Social emotional learning in the classroom:

Short-run effects from PERSPEKT 2.0

Anna Folke Larsen
The Rockwool Foundation
Intervention Unit

Marianne Simonsen
Department of Economics
and Business Economics
Aarhus University

August 2020

This paper combines a large-scale cluster randomized trial in Danish public schools with survey and register-based data to investigate the short-run effects of a universal social emotional learning intervention, PERSPEKT 2.0. The program was delivered to 4th and 5th graders in Danish elementary schools. While the program was well-liked by teachers and pupils and delivered as intended, PERSPEKT 2.0 did not affect the primary outcome, school social well-being, or any of the secondary outcomes including problem behavior, social emotional learning competencies, and emotional distress, neither for the population as a whole, nor for any pre-defined subgroups.

Keywords: Social emotional learning; randomized trial; school social well-being; problem behavior.

Acknowledgements: We appreciate outstanding research assistance from Mikkel Stahlschmidt. We are grateful to Simon Calmar Andersen and Vibeke Myrup Jensen for their valuable and constructive feedback. The project has been registered with the Social Science Registry: https://www.socialscienceregistry.org/trials/3565, RCT ID: AEARCTR-0003565. The usual disclaimer applies.

1. Introduction

Children spend a large fraction of their time in school, where the main objective traditionally has been to enhance academic skills. There is a growing consensus among educational policy makers and human development researchers about the importance of integrating social and emotional learning with academic learning to improve overall pupil outcomes (Corcoran and Slavin, 2016; Jones and Kahn, 2017). Though this idea is appealing, many countries are still short of evidence-based social emotional learning (SEL) programs to include as part of their suggested elementary school curricula. This paper uses a randomized controlled trial to study such a social emotional learning program, PERSPEKT 2.0, developed for a Danish elementary school context, while combining the experimental data with population-wide register-based data. In contrast to expectations, the estimated effects of PERSPEKT 2.0 are not significantly different from zero for any of the child outcomes studied.

SEL programs are tools for educators to support the development of social emotional skills among pupils in a school setting. According to the Collaborative for Academic, Social and Emotional Learning (CASEL), social and emotional learning refers to the process through which people understand and manage emotions, set and achieve positive goals, appreciate the perspective of others, feel and show empathy for others, establish and maintain positive relationships, and make responsible decisions (CASEL, 2013). CASEL has identified five interrelated sets of cognitive, social, and emotional competencies, some or all of which different SEL programs seek to strengthen: Self-awareness, self-management, social awareness, relationship skills, and responsible decision-making. Various sources of theory and evidence suggest that these social and emotional skills may be related to academic performance because they strengthen the executive control, well-being, ability to cope and engagement of the child (Corcoran and Slavin, 2016). A similar survey of the economics of education literature finds a link between so-called ‘soft skills’, i.e. personality traits not adequately measured by achievement tests, and educational achievement (Koch et al., 2015).

Several papers review the impact of SEL programs on a variety of outcomes, also academic attainment, including recent overviews and meta-analyses (Corcoran and Slavin, 2016; Connolly et al., 2016; Taylor et al., 2017). A much-cited study finds that universal school-based SEL programs, which are both well designed according to best-practice criteria for SELs and well implemented, lead to immediate improvements in cognitive, social and emotional competences (Durlak et al., 2011). They also prompt direct improvements in attitudes about self, others, and school. In turn, this drives

the improvements also found in well-being, positive social behavior, less emotional distress, fewer conduct problems, and increased academic success. Improvements were later found on a smaller sample of studies to last also beyond the intervention period, both in terms of the social and emotional competencies, well-being, and academic attainment; Taylor et al. (2017).

Although there is a growing number of rigorous SEL impact studies, little is known about their effectiveness outside of a US contexts, and sub-group analyses are needed more broadly. This paper provides such evidence by studying the impact of a universal classroom-based and teacher-instructed Danish SEL program, PERSPEKT 2.0, among fourth and fifth graders (10-12-year-olds) in Danish elementary schools. PERSPEKT 2.0 is a SEL program developed for the Danish school context.[1] Through a cluster randomized controlled trial, we analyze the impact of PERSPEKT 2.0 on a set of outcomes similar to those reviewed previously (Payton et al., 2008; Durlak et al., 2011; Taylor et al., 2017). Our outcomes and analysis strategies are described in detail in the protocol by Klejnstrup et al. (2018). Our primary outcome measures school social well-being (henceforth merely "social well-being") and is calculated based on nationally collected indicators of elementary and lower secondary school well-being, developed by the Danish Ministry of Education with the purpose of tracking pupils' well-being. Companion work (Larsen et al., 2020) documents that our primary outcome correlates meaningfully with standard measures of disadvantage at the pupil and parental level; that it exhibits high degrees of persistence over time; and that it is positively associated with academic performance and negatively associated with absence from school. Our secondary outcomes consist of measures of problem behavior, SEL competencies, and emotional distress. We also carry out subgroup analyses on gender, grade, ethnicity, and parental education, just as we investigate heterogeneity in effects across the distribution of our primary outcome measured at baseline and supplement with machine-learning approaches to choose relevant subgroups.

Despite a very positive qualitative implementation evaluation (DCUM, 2020) and in contrast to existing studies from other contexts, we do not find evidence that this SEL program improved child outcomes, just as we do not find evidence of the opposite. Effects are all small and precisely estimated. However, auxiliary teacher surveys document that supportive teaching is already very common in both treatment and control classrooms, while most often delivered in an informal fashion

[1] The first version of the program, PERSPEKT, was originally developed by Allan Knægt and Jane Vinter with financial support from the ROCKWOOL Foundation. In 2017 the program was adapted for implementation at scale by DCUM leading to PERSPEKT 2.0. The updated version included a section on digital communication and was adapted such that the material could be employed without further development of teacher competency.

and absent *ex ante* research about its effectiveness. In fact, almost all teachers use some type of tool, initiative, and/or method, and many use several. Hence, PERSPEKT 2.0 was implemented on top of an already rich supportive teaching environment that may have alieved some issues with pupil school well-being and problem behaviors, for example.

We structure the remainder of the paper as follows: Section 2 describes the institutional setting, and Section 3 provides details about the intervention, recruitment, and randomization. Section 4 shows our data, and Section 5 presents the results of the evaluation. Finally, Section 6 reflects upon the findings, and Section 7 concludes.

2. Institutional setting

In Denmark, compulsory education comprises primary and lower secondary education (ISCED 1 and 2) and lasts 10 years, from grade 0 to grade 9, with the possibility of attending an optional 11th year (grade 10). Children enter school in the year they turn six years old.

According to Statistics Denmark (2018), in 2017, 79 percent of children enrolled in grades 0-9 attended the municipal public school, *Folkeskolen.* In public schools, as well as in the majority of private schools, children are divided into classes of maximum 28 pupils during grade 0. Typically, children stay together in these classes until they leave school. With the exception of a few elective courses, a class receives education in all subjects together, and is headed by a "class teacher", who follows the class for around three years. This teacher, who is usually also the Danish or Math teacher of the class, coordinates the activities of the group of subject teachers associated with the class, and is the primary point person in cases of academic, behavioral or social problems. While a teacher is usually only class teacher for one class, subject teachers teach their subjects to several classes. It is common for classes within a grade to share subject teachers, and for class- and subject teachers to work together in grade-teams. In 2017, the average class size in public schools was 21.5 (StatBank Denmark, 2018).

The majority of public schools are divided into three, often physically separated, sections: the preparatory section, *indskolingen*, which encompasses grades 0 to 3; the intermediate section, *mellemtrinnet*, which encompasses grades 4-6; and the lower secondary stage, *udskolingen*, which encompasses grades 7 to 9. Each section will typically have a section leader and/or coordinator and teachers primarily teach classes within one section.

3. PERSPEKT 2.0, recruitment, and the experimental set-up

3.1 PERSPEKT 2.0 versus treatment as usual

PERSPEKT 2.0

Treatment classes received instruction using PERSPEKT 2.0; a set of teaching materials aimed at training pupils' emotional, personal and social skills to improve individual well-being as well as the social and learning environment in the classroom. It fulfills the four criteria for best implementation practice (SAFE). It is **S**equenced, in that there is coordinated progression of activities and practices to build competencies of the pupils; it is **A**ctive, as it includes a number of participatory elements, such as role plays; it is **F**ocused in terms of having allocated specific time and program elements to build specific SEL competencies; and it is **E**xplicit in terms of having identified specific SEL competencies, that it aims to strengthen (Durlak et al., 2010). The material bears resemblance to PATHS (Promoting Alternative Thinking Strategies) and Second Step, both widely used social-emotional learning programs developed in the US, which have been subjected to several RCT based evaluations with positive results, particularly for the PATHS program (Korpershoek et al., 2016).

PERSPEKT 2.0 exists in three age-appropriate modules (Module I, II and III), targeting grades 0-3, 4-6, and 7-9. Treatment classes in our evaluation were grade 4 and 5, and hence received instruction based on Module II. This module consists of 15 chapters, each of which is designed to take 45-60 minutes to complete. Table 1 shows an overview of the chapters and their objectives. Exercises in the material include conversations, classroom exercises, and small group activities. Some chapters offer specific tools, such as key phrases or steps, for children to use in different situations. Roleplaying and games are included as a means of drawing attention to and practicing different skills.

In treatment classes, instruction in PERSPEKT 2.0 was initiated in August 2018, at the beginning of the school year. To the extent possible, instruction in successive chapters was to be spaced by one week, however schools were allowed some flexibility in timing, in order to accommodate other planned activities (e.g. thematic weeks or class trips) and teacher absences. The entire course was completed by the end of February 2019 in the majority of treatment classes. Instruction was provided by either teachers or pedagogues associated with the class. While it was recommended that the same instructor – typically the class teacher – teaches the entire course, up to two teachers were, under special circumstances, involved.

