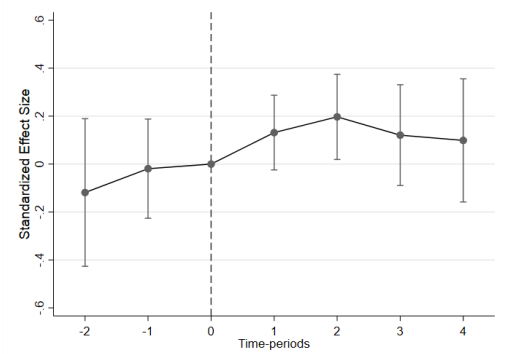
(a) Conscientiousness



(b) Agreeableness



(c) Emotional Stability

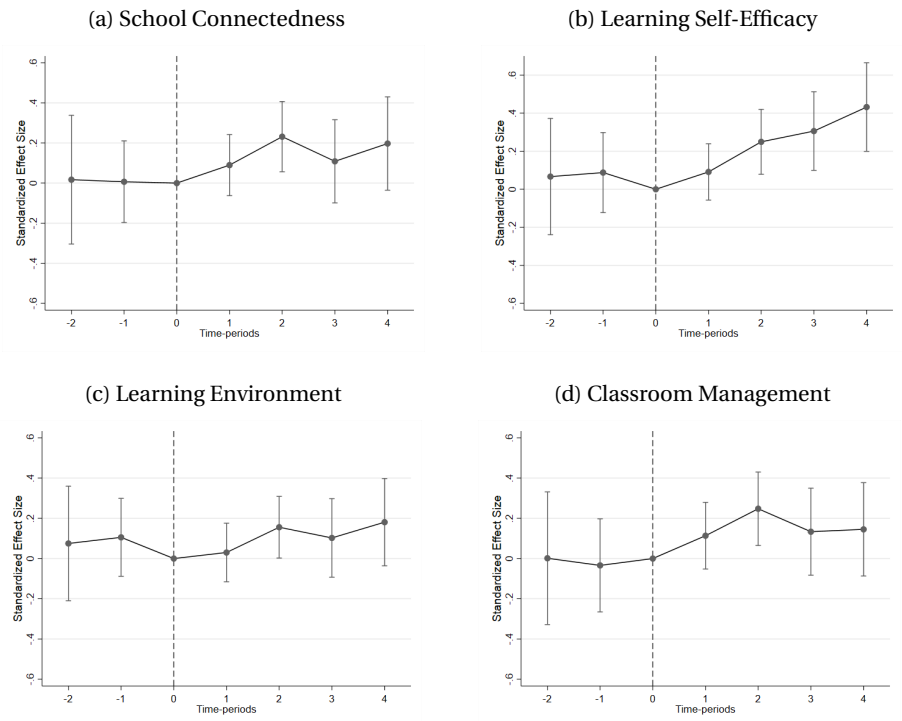


Notes: The figure presents time-varying treatment effects from the main pupils fixed effect model combined with entropy balancing. Time-period, t=1, is the school year of RCCD participation. The vertical lines show the 95% confidence interval.

slope is downwards trending. For agreeableness (b) the two groups' pre-treatment difference is smaller and insignificant. For emotional stability (c) the pre-trends are insignificant, however, upward sloping. This is a potential threat to the research design because it questions that the treatment effects is solely due to RCCD participation. Thus, conclusions on RCCD effects on emotional stability must be carefully interpret. 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 deviation. 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. Notably, the effects are only borderline significantly different from the pre-trends, indicating the treatment effect is an upper bound. 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. The pre-tends makes it difficult to believe the effect occur solely due to RCCD. Thus, it is hard to conclude that RCCD significantly increases emotional stability. Interestingly, however, the fade in the effects occurs in the time-period where the pupils are on their own and no longer in contact with the RCCD consultants.

Next, in figure 3.3 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 3.3 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. 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 deviation. 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 deviation. Similar to conscientiousness, I observe a slight decrease in the RCCD pupils learning self-efficacy in the time up to the intervention start. However, in the year after RCCD participation the effects are significantly different from their pre-RCCD counterparts. 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 show borderline significant effect the year after RCCD, and this effect seems to remain. However, there is larger, though insignificant pre-trends, and therefore, this effect must be interpreted carefully. Finally, in the lower right

Figure 3.3: 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 RCCD participation. The vertical lines show the 95% confidence interval.

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. The pre-RCCD trends are flat indicating similar feeling of their classroom in the years up to RCCD.