PERSPEKT 2.0 was designed to require no special training of instructors. Instructors in treatment classes were introduced to the materials through a video that demonstrated classroom practice. The teaching material itself is available through a custom-built web application, though a printed version of the material is also available upon request. Instructors were equipped with personal usernames and passwords and once logged in, they could read the chapters and exercise instructions and display project exercise materials on a smartboard in the classroom. In addition, instructors could easily keep track of the progression of their class(es) through the material at the level of individual exercises. Only instructors in treatment classes had access to the material.

Table 1

Overview of PERSPEKT 2.0 chapters in Module II

Chapter: Title	**Objectives**
1: Thoughts and emotions	Pupils understand that different people may perceive the same situation differently, and how emotions are closely related to perceptions.
2: Body language	Pupils become aware of body language and its importance in communication.
3: Communication	Pupils are introduced to the concepts of passive, aggressive, and assertive communication, and learn that they can affect situations by actively choosing communication strategy. They are introduced to and practice a strategy for assertive communication.
4: Digital communication	Pupils learn that “faceless” communication places special requirements on both sender and recipient.
5: Saying “no”	Pupils are introduced to and practice a strategy for saying “no” in difficult situations involving peer pressure.
6. Facts and assumptions	Pupils understand the difference between facts and assumptions and learn how to identify facts. In addition, they learn why we sometimes need to rely on assumptions and why it is important to be aware that they are not facts
7: Opinions	Pupils understand how opinions differ from assumptions and facts and practice distinguishing between the three.
8: From thoughts to emotion and action	Pupils gain awareness of the relationship between thoughts, emotions and actions and reflect on how their own thoughts and emotions are related to actions.
9: Consequences	Pupils reflect on how actions, including online behavior, as well as lack of action can have consequences – for oneself and for others.
10: Rules, agreements, and expectations	Pupils gain awareness of the role of rules in society and in the classroom and understand that rules are often created for the sake of the community.
11: Admitting something	Pupils are introduced to and practice a strategy for formulating an apology if, for example, rules, agreements, or expectations have been broken.
12: Roles	Pupils gain awareness of how people can have different roles in different contexts, and how this influences behaviors and expectations.

13: Other people's point of view	Pupils gain awareness of the importance of taking other people's viewpoints into consideration, and practice understanding other people's points of view.
14: Negotiation and compromise	Pupils practice negotiation and compromise and learn that sometimes we have to set aside our own wishes for the sake of the community.
15: Completion / summary	Pupils reflect on what they have learned through the course.

Treatment as usual

Classrooms allocated to the control group received "treatment as usual" (henceforth TAU). The content of this varied across schools as well as classrooms within schools, as there is no national curriculum or common goals for social skills training. The Danish Education Act stipulates that teaching of obligatory subjects and themes must be supplemented by "supportive teaching", which may include courses or activities aimed at strengthening social skills and well-being (Danish Ministry of Education, 2017a). However, the act does not include specific requirements regarding form, content, or extent. Similarly, a national Act on Educational Environment stipulates that schools undertake assessments of the educational environment at least every third year and formulate a set of school values, including an anti-bullying strategy, but requirements for content are minimal (Danish Ministry of Education, 2017b).

At participating schools, all classes that were not part of the trial (i.e. those that are grades 0-3 or 6-9 in the 2018/2019 school year) were allowed to implement PERSPEKT 2.0 throughout the trial period. We will release PERSPEKT 2.0 for use in all schools and across all classes from the beginning of the school year 2020/21.

We explore the actual implementation of PERSPEKT 2.0 as well as the use of supportive teaching in treatment and control classrooms in details below.

3.2 Recruitment

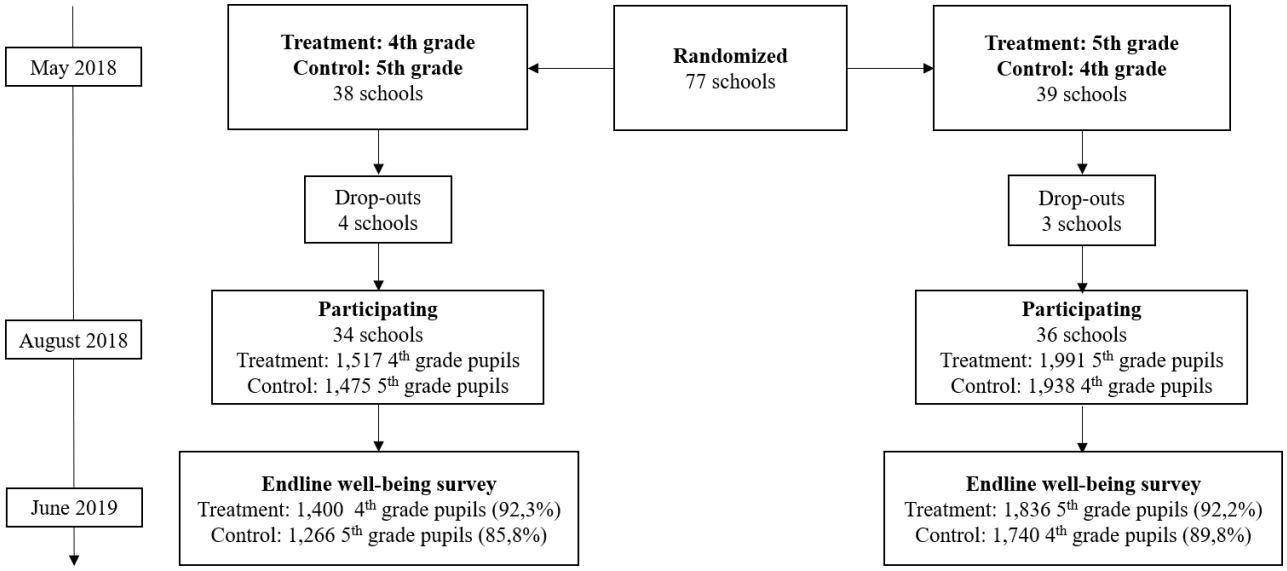
Recruitment was carried out by the DCUM during spring to fall of 2017. School level participation was voluntary and the decision to enroll was made by school principals.[2] Figure 1 illustrates randomization and school-level attrition. Seventy-seven schools agreed to participate, signed the final data agreements, and had cohorts of classrooms randomized. We randomized 38 schools to 4th grade treatment and 5th grade control and 39 schools to 4th grade control and 5th grade treatment. After

[2] Further details about the recruitment process can be found in the study protocol, Kleinstrup et al. (2018)

randomization, four schools dropped out from the former arm and three from the latter. The remaining 70 schools together enrolled 6,921 pupils in 4th or 5th grade at the time of implementation with 3,508 pupils in the treatment group and 3,413 pupils in the control group.[3] Response rates on the primary outcome measure, described in detail below, were generally high but slightly higher in the treatment classrooms. We explicitly explore this in our analyses below.

Figure 1

Flow chart of participating schools



3.3 Randomization

We employed a two-level cluster randomized trial for children in two adjacent school cohorts (fourth and fifth grades) within the same school. There is otherwise no blocking. This means that we have randomly allocated schools into teaching PERSPEKT 2.0 in either 4th grade or 5th grade such that all schools implement PERSPEKT 2.0 in only one of the two grade levels. We chose to randomize at the grade level instead of the class level to minimize spill-overs from treatment to control as teachers typically work together in grade-teams and some teachers teach multiple classes within the same grade.

In May 2018, we informed DCUM of which schools were randomly allocated to implement PERSPEKT 2.0 in 4th grade, and which schools were allocated to implement the program in 5th grade. DCUM immediately informed the schools and subsequently followed up with them to ensure that

[3] Based on administrative registers.

there had been no miscommunications, and that implementation of PERSPEKT 2.0 would be taking place in the correct grades.

4. Data

We make use of a series of data sources with individual level information about children, their families, and teachers. These include 1) nationally administered well-being surveys developed by the Danish Ministry of Education, 2) nationally administered IT-based tests of Danish reading skills and Math, 3) register-based data maintained by Statistics Denmark, 4) administrative data linking instructors to classes, 5) data from a pupil survey developed specifically for this trial (see questions in Table 3 below), and 6) teacher survey data.

The first four categories of data are available for all public schools and pupils, regardless of whether they participate in the study. Data in the last two categories are only gathered at participating schools. To minimize interference with regular school activities and promote high response rates, schools were strongly encouraged to implement the trial specific survey concurrently with the compulsory national well-being survey. In practice, the survey was administered to pupils by a teacher during school hours. It is web based and was created using the survey tool SurveyXact. Pupils accessed the survey through a common link and by way of their national pupil IDs (UNI-Login). This enabled us to link responses to personal identification numbers (CPR) and thereby to data from the other three categories.

4.1 Outcomes

<u>Primary study outcome</u>

Our primary outcome is a measure of attitudes towards school and emotional well-being in the classroom. As described in our protocol (Klejnstrup et al., 2018), we base the measure on the recently implemented national well-being indicators (Andersen et al., 2015; Danish Ministry of Education, 2018). For our primary outcome, we use responses to the survey collected in the second quarter of 2019 (May-June), i.e. following the provision of treatment. Among the full list of 40 questions in the national well-being survey for grade 4 to 9, we use only the ten questions that enter the Social Well-being subscale (Danish Ministry of Education, 2018). The answers to all questions are coded to range from one to five, with five being the most positive. We present the ten included questions in Table 2. We subsequently calculate social well-being as the within-individual average of the answers

provided. Finally, we standardize social well-being at the grade level to have a mean of zero and a standard deviation of one.

Pupils in grade 0-3 also answer a well-being survey with slightly different questions that are coded to range from one to three with three being the most positive. In order to construct baseline measures of social well-being for the 4th grade pupils in the sample, we have found eight questions in the questionnaire for grade 0-3 that correspond well to those included in the grade 4-9 social well-being subscale. We use these to construct a similar social well-being baseline measure for the 4th grade pupils. We standardize by grade level to make the baseline measure comparable. See Larsen et al. (2020) for further details about the social well-being measure.

Table 2

Social Well-being indicator questions included in primary study outcome

Question:	**Answers range from 1 to 5**
Do you like your school?	Never (1), rarely, sometimes, often, very often (5)
Do you like the other children in your classroom?	
Are you afraid of being ridiculed at school?	Always (1), mostly, sometimes, rarely, never (5)
How often do you feel safe at school?	Never (1), rarely, sometimes, mostly, always (5)
Do you feel lonely?	Very often (1), often, sometimes, rarely, never (5)
Since the start of the school year, did anyone bully you?	
I feel I belong at my school.	Strongly disagree (1), disagree, neither agree nor disagree, agree, strongly agree (5)
I like the breaks at school.	
Most of the pupils in my classroom are kind and helpful.	
Other pupils accept me as I am.	

Secondary study outcomes

Problem behavior. Our first measure of problem behavior is number of days absent from school relative to the number of school days during the intervention period (the 2018-2019 academic year). It is based on monthly school reports at the pupil level. Our second measure stems from the pupil survey, collected concurrently with the social well-being indicator. The item we use asks about the degree to which the child is likely to pick up quarrels with other children. Specifically, we ask the child to state to what extent he or she agrees with the statement "I am the type of person who initiates

quarrels with others". Response categories are: strongly disagree (1), disagree (2), neither agree nor disagree (3), agree (4), strongly agree (5).

Social emotional learning skills. PERSPEKT 2.0 seeks to improve the five SEL skills. In the PERSPEKT curriculum, the focus is especially on strengthening relationship skills and responsible decision making and certain aspects of self-awareness, self-management and social awareness. Table 3 lists each of the SEL skills, how they are described in the SEL literature, the degree to which they are covered in the PERSPEKT curriculum, and how we measure each of the five skill areas in the pupil survey. We have designed the survey questions specifically for this study. In practice, we construct scores within each skill that sum the answers from the separate items.

Emotional distress. Emotional distress is to some extent covered by our primary outcome, for example by the questions related to loneliness, the feeling of being safe, and to exposure to bullying. In the pupil survey, we ask the children to rate two further statements: 1) I am the type of person who often worries and 2) I am the type of person who often feels blue and sad. Response categories are: strongly disagree (1), disagree (2), neither agree nor disagree (3), agree (4), and strongly agree (5).

Table 3

Measuring social emotional learning, CASEL (2013)

SEL skill:	Description of skill:	PERSPEKT 2.0:	Questions: Response categories: Never (1), rarely (2), sometimes (3), mostly (4), always (5)
Self-awareness	The ability to accurately recognize one's own emotions, thoughts, and values and how they influence behavior. The ability to accurately assess one's strengths and limitations, with a well-grounded sense of confidence, optimism	PERSPEKT covers the ability to accurately recognize one's own emotions and thoughts and how they influence behavior. PERSPEKT does not directly cover the ability to assess one's strengths and weaknesses	• I reach out to an adult if I need help during sessions • I reach out to other children if I need help during sessions • If I am sad, I keep my thoughts and feelings to myself *

Self-management	The ability to successfully regulate one's emotions, thoughts, and behaviors in different situations — effectively managing stress, controlling impulses, and motivating oneself, setting and working towards achieving personal and academic goals.	PERSPEKT covers the ability to regulate one's emotions, thoughts, and behaviors in different situations, and it also covers the ability to control one's impulses. PERSPEKT does not cover the ability to manage stress or motivate oneself or to set personal and academic goals.	• I stay calm if someone says or writes something negative about me • If I get angry, I think before I react • I am the type of person who quickly forgets if something bad happens ** • I stay calm even though other children are upset
Social awareness	The ability to take the perspective of and empathize with others, including those from diverse backgrounds and cultures, to understand social and ethical norms for behavior and to recognize family, school, and community resources and supports.	PERSPEKT covers the ability to take the perspective of and empathize with others. PERSPEKT does not focus specifically on diversity and different cultures or on recognizing family, school and community resources and support.	• I do my best to understand the other children in my classroom even if I disagree with them • I feel sad if other children in my classroom are sad
Relationship skills	The ability to establish and maintain healthy and rewarding relationships with diverse individuals and groups, including communicating clearly, listening actively, cooperating, resisting inappropriate social pressure, negotiating conflict constructively, and seeking and offering help when needed.	PERSPEKT covers the ability to maintain healthy relationships through clear communication and active listening, resisting inappropriate social pressure and negotiating conflict constructively. PERSPEKT has a strong focus on relationship skills. PERSPEKT does not directly cover seeking and offering help.	• I do my best to help the other children in my classroom whenever they have a problem • It is easy for me to find new friends in school • I do my best to help the other children in my classroom when they end up in conflicts with each other
Responsible decision-making	The ability to make constructive choices about personal behavior and social	PERSPEKT covers the ability to evaluate consequences of various	• I do my best to forgive the other children in my

	interactions based on ethical standards, safety concerns, and social norms. The realistic evaluation of consequences of various actions, and a consideration of the well-being of oneself and others.	actions, and the ability to make constructive choices about personal behavior.	classroom when they apologize • I reach out to an adult if someone bullies a child in my classroom • I reach out to an adult if someone misbehaves towards me

*: Response categories are reversed: Always (1), mostly (2), sometimes (3), rarely (4), never (5)

**: Response categories: Strongly disagree (1), disagree (2), neither agree nor disagree (3), agree (4), strongly agree (5)

Note: Survey questions are designed specifically for this studied and not validated by CASEL (2013).

Exploratory study outcomes

Pupil academic performance. We measure academic performance for the subset of children in Grade 4 using nationally administered IT-based tests of Danish reading skills. These have been shown to correlate highly with later higher-stakes tests (Beuchert and Nandrup, 2018).

Finally, we show summary statistics of our eleven outcome measures in Table A1. We show the variables in levels prior to standardization to illustrate the general levels. The baseline measure of social well-being is measured in 3rd and 4th grade, which means that it is measured on two different scales. Hence, we show baseline means of social well-being for the two grades separately. For both 3rd and 4th grade pupils in the sample, the average level of social well-being is quite high. Similarly, average values of the five SEL-skills range from 3.5 to 4.0 all above the midpoint of the scale. We also note that levels are low for the indicator for sadness and for initiating quarrels with other children. However, the pupils express some level of worry as the average value is 3.2. For the regression analysis, we standardize all baseline and endline outcomes by grade level to have a mean zero and a standard deviation of one.

5. Results

This section presents estimates of the effect of being offered the intervention for pupil-level outcomes; or intention-to-treat effects. In practice, we compare PERSPEKT 2.0 with TAU using linear regressions with and without control for pre-randomization variables. In versions that control for pre-randomization variables, we include an indicator for 5th grade enrollment, the baseline

measure of the outcome, interactions between these two, child gender, and month of birth dummies. All standard errors are clustered at the grade-by-school level.

Before initiating the formal analysis, we investigate the distribution of an enriched set of pupil and teacher characteristics and outcomes across treatment and control classroom prior to randomization. In practice, we perform regressions of the randomization indicator on each characteristic and baseline outcome separately while clustering at the grade-by-school level. We also run joint regressions and because the unit of observation varies, we run one version of the regression based on pupil level variables and another for teacher, class and school level variables. Results are shown in Table A2 and indicate, as should be expected, no obvious problems with balance; none of the mean variable-by-variable differences is large and only few are statistically significant. Due to the very large conditioning set aided by the large sample, however, our F-test rejects the null hypothesis that regression coefficients are jointly zero in the pupil level regression.

Table 4 continues to show estimated effects for the overall population on standardized primary and secondary outcomes. Contrary to initial expectations, we find no evidence that PERSPEKT 2.0 improved neither primary nor secondary outcomes.[4] Effects are all small and precisely estimated; we can reject even small positive (and negative) effects.

[4] Appendix Table A3 shows versions of these using non-standardized outcomes; conclusions remain unaltered.

Table 4

Effects of PERSPEKT 2.0 on standardized primary and secondary outcomes

	No controls	Basic controls
Social well-being	-0.044	-0.043
	(0.045)	(0.034)
	[6242]	[6242]
Self-awareness	-0.013	-0.002
	(0.035)	(0.032)
	[5425]	[5425]
Self-management	-0.055	-0.025
	(0.038)	(0.030)
	[5394]	[5394]
Social awareness	0.043	0.042
	(0.038)	(0.032)
	[5419]	[5419]
Relationship skills	-0.043	-0.025
	(0.039)	(0.032)
	[5416]	[5416]
Responsible decision-making	-0.006	0.004
	(0.042)	(0.038)
	[5380]	[5380]
I am often sad	0.043	0.021
	(0.033)	(0.032)
	[5169]	[5169]
I often worry	-0.009	-0.013
	(0.033)	(0.032)
	[5186]	[5186]
I initiate quarrels with others	0.051	0.046
	(0.037)	(0.034)
	[5126]	[5126]
Absenteeism rate	-0.023	-0.002
	(0.049)	(0.034)
	[6811]	[6811]
Danish. 4th grade national test	-0.024	-0.019
	(0.070)	(0.056)
	[3328]	[3328]

Notes: Each row presents the regression coefficient from a regression of the given outcome variable on a PERSPEKT 2.0 indicator. In column 2, we also control for an indicator for 5th grade enrollment, the baseline measure of the outcome, interactions between these two, child gender, and month of birth dummies. All outcomes are standardized by grade. Grade-by-school cluster-robust standard errors in parentheses and number of observations in squared parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5.1 Sensitivity analyses

First, we explore the sensitivity of the results to the choice of control variables. In Table 4 we control for a small set of pre-randomization variables as specified in the analysis plan in the protocol. These results are reproduced if we instead choose the set of relevant control variables based on the Post-Double-Selection method (Belloni et al., 2014). To obtain maximal precision and control for any slight imbalances between the treatment and control group, the method uses LASSO to choose control variables that predict the outcome variable and the treatment dummy, respectively, among a list of controls based on the variables presented in Table A2. The union of these two sets of variables is then included as control variables. The results are presented in Appendix Table A4. As the Post-Double-Selection method does not alter the results, we maintain the small set of pre-specified control variables in the further sensitivity analyses.