3.5.2 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 3.3-3.5 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[5] 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 3.3 presents the effects of RCCD on reading abilities. Panel A shows the results on language comprehension, panel B shows results on text comprehension, and panel C the results on decoding. 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 0.03 SD for language comprehension. Thus, pupils enrolled in RCCD improves their language comprehension by 0.24 SD and text comprehension by 0.21 SD.

In table 3.4, 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 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 0.17 SD on emotional stability. Importantly, the effects on emotional stability should be carefully interpreted due to the pre-trend illustrated previously. For agreeableness, the estimates are positive but statistically insignificant.

The effects on school well-being are presented in table 3.5. 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

[5]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 3.3: 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**	**0.272**	**0.237**	**0.237**	**0.245**
	(0.061)	(0.060)	(0.060)	(0.062)	(0.061)	(0.061)
R-squared	0.004	0.023	0.023	0.021	0.029	0.032
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: Text Comprehension						
ATT	**0.278**	**0.277**	**0.278**	**0.214**	**0.212**	**0.216**
	(0.054)	(0.054)	(0.054)	(0.055)	(0.054)	(0.054)
R-squared	0.010	0.013	0.013	0.065	0.075	0.076
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: Decoding						
ATT	0.037	0.034	0.031	0.002	-0.001	0.002
	(0.051)	(0.051)	(0.051)	(0.052)	(0.051)	(0.051)
R-squared	0.001	0.013	0.013	0.003	0.017	0.020
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

panel D the effects on classroom management. Similar to table 3.4, 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 0.17 SD on classroom management.

3.5.3 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

Table 3.4: Main effects on Personality traits

	Fixed Effect (FE)			FE & Entropy Balance		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Conscientiousness						
ATT	**0.166**	**0.162**	**0.167**	**0.148**	**0.150**	**0.158**
	(0.066)	(0.065)	(0.065)	(0.066)	(0.066)	(0.066)
R-squared	0.002	0.005	0.008	0.013	0.018	0.021
Number of observations	99,528	99,528	99,528	99,528	99,528	99,528
Number of individuals	37,101	37,101	37,101	37,101	37,101	37,101
Panel B: Agreeableness						
ATT	0.046	0.048	0.048	0.034	0.040	0.047
	(0.075)	(0.075)	(0.075)	(0.076)	(0.075)	(0.077)
R-squared	0.001	0.002	0.002	0.009	0.016	0.018
Number of observations	100,504	100,504	100,504	100,504	100,504	100,504
Number of individuals	37,125	37,125	37,125	37,125	37,125	37,125
Panel C: Emotional Stability						
ATT	**0.146**	**0.143**	**0.151**	**0.160**	**0.163**	**0.170**
	(0.065)	(0.066)	(0.065)	(0.067)	(0.067)	(0.067)
R-squared	0.001	0.002	0.005	0.006	0.008	0.013
Number of observations	96,376	96,376	96,376	96,376	96,376	96,376
Number of individuals	36,772	36,772	36,772	36,772	36,772	36,772
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

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.[6] 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) Legal absence

[6] Absence is measured as percentage of school year with one of the three absence types.

Table 3.5: Main effects on School Well-being

	Fixed Effect (FE)			FE & Entropy Balance		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: School Connectedness						
ATT	**0.132**	**0.131**	**0.138**	**0.134**	**0.137**	**0.150**
	(0.065)	(0.066)	(0.065)	(0.067)	(0.067)	(0.065)
R-squared	0.004	0.006	0.009	0.006	0.007	0.016
Number of observations	92,949	92,949	92,949	92,949	92,949	92,949
Number of individuals	36,237	36,237	36,237	36,237	36,237	36,237
Panel B: Learning Self-Efficacy						
ATT	**0.149**	**0.151**	**0.155**	**0.141**	**0.146**	**0.153**
	(0.067)	(0.067)	(0.066)	(0.067)	(0.067)	(0.066)
R-squared	0.003	0.008	0.012	0.007	0.018	0.023
Number of observations	86,281	86,281	86,281	86,281	86,281	86,281
Number of individuals	35,354	35,354	35,354	35,354	35,354	35,354
Panel C: Learning Environment						
ATT	0.072	0.076	0.081	0.051	0.051	0.061
	(0.064)	(0.063)	(0.063)	(0.064)	(0.063)	(0.063)
R-squared	0.078	0.082	0.084	0.091	0.103	0.107
Number of observations	93,164	93,164	93,164	93,164	93,164	93,164
Number of individuals	36,197	36,197	36,197	36,197	36,197	36,197
Panel D: Classroom Management						
ATT	**0.187**	**0.188**	**0.194**	**0.174**	**0.168**	**0.171**
	(0.073)	(0.074)	(0.074)	(0.075)	(0.074)	(0.074)
R-squared	0.001	0.002	0.004	0.009	0.018	0.020
Number of observations	100,133	100,133	100,133	100,133	100,133	100,133
Number of individuals	37,148	37,148	37,148	37,148	37,148	37,148
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 clustered at the individual level

- absence with permission from the school principal, b) Sick absence, and c) Illegal absence – absence without 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. 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 between 0.03 to 0.14 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 3.6.