Next, we supplement the results from Table 4 with a range of sensitivity analyses concerned with missing values of the outcome under consideration. Higher response rates in intervention classrooms could, for example, lead to underestimation of effects if the marginal respondent lies in the lower tail of the distribution. In one version of our estimations, therefore, we replace missing values with the lowest value of the outcome ("lower bound"), thus assuming that those who do not complete the survey belong to the most disadvantaged group; in another with the highest value of the outcome ("upper bound); and in a final version with the baseline value ("baseline"). Results are shown in Table A5. Most results are not sensitive to these changes but the effect on the primary outcome does become positive, though still small, and borderline significant in the lower bound model. More generally, however, we conclude that even extreme assumptions about missing values of the outcome cannot generate large gains or disadvantages arising from program participation.

Third, we explore effects on each of the sub-questions composing the primary outcome. To the extent that averaging across sub-questions masked impacts in particular dimensions of social well-being, this exercise will reveal these. Yet, as seen in Table A6, PERSPEKT 2.0 did not affect any of these either.

5.2 Effects on subgroups

We next investigate whether effects of exposure to PERSPEKT 2.0 vary across subgroups. We start by investigating effects by the pre-specified subgroups, again following Klejnstrup et al. (2018). We

split children into subgroups based on their baseline value of the primary outcome. These results are shown in Table 5. Specifically, we consider children above and below the median value and children in the 1st and 2nd quartile of the distribution separately. In addition, Table 6 considers subgroups based on gender, parents' country of origin (both non-Danish versus at least one parent Danish), and mother's education level (high school/less than high school versus more than high school). In versions that control for baseline characteristics, we find indications that PERSPEKT 2.0 increases social awareness slightly among boys. At the same time, however, among subgroups with lower values of the primary outcome measured at baseline as well as among boys and children with immigrant backgrounds, we detect increases in the tendency to (report to) initiate quarrels with others. One interpretation for the latter result could be the increased social awareness, of course.

To explore heterogeneity in the effects of PERSPEKT 2.0 beyond the pre-registered subgroups, we also employ machine learning inference on heterogenous treatment effects in line with Chernozhukov et al. (2018). The method allows us to test whether there is a significant difference in the group average treatment effect between the 20 percent most and least affected group, and whether the group average treatment effects for the most and least affected group are significantly different from zero, respectively. The effect sizes for the most and least affected groups are found by proxying a high-dimensional set of conditional average treatment effects using machine learning methods and then sorting groups according to effect size. We explore results from four different machine learning methods in this exercise. Here, we present results from the elastic net method, which is the machine learning proxy that maximizes the correlation between predicted and true values. Details can be found in Appendix B.

We find a significant heterogeneity in effect size between the most and least affected groups, but none of the group average treatment effects are significantly different from zero—neither for the most nor least affected group. So even using the most flexible methods to detect effects of PERSPEKT 2.0 on sub-groups of pupils, we do not find consistent evidence that PERSPEKT 2.0 has affected social well-being for any groups of pupils neither positively nor negatively.

Given these findings, there are no clear indications that PERSPEKT 2.0 improves (or worsens) the range of short-run child outcomes considered in this study.

Table 5

Effects of PERSPEKT 2.0 on standardized primary and secondary outcomes, by subgroups based on baseline value of primary outcome

	Below median	1st quartile	2nd quartile	
Social well-being	-0.033	-0.038	-0.025	
	(0.051)	(0.070)	(0.061)	
	[2578]	[1559]	[1019]	
Self-awareness	0.061	0.024	0.129	*
	(0.044)	(0.061)	(0.066)	
	[2205]	[1316]	[889]	
Self-management	0.034	0.066	-0.010	
	(0.042)	(0.055)	(0.064)	
	[2200]	[1318]	[882]	
Social awareness	0.007	0.021	0.004	
	(0.050)	(0.058)	(0.074)	
	[2203]	[1316]	[887]	
Relationship skills	-0.015	-0.008	-0.017	
	(0.048)	(0.062)	(0.064)	
	[2196]	[1313]	[883]	
Responsible decision-making	0.059	0.098	0.006	
	(0.055)	(0.071)	(0.071)	
	[2183]	[1305]	[878]	
I am often sad	-0.013	0.020	-0.086	
	(0.043)	(0.050)	(0.063)	
	[2085]	[1243]	[842]	
I often worry	-0.031	0.001	-0.081	
	(0.047)	(0.062)	(0.063)	
	[2091]	[1254]	[837]	
I initiate quarrels with others	0.032	0.006	0.071	
	(0.044)	(0.050)	(0.066)	
	[2045]	[1220]	[825]	
Absenteeism rate	-0.032	-0.057	0.001	
	(0.041)	(0.054)	(0.048)	
	[2820]	[1725]	[1095]	
Danish, 4th grade national test	0.010	-0.044	0.115	
	(0.070)	(0.085)	(0.076)	
	[1336]	[898]	[438]	

Notes: Table presents effects of PERSPEKT 2.0 on pupil level outcomes from regressions that control for baseline characteristics as in Table 4. 'Below median' indicates subgroup with baseline value of primary outcome below the median; 1st and 2nd quartile are defined analogously. Grade-by-school cluster-robust standard errors in parentheses and number of observations in squared parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6

Effects of PERSPEKT 2.0 on standardized primary and secondary outcomes, by subgroups based on characteristics measured at baseline

	Boys		Girls	Danish parents	Non-Danish parent(s)		Mother prim/sec educ.	Mother tertiary educ.
Social well-being	-0.039		-0.047	-0.054	0.075		-0.058	-0.023
	(0.040)		(0.043)	(0.035)	(0.075)		(0.040)	(0.044)
	[3180]		[3062]	[5683]	[559]		[3401]	[2841]
Self-awareness	0.020		-0.026	-0.009	0.092		-0.012	0.007
	(0.044)		(0.041)	(0.032)	(0.100)		(0.038)	(0.040)
	[2782]		[2643]	[4952]	[473]		[2928]	[2497]
Self-management	-0.017		-0.035	-0.030	0.041		-0.051	0.011
	(0.039)		(0.040)	(0.031)	(0.092)		(0.043)	(0.034)
	[2769]		[2625]	[4926]	[468]		[2916]	[2478]
Social awareness	0.091	**	-0.012	0.048	-0.028		0.041	0.049
	(0.045)		(0.040)	(0.034)	(0.088)		(0.045)	(0.037)
	[2780]		[2639]	[4946]	[473]		[2928]	[2491]
Relationship skills	0.000		-0.050	-0.020	-0.091		-0.028	-0.018
	(0.043)		(0.041)	(0.033)	(0.118)		(0.044)	(0.036)
	[2781]		[2635]	[4941]	[475]		[2926]	[2490]
Responsible decision-making	0.017		-0.013	0.007	-0.006		0.007	0.001
	(0.047)		(0.047)	(0.037)	(0.106)		(0.050)	(0.042)
	[2759]		[2621]	[4912]	[468]		[2905]	[2475]
I am often sad	0.016		0.025	0.022	-0.001		0.026	0.007
	(0.042)		(0.044)	(0.033)	(0.090)		(0.039)	(0.039)
	[2648]		[2521]	[4730]	[439]		[2765]	[2404]
I often worry	-0.022		-0.004	-0.011	-0.099		0.015	-0.041
	(0.041)		(0.037)	(0.032)	(0.089)		(0.039)	(0.046)
	[2647]		[2539]	[4745]	[441]		[2775]	[2411]
I initiate quarrels with others	0.084	*	0.005	0.036	0.163	*	0.038	0.051
	(0.042)		(0.046)	(0.035)	(0.092)		(0.045)	(0.041)
	[2636]		[2490]	[4692]	[434]		[2739]	[2387]
Absenteeism rate	-0.022		0.017	0.001	-0.038		0.016	-0.027
	(0.042)		(0.037)	(0.035)	(0.056)		(0.043)	(0.037)
	[3496]		[3315]	[6179]	[632]		[3731]	[3080]
Danish, 4th grade national test	0.007		-0.048	-0.034	0.048		0.000	-0.030
	(0.068)		(0.056)	(0.056)	(0.110)		(0.057)	(0.067)
	[1729]		[1599]	[3021]	[307]		[1768]	[1560]

Notes: Table presents effects of PERSPEKT 2.0 on pupil level outcomes from regressions that control for baseline characteristics as in Table 4. Grade-by-school cluster-robust standard errors in parentheses and number of observations in squared parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

6. Discussion and perspectives: reconciling the evidence?

Our quantitative findings are in stark contrast to a very positive and thorough qualitative, implementation evaluation (DCUM, 2020) that indicated that the program was delivered as intended: The implementation evaluation showed that the teachers found the material to be accessible and well structured, just as the themes and core elements were found to be meaningful and relevant. Teachers experienced that pupils benefitted from program participation. Teachers were primarily skeptical about the rather inflexible structure of PERSPEKT 2.0 but the skepsis wore off as teachers gained experience with the format. Fidelity was high: 82% of the classes completed chapters 1-13 and 81% the entire material in chapters 1-15. As few as 3% of teachers report that they did not prepare for the teaching and teachers spent on average roughly 30 minutes on preparing each chapter. Pupils were equally positive towards the program, though with some heterogeneity in attitudes towards the various components and types of exercises involved.