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.

3.5.4 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 3.4 and 3.5. 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 small difference across gender on text comprehension with effects larger for girls. However, for language comprehension the effect is of similar size, however, for boys. When I observe personality traits it is clear that the main effect on conscientiousness is driven by the girls (effect size of 0.23). However, for emotional stability, I observe larger effect for the boys. 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 0.27, 0.16 and 0.16). 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.

3.6 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.

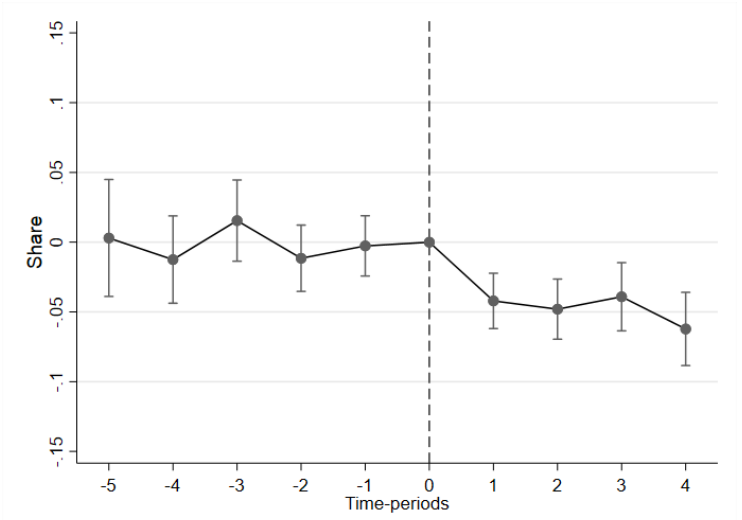
3.6.1 Economic significance

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 18,000USD 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.

Figure 3.4: Event study on special needs teaching



Notes: The figure presents time-varying treatment effects on the 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 they receive minimum 9 hours of special needs teaching per week. Time-period, t=1, is the school year of RCCD participation. The vertical lines show the 95% confidence interval.

Despite the high costs, investments in dyslexic learning program remain policy relevant because of the large effects for this extremely disadvantaged group. Figure 3.4 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.

3.6.2 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 3.2. 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*.

3.7 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 24% and 21% of a standard deviation. They, thus, perform well beyond other dyslexics and have closed the gap to non-dyslexics by 33%. In addition, RCCD participants also show a significant increase in one out of three measured personality traits and three out of four school well-being measures with effect sizes between 14% and 17% of an standard deviation. Significantly closing the gap to non-dyslexics. 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 seems 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.

Acknowledgements

I gratefully acknowledge helpful comments from Marianne Simonsen, Michael Rosholm, Helena Skyt Nielsen, Lars Skipper, Kristoffer Holst Ibsen, Bastien Michel, Louise Beuchert, participants in the EALE conference, and seminar participants at Aarhus University and the Norwegian School of Economics.

3.8 References

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

Alan, S., Boneva, T., Ertac, S., 2019. Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. The Quarterly Journal of Economics 134 (3), 1121–1162, publisher: Oxford University Press.

Alan, S., Ertac, S., 2018. Fostering patience in the classroom: Results from randomized educational intervention. Journal of Political Economy 126 (5), 1865–1911, publisher: University of Chicago Press Chicago, IL.

Andersen, S. C., Gensowski, M., Ludeke, S. G., John, O. P., Oct. 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 (5), 925–939.
URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jopy.12538

Andersen, S. C., Humlum, M. K., Nandrup, A. B., Jul. 2016. Increasing instruction time in school does increase learning. Proceedings of the National Academy of Sciences 113 (27), 7481–7484.
URL http://www.pnas.org/lookup/doi/10.1073/pnas.1516686113

Ballis, B., Heath, K., Nov. 2021. The Long-Run Impacts of Special Education. American Economic Journal: Economic Policy 13 (4), 72–111.
URL https://pubs.aeaweb.org/doi/10.1257/pol.20190603

Bertrand, M., Duflo, E., Mullainathan, S., 2004. How much should we trust differences-in-differences estimates? Quarterly journal of economics, 27.

Beuchert, L., Nandrup, A., 2017. The Danish National Tests at a Glance. SSRN Electronic Journal.
URL https://www.ssrn.com/abstract=2996431

Blundell, R., Dias, M. C., Meghir, C., van Reenen, J., Jun. 2004. Evaluating the Employment Impact of a Mandatory Job Search Program. Journal of the European Economic Association 2 (4), 569–606.
URL https://academic.oup.com/jeea/article/2280805/Evaluating

Dietrichson, J., Filges, T., Klokker, R. H., Nielsen, B. C. V., Bøg, M., Jensen, U. H., 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 (2), e1081.

Dietrichson, J., Filges, T., Seerup, J. K., Klokker, R. H., Viinholt, B. A., Bøg, M., Eiberg, M., 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), 1–60.

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

Einar, H., Anders, L., Jim, S., Kenneth, H., Jan. 2001. Behaviour problems in children with dyslexia. Nordic Journal of Psychiatry 55 (4), 251–256.
URL https://doi.org/10.1080/080394801681019101

Epnion, Nov. 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 (1), 25–46.
URL https://www.cambridge.org/core/product/identifier/S1047198700012997/type/journal_article

Hainmueller, J., Xu, Y., Sep. 2013. ebalance: A Stata Package for Entropy Balancing. Journal of Statistical Software 54 (1), 1–18.
URL https://www.jstatsoft.org/index.php/jss/article/view/v054i07

Hanushek, E. A., Kain, J. F., Rivkin, S. G., Nov. 2002. Inferring Program Effects for Special Populations: Does Special Education Raise Achievement for Students with Disabilities? Review of Economics and Statistics 84 (4), 584–599.
URL http://www.mitpressjournals.org/doi/10.1162/003465302760556431

Haßler, B., Major, L., Hennessy, S., 2016. Tablet use in schools: a critical review of the evidence for learning outcomes. Journal of Computer Assisted Learning 32 (2), 139–156.
URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jcal.12123

Hebert, M., Kearns, D. M., Hayes, J. B., Bazis, P., Cooper, S., Oct. 2018. Why Children With Dyslexia Struggle With Writing and How to Help Them. Language, Speech, and Hearing Services in Schools 49 (4), 843–863.
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6430506/

Heckman, J. J., Ichimura, H., Todd, P. E., Oct. 1997. Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. The

Review of Economic Studies 64 (4), 605–654.
URL https://academic.oup.com/restud/article-lookup/doi/10.2307/2971733

Hurwitz, S., Perry, B., Cohen, E. D., Skiba, R., Apr. 2020. Special Education and Individualized Academic Growth: A Longitudinal Assessment of Outcomes for Students With Disabilities. American Educational Research Journal 57 (2), 576–611, publisher: American Educational Research Association.
URL https://doi.org/10.3102/0002831219857054

Jones, D., Molitor, D., Reif, J., Nov. 2019. What do Workplace Wellness Programs do? Evidence from the Illinois Workplace Wellness Study*. The Quarterly Journal of Economics 134 (4), 1747–1791.
URL https://academic.oup.com/qje/article/134/4/1747/5550759

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

Kraft, M. A., May 2020. Interpreting Effect Sizes of Education Interventions. Educational Researcher 49 (4), 241–253.
URL http://journals.sagepub.com/doi/10.3102/0013189X20912798

Lindeblad, E., Nilsson, S., Gustafson, S., Svensson, I., Oct. 2017. Assistive technology as reading interventions for children with reading impairments with a one-year follow-up. Disability and Rehabilitation. Assistive Technology 12 (7), 713–724.

Morgan, P. L., Frisco, M., Farkas, G., Hibel, J., Feb. 2010. A Propensity Score Matching Analysis of the Effects of Special Education Services. The Journal of special education 43 (4), 236–254.
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3630079/

Niclasen, J., Keilow, M., Obel, C., Dec. 2018. Psychometric properties of the Danish student well-being questionnaire assessed in >250,000 student responders. Scandinavian Journal of Public Health 46 (8), 877–885.
URL http://journals.sagepub.com/doi/10.1177/1403494818772645

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

Nørgaard, E., Schøler Kollin, M., Panduro, B., 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.
URL https://www.vive.dk/media/pure/7029/2300521

Perelmutter, B., McGregor, K. K., Gordon, K. R., Nov. 2017. Assistive technology interventions for adolescents and adults with learning disabilities: An evidence-based systematic review and meta-analysis. Computers & Education 114, 139–163.
URL https://linkinghub.elsevier.com/retrieve/pii/S0360131517301410