Why, then, was PERSPEKT 2.0 not successful in improving pupil outcomes? One important reason for the lack of results might lie in the use of other supportive teaching in control as well as treatment classrooms. To document the implementation of PERSPEKT 2.0 as well as the content of treatment as usual, we collected information on activities and courses undertaken to improve well-being and socio-emotional skills through a survey targeted at teachers responsible for class-well-being. Surveys were administered to class teachers once a year during the study period, concurrently with the administration of a well-being survey to pupils. In practice, as seen in Table 7, in the end-line teacher survey, PERSPEKT 2.0 teachers reported that program delivery most often occurred during Danish language lessons (61%) and/or assisted learning lessons (48%). With this is mind, it is reassuring that we do not detect negative effects on national test scores in Danish reading. Only 13% of teachers reported that the program was taught during math lessons. Thus, PERSPEKT 2.0 was often delivered during lessons in which the usual supportive teaching would take place. Moreover, supportive teaching is very common in both treatment and control classrooms as illustrated in Table 7; almost all teachers use some type of tool, initiative, and/or method, and many use several. With the exception of PERSPEKT 2.0, of course, most activities appear individually driven and informal, rather than being formal social emotional learning programs. Hence, PERSPEKT 2.0 should be contrasted to an already rich supportive (but more informal) teaching environment that may already have alieved some issues with pupil well-being and problem behaviors, for example. PERSPEKT 2.0 did only replace other activities to a smaller extent.

Table 7

Use of supportive teaching in PERSPEKT 2.0 and control classrooms

	Perspekt			Control	
	# Obs.	Share		# Obs.	Share
During which lessons have you taught Perspekt 2.0?	132		During which lessons have you taught Perspekt 2.0?	0	
Danish language lessons	80	0.61	Danish language lessons	0	
Assisted learning	64	0.48	Assisted learning	0	
Math	17	0.13	Math	0	
Others	44	0.33	Others	0	
Which tools have you used during this academic year?	132		Which tools have you used during this academic year?	128	
Any	128	0.97	Any	124	0.97
Most common: Class rules or well-being goals developed by the class	107	0.81	Most common: Class rules or well-being goals developed by the class	104	0.81
Second most common: Individual conversations regarding well-being between teacher and student	94	0.71	Second most common: Individual conservations regarding well-being between teacher and student	104	0.81
Third most common: Regular contact with parents	78	0.59	Third most common: Regular contact with parents	92	0.72
Which initiatives have you used during this academic year?	132		Which initiatives have you used during this academic year?	128	
Any	127	0.96	Any	124	0.97
Most common: Individual conversations with students	100	0.76	Most common: Individual conversations with students	101	0.79
Second most common: Joint play in recess	75	0.57	Second most common: Joint play in recess	75	0.59
Third most common: Class meetings with well-being themes (e.g. digital culture, conflict solution, bullying etc.)	68	0.52	Third most common: Social happenings outside school	75	0.59
Which methods have you used during this academic year?	132		Which methods have you used during this academic year?	128	
Any	124	0.94	Any	123	0.96
Most common: Educational material in social media	66	0.50	Most common: Educational material in social media	68	0.53
Second most common: Self-made course	55	0.42	Second most common: Self-made course	63	0.49
Third most common: Uge 6	47	0.36	Third most common: Uge 6	56	0.44
Which specific courses have you used during this academic year?	132		Which specific courses have you used during this academic year?	128	
Any	112	0.85	Any	59	0.46
Most common: Perspekt	104	0.79	Most common: We haven't and doesn't plan on completing any courses	34	0.27
Second most common: Fri for mobberi	20	0.15	Second most common: Stærke Sammen (Red Barnet)	16	0.13
Third most common: Trin-for-trin	11	0.08	Third most common: Fri for mobberi	11	0.09

Notes: Endline teacher survey based on 160 PERSPEKT 2.0 classrooms and 155 control classrooms.

Another reason behind the lack of findings on school social well-being could be the initial high average levels demonstrated by Larsen et al. (2020) and Knoop et al. (2017). When only a small share of pupils experiences low levels of well-being, the potential for improvement is obviously small. Still, as we demonstrate above, we do not detect benefits in the subgroups with low levels of baseline well-being either.

Finally, a natural question is to what extent these results will generalize to other schools. This is inherently difficult to answer; schools who chose to participate in the evaluation may be different from other schools in a range of dimensions. Table A7 demonstrates, however, that neither the pupil body, nor school and teacher characteristics are strikingly different across the two types of schools.

7. Conclusion

This paper uses a large-scale randomized trial to evaluate the short-run effects of a structured social emotional learning intervention, PERSPEKT 2.0, on pupil outcomes in Danish public schools. The intervention is similar in nature to various international programs that have previously shown positive effects. The quantitative evaluation combines survey data with register-based data that allows for tracking of participant outcomes with minimal risk of attrition. Despite a positive implementation evaluation based on both teacher and pupil informants, we find no evidence that PERSPEKT 2.0 lead to improvements in our primary outcome, namely pupils' school social well-being, nor do we detect effects on secondary outcomes such as measures of problem behavior, SEL competencies, and emotional distress. These conclusions hold in the population as a whole as well as in all pre-defined subgroups. Auxiliary survey data show that the use of more informal supportive teaching is widespread in both treatment and control classrooms. We conjecture that this could be the main reason behind the lack of positive effects.

Literature

Ahrens, A., C. B. Hansen, and M. E. Schaffer (2018). PDSLASSO: Stata module for post-selection and post-regularization OLS or IV estimation and inference, Statistical Software Components S458459, Boston College Department of Economics.

Andersen, S. C., M. Gensowski, S. Ludeke and J. H. Pedersen (2015). Evaluering af den nationale trivselsmåling for folkeskoler – og forslag til justeringer. [Evaluation of the national social well-being measurement – and suggestions for adjustments]. Trygfondens Børneforskningcenter, Aarhus Universitet.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies* 81(2), 608–650.

Beuchert, L. V. and A. B. Nandrup (2018). "The Danish National Tests at a Glance." *Danish Journal of Economics* 1, 1–37.

Chernozhukov, V., M. Demirer, E. Duflo and I. Fernández-Val (2018). Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments. NBER working paper 24678, https://www.nber.org/papers/w24678.pdf. Accessed 30 Jul 2020.

Collaborative for Academic, Social and Emotional Learning, CASEL (2013). Effective Social and Emotional Learning Programs: Preschool and Elementary School Edition. In: CASEL guide. 2013. http://casel.org/wp-content/uploads/2016/01/2013-casel-guide-1.pdf . Accessed 10 Jul 2018.

Connolly P., S. Miller, J. Mooney, S. Sloan and J. Hanratty (2016). Universal School-Based Programmes for Improving Social and Emotional Outcomes in Children Aged 3-11 years: A Systematic Review and Meta-Analysis. Campbell Collaboration.

Corcoran R.P and R. E. Slavin (2016). Effective programs for social and emotional learning (SEL): A systematic review. Campbell Systematic Reviews.

Danish Ministry of Education (2017a). Bekendtgørelse af lov om Folkeskolen, §16a. LBK nr. 1510 from 14/12/2017. [The Danish Education Act]. 2017. Danish Available at: https://www.retsinformation.dk/Forms/R0710.aspx?id=196651. Accessed 31 Aug 2018.

Danish Ministry of Education (2017b). Bekendtgørelse af lov om Folkeskolen, §1a and §6. LBK nr. 316 from 14/12/2017. [The Danish Education Act]. 2017. Danish Available at: https://www.retsinformation.dk/Forms/R0710.aspx?id=188636. Accessed 31 Aug 2018.

Dansk center for undervisningsmiljø (DCUM), and Danish Centre for Social Science Research (VIVE) (2020). Perspekt 2.0 – Implementeringsevaluering. Dansk center for undervisningsmiljø, August 2020.

Durlak J. A., R. P. Weissberg, A. B. Dymnicki, R. D. Taylorn and K. B. Schellinger (2011). The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. *Child Development* 82(1), 405-32.

Jones S. M. and J. Kahn (2017). The Evidence Base for How We Learn: Supporting Students' Social, Emotional, and Academic Development. *The WERA Educational Journal* 10(1), 5-20.

Klejnstrup, N., A. F. Larsen, H. Lilleør and M. Simonsen (2018). Social emotional learning in the classroom: Study protocol for a randomized controlled trial of PERSPEKT 2.0. Economics Working Papers, Aarhus University, 2018-11.

Koch A., J. Nafziger and H. S. Nielsen (2015). Behavioral Economics of Education. *Journal of Economic Behavior & Organization* 115(C), 3-17.

Korpershoek H., T. Harms, H. de Boer, M. van Kuijk and S. Doolaard (2016). A Meta-Analysis of the Effects of Classroom Management Strategies and Classroom Management Programs on Students' Academic, Behavioral, Emotional, and Motivational Outcomes. *Review of Educational Research* 86(3), 643–680.

Knoop, H. H., B. E. Holstein, H. Viskum, and J. M. Lindskov (2017). "Elevernes fællesskab og trivsel i skolen. Analyser af Den Nationale Trivselsmåling". DCUM report.

Larsen, A. F., A. S. C. Leme and M. Simonsen (2020). Pupil Well-being in Danish Primary and Lower Secondary School. Mimeo, Aarhus University.

Payton, J., R. P. Weissberg, J. A. Durlak, A. B. Dymnicki, R. D. Taylor, R.D., K. B. Schellinger, and M. Pachan (2008). "The positive impact of social and emotional learning for kindergarten to eighth-grade students: Findings from three scientifi c reviews. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning (CASEL).