Reynolds, A. J., Wolfe, B., Sep. 1999. Special Education and School Achievement: An Exploratory Analysis With a Central-City Sample. Educational Evaluation and Policy Analysis 21 (3).
URL https://doi.org/10.3102/01623737021003249

Rosenbaum, P. R., Rubin, D. B., Apr. 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70 (1), 41–55.
URL https://academic.oup.com/biomet/article/70/1/41/240879

Rosholm, M., Paul, A., Bleses, D., Højen, A., S. Dale, P., Jensen, P., M. Justice, L., Svarer, M., Calmar Andersen, S., Feb. 2021. Are impacts of early interventions in the scandinavian welfare state consistent with a heckman curve? a meta-analysus. Journal of Economic Surveys 35 (1), 106–140.
URL https://onlinelibrary.wiley.com/doi/10.1111/joes.12400

Ryan, A. M., Kontopantelis, E., Linden, A., Burgess, J. F., 2018. Now trending: Coping with non-parallel trends in difference-in-differences analysis. Statistical Methods in Medical Research.
URL https://www.research.manchester.ac.uk/portal/en/publications/now-trending(89a62c4f-c5ef-4502-af22-fb39907b4d5d).html

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

Schwartz, A. E., Hopkins, B. G., Stiefel, L., Jan. 2021. The Effects of Special Education on the Academic Performance of Students with Learning Disabilities. Journal of Policy Analysis and Management, 41.
URL https://onlinelibrary.wiley.com/doi/10.1002/pam.22282

Snowling, M. J., Jan. 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 (1), 7–14.
URL http://doi.wiley.com/10.1111/j.1471-3802.2012.01262.x

Snowling, M. J., Hulme, C., Nation, K., Jul. 2020. Defining and understanding dyslexia: past, present and future. Oxford Review of Education 46 (4).
URL https://doi.org/10.1080/03054985.2020.1765756

Svendsen, H. B., Jan. 2017. Et didaktisk spændingsfelt. Learning Tech (2), 110–136.
URL https://tidsskrift.dk/learningtech/article/view/107732

Toffalini, E., Giofrè, D., Pastore, M., Carretti, B., Fraccadori, F., Szűcs, D., Oct. 2021. Dyslexia treatment studies: A systematic review and suggestions on testing treatment efficacy with small effects and small samples. Behavior Research Methods 53 (5), 1954–1972.
URL https://link.springer.com/10.3758/s13428-021-01549-x

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 (4).
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/dys.384

Westfall, P. H., Young, S. S., Wright, S. P., Sep. 1993. On Adjusting P-Values for Multiplicity. Biometrics 49 (3), 941.
URL https://www.jstor.org/stable/2532216?origin=crossref

A.1 Appendix

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

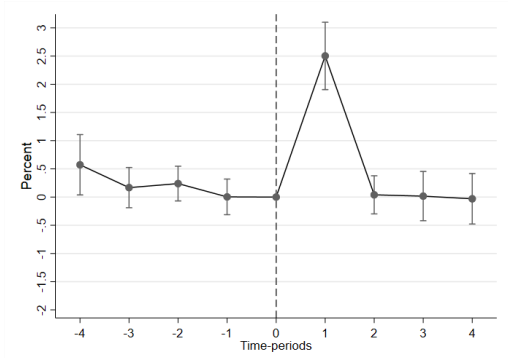
- Academy higher education (1/0)	0.109	(0.312)	0.063	(0.244)	**0.046**	0.109	(0.312)	0.000
- College (1/0)	0.127	(0.333)	0.077	(0.266)	**0.050**	0.127	(0.333)	0.000
- University (1/0)	0.113	(0.317)	0.056	(0.230)	**0.057**	0.113	(0.317)	0.000
Birth cohort								
- 1996	0.014	(0.116)	0.016	(0.125)	-0.002	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
- Text Comprehension (std.)	0.362	(1.226)	0.096	(1.094)	**0.266**	0.362	(1.110)	0.000
- Decoding (std.)	-0.030	(0.966)	-0.024	(0.907)	-0.006	-0.030	(0.901)	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					

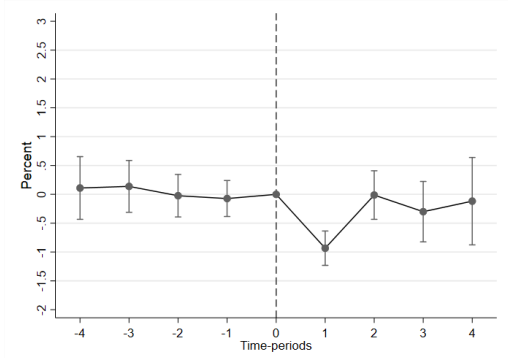
Notes: The table shows the descriptive statistics of the variables used in the balancing method. The mean, standard derivation and differences in means of the covariates are reported for the treatment group, for the comparison group before balancing, and for the comparison group after balancing. The comparison group is reweighted such that it mimics the treatment group within each grade. 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