Statistics Denmark (2018). Grundskole. http://www.dst.dk/da/Statistik/emner/uddannelse-og viden/fuldtidsuddannelser/grundskole. Accessed 17 Jul 2018.

StatBank Denmark (2018). Kvotien: Klassekvontienter i grundskolen efter område, klassetrin og skoletype. http://www.statistikbanken.dk/KVOTIEN. Accessed 17 Jul 2018.

Taylor R. D., E. Oberle, J. A. Durlak and R. P. Weissberg (2017). Promoting Positive Youth Development through School-Based Social and Emotional Learning Interventions: A Meta-Analysis of Follow-Up Effects. *Child Development* 88(4), 1156–1171.

Appendix A. Tables and Figures

Table A1

Summary statistics of baseline outcomes, non-standardized

	Mean (std.dev.) [obs.]
Social well-being, 3rd grade	2.576 (0.346) [3240]
Social well-being, 4th grade	4.096 (0.637) [3228]
Self-awareness	3.415 (0.638) [6140]
Self-management	3.521 (0.775) [6096]
Social awareness	3.993 (0.761) [6168]
Relationship skills	3.728 (0.689) [6143]
Responsible decision-making	3.912 (0.711) [6127]
I am often sad	2.031 (1.057) [5699]
I often worry	3.068 (1.158) [5592]
I initiate quarrels with other children	2.187 (1.008) [5550]
Absenteeism rate	0.055 (0.051) [6882]
Danish, 2nd grade national test	0.050 (0.962) [6660]

Notes: Table shows full population means, standard deviations in parentheses and number of observations in squared brackets. Summary statistics are calculated on non-standardized baseline outcomes, except from Danish 2nd grade

national test, which is a standardized test score by construction. Social well-being is presented separately by grade, as the scale is 1-3 for 3rd grade students, and 1-5 for 4th grade students.

Table A2

Pre-randomization balance

	Treatment	Control	Balance test	
Panel A: Characteristics of students				
Baseline outcomes				
Social well-being, std.	0.000	0.000	0.000	
	(1.010)	(0.990)	(0.012)	
	[3269]	[3138]	[6407]	
Self-awareness, std.	-0.021	0.022	-0.011	
	(1.018)	(0.981)	(0.010)	
	[3089]	[3002]	[6091]	
Self-management, std.	-0.028	0.028	-0.014	
	(0.991)	(1.008)	(0.010)	
	[3069]	[2978]	[6047]	
Social awareness, std.	0.002	-0.002	0.001	
	(1.004)	(0.995)	(0.011)	
	[3100]	[3019]	[6119]	
Relationship skills, std.	-0.028	0.029	-0.014	
	(0.999)	(1.000)	(0.011)	
	[3090]	[3004]	[6094]	
Responsible decision-making, std.	-0.022	0.023	-0.011	
	(1.015)	(0.984)	(0.010)	
	[3088]	[2991]	[6079]	
I am often sad, std.	0.025	-0.026	0.013	*
	(1.009)	(0.990)	(0.007)	
	[2873]	[2780]	[5653]	
I often worry, std.	0.010	-0.010	0.005	
	(0.992)	(1.008)	(0.008)	
	[2825]	[2724]	[5549]	
I initiate quarrels with others, std.	0.027	-0.028	0.014	
	(0.998)	(1.001)	(0.008)	
	[2810]	[2697]	[5507]	
Absenteeism rate, std.	-0.022	0.023	-0.011	
	(1.025)	(0.973)	(0.014)	
	[3451]	[3350]	[6801]	
Danish, 4th grade national test	-0.019	0.020	-0.010	
	(1.001)	(0.999)	(0.015)	
	[3363]	[3238]	[6601]	
Student characteristics				
Male	0.508	0.521	-0.013	
	(0.500)	(0.500)	(0.013)	
	[3471]	[3369]	[6840]	
Indicator if enrolled in 5th grade	0.567	0.431	0.136	
	(0.496)	(0.495)	(0.095)	
	[3471]	[3369]	[6840]	
Indicator if student has special needs	0.005	0.008	-0.108	

	(0.072)	(0.089)	(0.092)	
	[3471]	[3369]	[6840]	
Average yearly number of hospital visits, from 2013 to 2016	0.331	0.321	0.010	
	(0.513)	(0.477)	(0.021)	
	[3471]	[3369]	[6840]	
Child has been in contact with psychiatric hospital since birth	0.034	0.030	0.033	
	(0.181)	(0.170)	(0.035)	
	[3471]	[3369]	[6840]	
Birthweight under 2500 grams	0.044	0.050	-0.037	
	(0.205)	(0.219)	(0.032)	
	[3471]	[3369]	[6840]	
Started school year after child turned 6	0.071	0.066	0.022	
	(0.257)	(0.248)	(0.032)	
	[3471]	[3369]	[6840]	
Placed outside home, 2010 to 2017	0.006	0.010	-0.134	*
	(0.078)	(0.101)	(0.072)	
	[3471]	[3369]	[6840]	
Received social preventive measures by municipality, since birth	0.044	0.050	-0.037	
	(0.205)	(0.219)	(0.036)	
	[3471]	[3369]	[6840]	
Repeated a grade	0.033	0.036	-0.021	
	(0.179)	(0.186)	(0.050)	
	[3471]	[3369]	[6840]	
Child switched school but not address, any time	0.114	0.113	0.001	
	(0.318)	(0.317)	(0.067)	
	[3471]	[3368]	[6839]	
Switched school, any time	0.173	0.194	-0.035	
	(0.378)	(0.396)	(0.050)	
	[3471]	[3368]	[6839]	
Switched adress, any time	0.290	0.294	-0.004	
	(0.454)	(0.455)	(0.019)	
	[3471]	[3368]	[6839]	
Household characteristics				
Single provider household	0.195	0.199	-0.007	
	(0.396)	(0.399)	(0.022)	
	[3471]	[3369]	[6840]	
Parents have lived apart for two consecutive years	0.245	0.241	0.006	
	(0.430)	(0.428)	(0.023)	
	[3471]	[3369]	[6840]	
Number of siblings	1.709	1.717	-0.002	
	(1.128)	(1.092)	(0.009)	
	[3471]	[3369]	[6840]	
Birth order on mother's side	1.858	1.855	0.001	
	(0.965)	(0.935)	(0.008)	
	[3471]	[3369]	[6840]	
Paternal characteristics				
Age at birth of child	32.536	32.554	0.000	
	(6.933)	(7.089)	(0.001)	
	[3471]	[3369]	[6840]	
Unskilled	0.162	0.164	-0.005	
	(0.368)	(0.371)	(0.027)	
	[3471]	[3369]	[6840]	
Highschool	0.041	0.045	-0.024	
	(0.199)	(0.208)	(0.035)	

	[3471]	[3369]	[6840]	
Vocational	0.452	0.434	0.018	
	(0.498)	(0.496)	(0.022)	
	[3471]	[3369]	[6840]	
Short further	0.078	0.071	0.025	
	(0.268)	(0.257)	(0.024)	
	[3471]	[3369]	[6840]	
Medium further	0.131	0.131	0.001	
	(0.338)	(0.337)	(0.023)	
	[3471]	[3369]	[6840]	
Long further or PhD	0.103	0.113	-0.027	
	(0.303)	(0.316)	(0.048)	
	[3471]	[3369]	[6840]	
2nd quartile of taxable income, 2017	0.243	0.249	-0.008	
	(0.429)	(0.432)	(0.024)	
	[3471]	[3369]	[6840]	
3rd quartile of taxable income, 2017	0.247	0.241	0.008	
	(0.431)	(0.428)	(0.016)	
	[3471]	[3369]	[6840]	
4th quartile of taxable income, 2017	0.243	0.246	-0.003	
	(0.429)	(0.431)	(0.036)	
	[3471]	[3369]	[6840]	
Committed crime after birth of child	0.072	0.074	-0.006	
	(0.259)	(0.261)	(0.030)	
	[3471]	[3369]	[6840]	
Received any unemployment assistance, 2017	0.203	0.190	0.020	
	(0.402)	(0.392)	(0.023)	
	[3419]	[3311]	[6730]	
Received any unemployment insurance, 2017	0.054	0.058	-0.021	
	(0.226)	(0.234)	(0.029)	
	[3419]	[3311]	[6730]	
Avg. yearly number of visits to somatic hospitals, 2013 to 2016	0.544	0.560	-0.007	
	(0.732)	(0.764)	(0.012)	
	[3471]	[3369]	[6840]	
Visited psychiatric hospital since birth of child	0.068	0.063	0.021	
	(0.252)	(0.243)	(0.029)	
	[3471]	[3369]	[6840]	
Maternal characteristics				
Age at birth of child	30.267	30.311	0.000	
	(5.106)	(5.115)	(0.002)	
	[3471]	[3369]	[6840]	
Unskilled	0.122	0.131	-0.020	
	(0.328)	(0.338)	(0.030)	
	[3471]	[3369]	[6840]	
Highschool	0.042	0.046	-0.022	
	(0.201)	(0.209)	(0.029)	
	[3471]	[3369]	[6840]	
Vocational	0.377	0.343	0.037	*
	(0.485)	(0.475)	(0.019)	
	[3471]	[3369]	[6840]	
Short further	0.055	0.060	-0.024	
	(0.228)	(0.237)	(0.030)	
	[3471]	[3369]	[6840]	