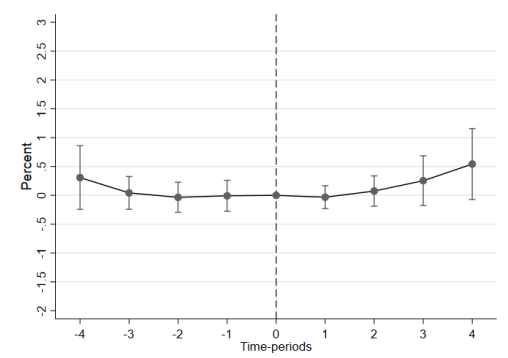
(a) Legal absence



(b) Sick absence



(c) Illegal 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	Text Comprehension	Decoding
	(1)	(2)	(3)
Panel A: Member of NOTA			
ATT	**0.237**	**0.212**	-0.001
	(0.061)	(0.054)	(0.051)
R-squared	0.029	0.075	0.017
Number of observations	134,171	134,171	134,171
Number of individuals	45,901	45,901	45,901
Panel B: Tested dyslexic			
ATT	**0.241**	**0.248**	0.028
	(0.062)	(0.055)	(0.052)
R-squared	0.031	0.065	0.017
Number of observations	97,419	97,419	97,419
Number of individuals	30,754	30,754	30,754
Panel C: All			
ATT	**0.300**	**0.351**	0.023
	(0.060)	(0.053)	(0.050)
R-squared	0.026	0.054	0.013
Number of observations	1,383,547	1,383,547	1,383,547
Number of individuals	474,743	474,743	474,743
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	**0.150**	0.040	**0.163**
	(0.066)	(0.075)	(0.067)
R-squared	0.018	0.016	0.008
Number of observations	99,528	100,504	96,376
Number of individuals	37,101	37,125	36,772
Panel B: Tested dyslexic			
ATT	**0.161**	0.050	**0.138**
	(0.065)	(0.075)	(0.066)
R-squared	0.017	0.012	0.008
Number of observations	84,914	85,814	82,249
Number of individuals	29,336	29,384	29,168
Panel C: All			
ATT	**0.244**	0.050	**0.198**
	(0.064)	(0.074)	(0.065)
R-squared	0.018	0.012	0.009
Number of observations	983,296	988,837	961,813
Number of individuals	365,726	365,900	363,360
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	**0.137**	**0.146**	0.051	**0.168**
	(0.067)	(0.067)	(0.063)	(0.074)
R-squared	0.007	0.018	0.103	0.018
Number of observations	92,949	86,281	93,164	100,133
Number of individuals	36,237	35,354	36,197	37,148
Panel B: Tested dyslexic				
ATT	*0.122*	**0.139**	0.074	**0.176**
	(0.066)	(0.066)	(0.062)	(0.073)
R-squared	0.007	0.017	0.101	0.017
Number of observations	79,366	73,439	79,540	85,450
Number of individuals	28,824	28,122	28,801	29,399
Panel C: All				
ATT	**0.170**	**0.172**	0.074	**0.182**
	(0.064)	(0.064)	(0.061)	(0.072)
R-squared	0.009	0.016	0.101	0.019
Number of observations	939,512	890,668	942,256	987,313
Number of individuals	360,185	353,904	360,045	366,127
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	**0.237**	**0.266**	**0.250**	**0.269**	**0.252**	**0.261**
	(0.061)	(0.060)	(0.064)	(0.068)	(0.066)	(0.063)
R-squared	0.029	0.032	0.036	0.028	0.029	0.028
Number of observations	134,171	134,171	109,513	5,732	8,293	14,107
Number of individuals	45,901	45,901	36,224	1,934	2,794	4,754
Panel B: Text Comprehension						
ATT	**0.212**	**0.282**	**0.254**	**0.196**	**0.203**	**0.213**
	(0.054)	(0.053)	(0.057)	(0.060)	(0.058)	(0.056)
R-squared	0.075	0.053	0.051	0.079	0.078	0.078
Number of observations	134,171	134,171	109,513	5,732	8,293	14,107
Number of individuals	45,901	45,901	36,224	1,934	2,794	4,754
Panel C: Decoding						
ATT	-0.001	0.032	0.054	-0.005	-0.017	0.009
	(0.051)	(0.050)	(0.054)	(0.057)	(0.054)	(0.053)
R-squared	0.017	0.018	0.018	0.017	0.018	0.015
Number of observations	134,171	134,171	109,513	5,732	8,293	14,107
Number of individuals	45,901	45,901	36,224	1,934	2,794	4,754
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	**0.150**	**0.161**	**0.163**	**0.234**	**0.195**	**0.194**
	(0.066)	(0.065)	(0.065)	(0.080)	(0.074)	(0.070)
R-squared	0.018	0.017	0.019	0.023	0.022	0.020
Number of observations	99,528	99,528	80,913	4,112	6,048	10,387
Number of individuals	37,101	37,101	29,035	1,540	2,240	3,832
Panel B: Agreeableness						
ATT	0.040	0.050	0.051	0.024	0.001	0.026
	(0.075)	(0.074)	(0.074)	(0.087)	(0.082)	(0.079)
R-squared	0.016	0.011	0.011	0.017	0.017	0.014
Number of observations	100,504	100,504	81,725	4,169	6,128	10,536
Number of individuals	37,125	37,125	29,055	1,539	2,238	3,833
Panel C: Emotional Stability						
ATT	**0.163**	**0.142**	**0.144**	**0.190**	**0.181**	**0.172**
	(0.067)	(0.065)	(0.065)	(0.080)	(0.075)	(0.071)
R-squared	0.008	0.007	0.009	0.016	0.013	0.010
Number of observations	96,376	96,376	78,395	3,998	5,891	10,082
Number of individuals	36,772	36,772	28,807	1,534	2,226	3,804
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	**0.137**	*0.127*	*0.127*	*0.145*	*0.139*	*0.133*
	(0.067)	(0.065)	(0.065)	(0.078)	(0.074)	(0.070)
R-squared	0.007	0.006	0.005	0.013	0.011	0.008
Number of observations	92,949	92,949	75,716	3,874	5,714	9,777
Number of individuals	36,237	36,237	28,447	1,511	2,193	3,749
Panel B: Learning Self-Efficacy						
ATT	**0.146**	**0.153**	**0.155**	**0.236**	**0.197**	**0.178**
	(0.067)	(0.066)	(0.066)	(0.084)	(0.077)	(0.071)
R-squared	0.018	0.015	0.016	0.026	0.024	0.021
Number of observations	86,281	86,281	70,415	3,570	5,241	9,016
Number of individuals	35,354	35,354	27,859	1,486	2,160	3,680
Panel C: Learning Environment						
ATT	0.051	0.070	0.071	0.044	0.033	0.061
	(0.063)	(0.062)	(0.062)	(0.076)	(0.071)	(0.067)
R-squared	0.103	0.094	0.092	0.109	0.101	0.103
Number of observations	93,164	93,164	75,862	3,883	5,699	9,756
Number of individuals	36,197	36,197	28,397	1,505	2,180	3,737
Panel D: Classroom Management						
ATT	**0.168**	**0.178**	**0.178**	*0.152*	*0.151*	**0.175**
	(0.074)	(0.072)	(0.072)	(0.087)	(0.082)	(0.078)
R-squared	0.018	0.016	0.017	0.019	0.018	0.016
Number of observations	100,133	100,133	81,348	4,167	6,123	10,495
Number of individuals	37,148	37,148	29,051	1,547	2,245	3,836
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 (1)	Westfall-Young (2)
Reading abilities		
- Language comprehension	0.000	0.000
- Text comprehension	0.000	0.000
- Decoding	0.987	0.990
Personality traits		
- Conscientiousness	0.024	0.042
- Agreeableness	0.594	0.568
- Emotional stability	0.015	0.042
School well-being		
- School connectedness	0.040	0.106
- Learning self-efficacy	0.028	0.106
- Learning environment	0.418	0.422
- Classroom management	0.022	0.106