Medium further	0.281	0.292	-0.013
	(0.450)	(0.455)	(0.018)
	[3471]	[3369]	[6840]
Long further or PhD	0.107	0.112	-0.014
	(0.309)	(0.316)	(0.045)
	[3471]	[3369]	[6840]
2nd quartile of taxable income, 2017	0.247	0.251	-0.006
	(0.431)	(0.434)	(0.020)
	[3471]	[3369]	[6840]
3rd quartile of taxable income, 2017	0.245	0.254	-0.012
	(0.430)	(0.436)	(0.017)
	[3471]	[3369]	[6840]
4th quartile of taxable income, 2017	0.257	0.242	0.020
	(0.437)	(0.428)	(0.032)
	[3471]	[3369]	[6840]
Committed crime after birth of child	0.025	0.024	0.010
	(0.157)	(0.154)	(0.042)
	[3471]	[3369]	[6840]
Received any unemployment assistance, 2017	0.270	0.274	-0.005
	(0.444)	(0.446)	(0.024)
	[3455]	[3358]	[6813]
Received any unemployment insurance, 2017	0.080	0.087	-0.024
	(0.271)	(0.282)	(0.025)
	[3455]	[3358]	[6813]
Avg. yearly amount of visits to somatic hopsital, 2013 to 2016	0.804	0.836	-0.008
	(0.982)	(0.993)	(0.010)
	[3471]	[3369]	[6840]
Visited psychiatric hospital since birth of child	0.095	0.099	-0.011
	(0.293)	(0.298)	(0.025)
	[3471]	[3369]	[6840]
Panel B: Characteristics of teachers, classes and schools			
Age of teacher	43.677	44.441	-0.002
	(10.097)	(9.832)	(0.003)
	[130]	[127]	[257]
Male	0.238	0.197	0.061
	(0.428)	(0.399)	(0.076)
	[130]	[127]	[257]
Has degree in teaching or comparable	0.991	0.992	-0.015
	(0.093)	(0.091)	(0.357)
	[115]	[122]	[237]
Tenure is 1-2 years	0.261	0.303	-0.052
	(0.441)	(0.462)	(0.072)
	[115]	[122]	[237]
Tenure is 3-4 years	0.252	0.230	0.031
	(0.436)	(0.422)	(0.076)
	[115]	[122]	[237]
Tenure is 5 years or above	0.409	0.377	0.033
	(0.494)	(0.487)	(0.067)
	[115]	[122]	[237]
Years since degree obtained	15.612	14.683	0.002
	(10.104)	(10.661)	(0.003)
	[121]	[123]	[244]
Class size	21.706	21.742	-0.001
	(3.923)	(3.625)	(0.007)

	[160]	[155]	[315]	
Turnover rate of math and danish teachers, 2016 to 2017	16.488	16.573	0.000	
	(10.569)	(10.645)	(0.003)	
	[160]	[155]	[315]	
Turnover rate of all teachers, 2016 to 2017	17.881	18.022	0.000	
	(11.243)	(11.395)	(0.003)	
	[160]	[155]	[315]	
Joint test of significance for Panel A:		F(58,138) =	2.342	***
Joint test of significance for Panel B:		F(10,221) =	0.372	

Notes: Table shows variable-by-variable means and standard deviations (columns 1-2, standard deviations in parentheses and number of observations in squared brackets) and regression coefficients and standard errors (column 3) from regressions of the treatment indicator on each variable separately. In the bottom, we present F-tests for joint significance from a regression with all variables in panel A and panel B separately. Separate analyses for each panel. Standard errors in regressions are school-grade cluster robust for panel A and Huber-White heteroskedastic robust for panel B. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A3

Effects of PERSPEKT 2.0 on primary and secondary outcomes, non-standardized outcomes

	No controls	Basic controls
Social well-being	-0.028	-0.028
	(0.029)	(0.022)
	[6242]	[6242]
Self-awareness	-0.003	-0.003
	(0.022)	(0.019)
	[5425]	[5425]
Self-management	-0.034	-0.019
	(0.029)	(0.022)
	[5394]	[5394]
Social awareness	0.033	0.029
	(0.027)	(0.023)
	[5419]	[5419]
Relationship skills	-0.025	-0.018
	(0.025)	(0.021)
	[5416]	[5416]
Responsible decision-making	-0.008	0.001
	(0.029)	(0.026)
	[5380]	[5380]
I am often sad	0.041	0.022
	(0.036)	(0.034)
	[5169]	[5169]
I often worry	0.000	-0.016
	(0.037)	(0.035)
	[5186]	[5186]
I initiate quarrels with others	0.045	0.041
	(0.034)	(0.031)
	[5126]	[5126]
Absenteeism rate	0.000	0.000
	(0.003)	(0.002)
	[6811]	[6811]
Danish, 4th grade national test	-0.023	-0.018
	(0.067)	(0.053)
	[3329]	[3329]

Notes: Each row presents the regression coefficient from a regression of the given outcome variable on a PERSPEKT 2.0 indicator. In column 2, we also control for an indicator for 5th grade enrollment, the baseline measure of the outcome, interactions between these two, child gender, and month of birth dummies. Grade-by-school cluster-robust standard errors in parentheses and number of observations in squared parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4

Treatment effects using Post-Double Selection LASSO

	No controls	Basic controls	Post-Double Selection
Social well-being	-0.044	-0.043	-0.055
	(0.045)	(0.034)	(0.046)
	[6242]	[6242]	[6242]
Self-awareness	-0.013	-0.002	-0.014
	(0.035)	(0.032)	(0.034)
	[5425]	[5425]	[5425]
Self-management	-0.055	-0.025	-0.028
	(0.038)	(0.030)	(0.031)
	[5394]	[5394]	[5394]
Social awareness	0.043	0.042	0.050
	(0.038)	(0.032)	(0.031)
	[5419]	[5419]	[5419]
Relationship skills	-0.043	-0.025	-0.025
	(0.039)	(0.032)	(0.033)
	[5416]	[5416]	[5416]
Responsible decision-making	-0.006	0.004	-0.002
	(0.042)	(0.038)	(0.039)
	[5380]	[5380]	[5380]
I am often sad	0.043	0.021	0.017
	(0.033)	(0.032)	(0.031)
	[5169]	[5169]	[5169]
I often worry	-0.009	-0.013	-0.020
	(0.033)	(0.032)	(0.032)
	[5186]	[5186]	[5186]
I initiate quarrels with others	0.051	0.046	0.033
	(0.037)	(0.034)	(0.033)
	[5126]	[5126]	[5126]
Absenteeism rate	-0.023	-0.002	0.009
	(0.049)	(0.034)	(0.033)
	[6811]	[6811]	[6811]
Danish, 4th grade national test	-0.024	-0.019	-0.005
	(0.070)	(0.056)	(0.050)
	[3328]	[3328]	[3328]

Note: Table presents effects of PERSPEKT 2.0 on pupil level outcomes from OLS regressions. Column 1 and 2 are identical to Table 4 in the main paper. In column 3 control variables are chosen using Post-Double-Selection LASSO (Belloni et al., 2014). Control variables are chosen among the set of pre-randomization variables presented in Table A2 along with squared terms of all non-binary variables, missing indicators, first-order interaction terms and standardization of all variables. We use two sets of LASSO procedures to select variables that correlate with the outcome variable and the treatment dummy, respectively. The union of these variables is included in the final regression. The procedure is implemented in Stata using the command pdslasso (Ahrens et al., 2018).

Table A5

Effects of PERSPEKT 2.0 on primary and secondary outcomes

Sensitivity to assumptions about missing values of the outcome

	Lower bound		Upper bound		Baseline	
	No controls	Basic controls	No controls	Basic controls	No controls	Basic controls
SWB in 2019	0.103 *	0.112 **	-0.064 **	-0.065 ***	0.005	-0.010
	(0.057)	(0.054)	(0.028)	(0.024)	(0.035)	(0.025)
	[6921]	[6921]	[6921]	[6921]	[6839]	[6839]
Self-awareness	0.093	0.097	-0.062	-0.064	-0.003	0.002
	(0.089)	(0.082)	(0.053)	(0.050)	(0.022)	(0.017)
	[6921]	[6921]	[6921]	[6921]	[6625]	[6625]
Self-management	0.069	0.078	-0.081	-0.081 *	-0.021	-0.008
	(0.093)	(0.085)	(0.052)	(0.048)	(0.027)	(0.018)
	[6921]	[6921]	[6921]	[6921]	[6599]	[6599]
Social awareness	0.140	0.138	-0.013	-0.020	0.038	0.030
	(0.107)	(0.096)	(0.040)	(0.038)	(0.027)	(0.020)
	[6921]	[6921]	[6921]	[6921]	[6632]	[6632]
Relationship skills	0.088	0.099	-0.066	-0.062	-0.024	-0.010
	(0.100)	(0.091)	(0.044)	(0.041)	(0.025)	(0.017)
	[6921]	[6921]	[6921]	[6921]	[6617]	[6617]
Responsible decision-making	0.103	0.120	-0.052	-0.043	-0.019	0.000
	(0.102)	(0.093)	(0.043)	(0.041)	(0.027)	(0.022)
	[6921]	[6921]	[6921]	[6921]	[6615]	[6615]
I am often sad	-0.074	-0.083	-0.074	-0.083	0.044	0.020
	(0.106)	(0.100)	(0.106)	(0.100)	(0.032)	(0.028)
	[6921]	[6921]	[6921]	[6921]	[6446]	[6446]
I often worry	-0.065	-0.079	-0.065	-0.079	0.036	0.010
	(0.066)	(0.063)	(0.066)	(0.063)	(0.034)	(0.030)
	[6921]	[6921]	[6921]	[6921]	[6410]	[6410]
I initiate quarrels with others	-0.047	-0.050	0.066	0.055	0.060 **	0.040
	(0.101)	(0.093)	(0.043)	(0.041)	(0.030)	(0.025)
	[6921]	[6921]	[6921]	[6921]	[6381]	[6381]
Absenteeism rate	-0.005	-0.004	0.000	0.000	0.000	0.000
	(0.004)	(0.004)	(0.003)	(0.002)	(0.003)	(0.002)
	[6921]	[6921]	[6921]	[6921]	[6897]	[6897]
Danish, 4th grade national test	-0.644	0.061	0.407	-0.058 *	-0.037	-0.005
	(0.497)	(0.046)	(0.327)	(0.033)	(0.052)	(0.027)
	[6921]	[6921]	[6921]	[6921]	[6755]	[6755]