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	**0.237**	**0.274**	**0.208**	0.273	**0.193**
	(0.061)	(0.083)	(0.089)	(0.184)	(0.081)
R-squared	0.029	0.029	0.041	0.066	0.048
Number of observations	134,171	74,795	59,376	15,174	55,085
Number of individuals	45,901	25,690	20,211	5,091	17,608
Panel B: Text Comprehension					
ATT	**0.212**	**0.207**	**0.274**	0.163	**0.177**
	(0.054)	(0.074)	(0.078)	(0.168)	(0.070)
R-squared	0.075	0.072	0.069	0.094	0.064
Number of observations	134,171	74,795	59,376	15,174	55,085
Number of individuals	45,901	25,690	20,211	5,091	17,608
Panel C: Decoding					
ATT	-0.001	0.076	-0.065	0.155	-0.021
	(0.051)	(0.076)	(0.060)	(0.150)	(0.071)
R-squared	0.017	0.021	0.036	0.046	0.086
Number of observations	134,171	74,795	59,376	15,174	55,085
Number of individuals	45,901	25,690	20,211	5,091	17,608
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 derivation 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	**0.150**	0.074	**0.234**	0.240	0.141
	(0.066)	(0.092)	(0.090)	(0.225)	(0.086)
R-squared	0.018	0.020	0.025	0.039	0.030
Number of observations	99,528	56,618	42,910	11,106	40,078
Number of individuals	37,101	21,117	15,984	4,211	14,453
Panel B: Agreeableness					
ATT	0.040	-0.036	0.136	0.066	0.021
	(0.075)	(0.095)	(0.122)	(0.178)	(0.097)
R-squared	0.016	0.011	0.033	0.026	0.019
Number of observations	100,504	57,048	43,456	11,331	40,588
Number of individuals	37,125	21,095	16,030	4,237	14,486
Panel C: Emotional Stability					
ATT	**0.163**	**0.226**	0.093	-0.114	**0.249**
	(0.067)	(0.092)	(0.098)	(0.163)	(0.085)
R-squared	0.008	0.016	0.007	0.044	0.016
Number of observations	96,376	55,056	41,320	10,700	38,791
Number of individuals	36,772	20,954	15,818	4,185	14,344
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 derivation 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	**0.137**	0.128	0.147	-0.195	**0.196**
	(0.067)	(0.092)	(0.095)	(0.137)	(0.091)
R-squared	0.007	0.010	0.012	0.057	0.014
Number of observations	92,949	53,128	39,821	10,230	37,291
Number of individuals	36,237	20,648	15,589	4,117	14,148
Panel B: Learning Self-Efficacy					
ATT	**0.146**	0.120	*0.148*	0.037	*0.170*
	(0.067)	(0.094)	(0.086)	(0.169)	(0.092)
R-squared	0.018	0.025	0.027	0.065	0.032
Number of observations	86,281	49,060	37,221	9,464	34,368
Number of individuals	35,354	20,104	15,250	3,979	13,831
Panel C: Learning Enviroment					
ATT	0.051	0.031	0.037	-0.104	0.084
	(0.063)	(0.091)	(0.084)	(0.159)	(0.078)
R-squared	0.103	0.081	0.146	0.187	0.120
Number of observations	93,164	53,006	40,158	10,283	37,325
Number of individuals	36,197	20,562	15,635	4,087	14,137
Panel C: Classroom Management					
ATT	**0.168**	0.128	0.171	*0.352*	**0.235**
	(0.074)	(0.099)	(0.111)	(0.188)	(0.088)
R-squared	0.018	0.020	0.030	0.092	0.030
Number of observations	100,133	57,014	43,119	11,200	40,350
Number of individuals	37,148	21,121	16,027	4,226	14,486
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 derivation 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

AARHUS BSS SCHOOL OF BUSINESS AND SOCIAL SCIENCES
AARHUS UNIVERSITY

Declaration of co-authorship

Full name of the PhD student: Søren Albeck Nielsen

This declaration concerns the following article/manuscript:

Title:	An Intensive, School-Based Learning Camp Targeting Academic and Non-Cognitive Skills Evaluated in a Randomized Trial
Authors:	Charlotte Hvidman, Alexander Koch, Julia Nafziger, Michael Rosholm

The article/manuscript is: Published ☐ Accepted ☐ Submitted ☒ In preparation ☐

If published, state full reference:

If accepted or submitted, state journal: Economics of Education Review

Has the article/manuscript previously been used in other PhD or doctoral dissertations?

No ☒ Yes ☐ If yes, give details:

The PhD student has contributed to the elements of this article/manuscript as follows:

A. Has essentially done all the work
B. Major contribution
C. Equal contribution
D. Minor contribution
E. Not relevant

Element	Extent (A-E)
1. Formulation/identification of the scientific problem	D
2. Planning of the experiments/methodology design and development	B
3. Involvement in the experimental work/clinical studies/data collection	B
4. Interpretation of the results	B
5. Writing of the first draft of the manuscript	C
6. Finalization of the manuscript and submission	C

Signatures of the co-authors

Date	Name	Signature
16.11.2021	Charlotte Hvidman	
08.11.2021	Alexander Koch	
08.11.2021	Julia Nafziger	
07.11.2021	Michael Rosholm	

In case of further co-authors please attach appendix

Date: 22/11 2021

Signature of the PhD student