Notes: Table presents effects of PERSPEKT 2.0 on pupil level outcomes. 'Lower bound' column replaces missing values by the lowest value of the outcome; 'upper bound' replaces missing values by the highest value of the outcome; 'baseline' replaces missing values by the baseline value of the outcome. Grade-by-school cluster-robust standard errors in parentheses and number of observations in squared parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A6

Effects of PERSPEKT 2.0 on individual social well-being questions

	Non-standardized outcomes		Standardized outcomes	
	No controls	Basic controls	No controls	Basic controls
How well do you like your school?	-0.046	-0.051	-0.054	-0.056
	(0.044)	(0.037)	(0.049)	(0.041)
	[6206]	[6206]	[6206]	[6206]
How well do you like the other children in your classroom?	-0.027	-0.035	-0.041	-0.040
	(0.039)	(0.033)	(0.044)	(0.037)
	[6204]	[6204]	[6204]	[6204]
Do you feel lonely?	-0.029	-0.033	-0.039	-0.035
	(0.033)	(0.030)	(0.034)	(0.032)
	[6168]	[6168]	[6168]	[6168]
Are you afraid of being ridiculed at school?	0.007	0.017	0.009	0.015
	(0.036)	(0.035)	(0.030)	(0.029)
	[6132]	[6132]	[6132]	[6132]
Do you feel safe at school?	-0.046	-0.042	-0.044	-0.043
	(0.042)	(0.033)	(0.043)	(0.035)
	[6116]	[6116]	[6116]	[6116]
Since the start of the school year, did anyone bully you?	-0.009	-0.023	-0.025	-0.026
	(0.031)	(0.026)	(0.033)	(0.029)
	[6064]	[6064]	[6064]	[6064]
I feel I belong at my school	-0.031	-0.025	-0.024	-0.024
	(0.044)	(0.035)	(0.043)	(0.034)
	[6087]	[6087]	[6087]	[6087]
I like the breaks at school	0.007	0.016	0.016	0.019
	(0.031)	(0.028)	(0.036)	(0.033)
	[6221]	[6221]	[6221]	[6221]
Most of the pupils in my classroom are kind and helpful	-0.063	-0.064 *	-0.073	-0.072 *
	(0.039)	(0.034)	(0.044)	(0.038)
	[6157]	[6157]	[6157]	[6157]
Other pupils accept me as i am	-0.049	-0.045	-0.045	-0.047
	(0.035)	(0.030)	(0.037)	(0.032)
	[6030]	[6030]	[6030]	[6030]

Notes: Table presents effects of PERSPEKT 2.0 on individual questions entering the primary outcome. Grade-by-school cluster-robust standard errors in parentheses and number of observations in squared parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A7

Comparing experimental schools to other public schools

	Experimental schools	Other public schools	Balance test	
Average class size	21.223	20.870	0.000	
	(3.100)	(7.106)	(0.001)	
	[70]	[1169]	[1239]	
School size	501.143	429.073	0.000	**
	(258.947)	(255.709)	(0.000)	
	[70]	[1169]	[1239]	
Share of students below median of social well-being in 2017/2018	0.510	0.495	0.109	
	(0.081)	(0.089)	(0.072)	
	[70]	[1128]	[1198]	
Share of students in 1st quartile of social well-being in 2017/2018	0.269	0.262	0.077	
	(0.066)	(0.073)	(0.086)	
	[70]	[1128]	[1198]	
Share of students in 2nd quartile of social well-being in 2017/2018	0.241	0.233	0.204	
	(0.034)	(0.047)	(0.127)	
	[70]	[1128]	[1198]	
Share of students with special educational needs	0.032	0.035	-0.027	
	(0.047)	(0.082)	(0.048)	
	[70]	[1169]	[1239]	
Share of students with non-Danish parent(s)	0.093	0.115	-0.073	**
	(0.078)	(0.129)	(0.033)	
	[70]	[1169]	[1239]	
Average GPA of statutory exams in 9th grade	7.126	7.113	0.001	
	(0.789)	(0.949)	(0.008)	
	[59]	[800]	[859]	
Average std. test score of National Test in Danish for 2nd, 4th, 6th, and 8th grade	-0.008	-0.013	0.003	
	(0.250)	(0.297)	(0.020)	
	[70]	[1125]	[1195]	
Average std. test score of National Test in Math for 3rd, 6th, and 8th grade	0.019	-0.007	0.015	
	(0.304)	(0.312)	(0.021)	
	[70]	[1124]	[1194]	
Share of teachers with 0 years of tenure	0.100	0.120	-0.075	***
	(0.052)	(0.121)	(0.026)	
	[70]	[1159]	[1229]	
Share of teachers with 1-2 years of tenure	0.291	0.263	0.073	
	(0.174)	(0.143)	(0.054)	
	[70]	[1159]	[1229]	
Share of teachers with 3-4 years of tenure	0.266	0.305	-0.073	**
	(0.141)	(0.171)	(0.033)	
	[70]	[1159]	[1229]	
Share of teachers with 5 years or above of tenure	0.343	0.312	0.036	
	(0.208)	(0.213)	(0.030)	
	[70]	[1159]	[1229]	
Joint test of significance:		F(12,845) =	2,489	***

Notes: Table shows variable-by-variable means and standard deviations (columns 1-2, standard deviations in parentheses and number of schools in squared brackets) and regression coefficients and standard errors (column 3) from regressions of the treatment indicator on each variable separately. In the bottom, we present an F-test for joint significance from a regression with all variables. Standard errors in regression are robust. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix B: Heterogenous effects of PERSPEKT 2.0

In this appendix we describe how we explore the existence of heterogenous effects of PERSPEKT 2.0 using the methods developed by Chernozhukov, Demirer, Duflo and Fernandez-Val (2018), henceforth CDDF. We refer to CDDF for all technical details.

In the main paper we present results for a range of pre-specified subgroups. However, the choice of subgroups is delicate. On the one hand, choosing subgroups ex-ante as we have done, amounts to throwing away a large amount of potentially valuable information. On the other hand, choosing subgroups ex-post opens the possibility of overfitting and raises the problem of multiple hypothesis testing.

CDDF propose a disciplined strategy to ex-post discover any relevant heterogeneity in treatment effects by covariates (called conditional average treatment effects, CATE) while avoiding the risk of overfitting. The strategy builds on machine learning tools and require little guiding principles on which covariates that are likely to be relevant. The strategy makes it possible to find out if there is detectable heterogeneity in the treatment effect based on observables, and if there is any, what is the group average treatment effect for different bins of the effect size distribution. Finally, it is possible to make a characterization of the most and least affected groups based on covariates. Specifically, the strategy includes building a machine learning proxy predictor of CATE used to develop valid inference on three objects: first, the Best Linear Predictor (BLP) of CATE; second, the Sorted Group Average Treatment Effects (GATES) or average treatment effects by heterogeneity groups; and third, the Classification Analysis (CLAN) or the average characteristics of the most and least affected individuals. Estimation and inference rely on repeated data splitting to avoid overfitting and achieve validity. The inference methods account for both usual estimation uncertainty and splitting uncertainty by reporting the medians of the estimated key features over different random splits of the data. Similarly, confidence intervals and p-values are constructed by medians over the many random splits.

To implement the method by CDDF we build on their R-script available at co-author Mert Demirer's GitHub-page. For the proxy predictor of CATE we follow CDDF and employ four different machine learning methods: Random forest; elastic net; boosting; and neural network.

We run all analyses using the absolute social well-being score as outcome measure. As covariates we employ an extended list of the variables presented in Table A2, we standardize all variables and include missing variable indicators.

Table B1 presents two metrics that allow for comparison across the four machine learning methods using the same notation as CDDF. Loosely speaking, the metrics can be understood as the correlation between the machine learning proxy and the true values for the BLP and the GATES, respectively. We see that the elastic net method outperforms the three other methods, and we will focus on results from the elastic net in the following.

Table B1: Comparisons of ML methods

	Elastic net	Boosting	Neural net	Random Forest
Best BLP(Λ)	0.008	0.005	0.005	0.008
Best GATES($\overline{\Lambda}$)	0.077	0.042	0.042	0.061

Note: The table shows metrics for targeting the best BLP and best GATES, respectively. Medians over 100 splits. Formula can be found in CDDF.

Next, we show the results of the BLP of CATE using the Elastic net proxy in Table B2. We report the average treatment effect (ATE) and the heterogeneity loading (HET) parameters, respectively.

Table B2: Average treatment effect and heterogeneity loading from BLP using Elastic net

	ATE	HET
Social wellbeing	-0.031	0.515
	(-0.086,0.024)	(0.131,0.877)
	[0.543]	[0.018]

Note: The table shows median parameter estimates over 100 splits. 90 % confidence intervals in parentheses and p-values from testing against the null hypothesis are in square brackets. Formula can be found in CDDF.

The average treatment effect is very similar to the OLS estimate presented in the main paper and likewise not significantly different from zero. However, the heterogeneity parameter suggests that there are significant differences in the impact of PERSPEKT 2.0.

Hence, we calculate the group average treatment effect for the most and least affected groups as presented in Table B3. The estimated impacts of PERSPEKT 2.0 are not significantly different from zero, neither for the most nor least affected group. The last column presents the difference between the most and least affected groups, which is significantly different from zero. However, as none of the group average treatment effects are significantly different from zero, we conclude that PERSPEKT 2.0 has not consistently affected the social wellbeing of any larger subgroups of pupils.

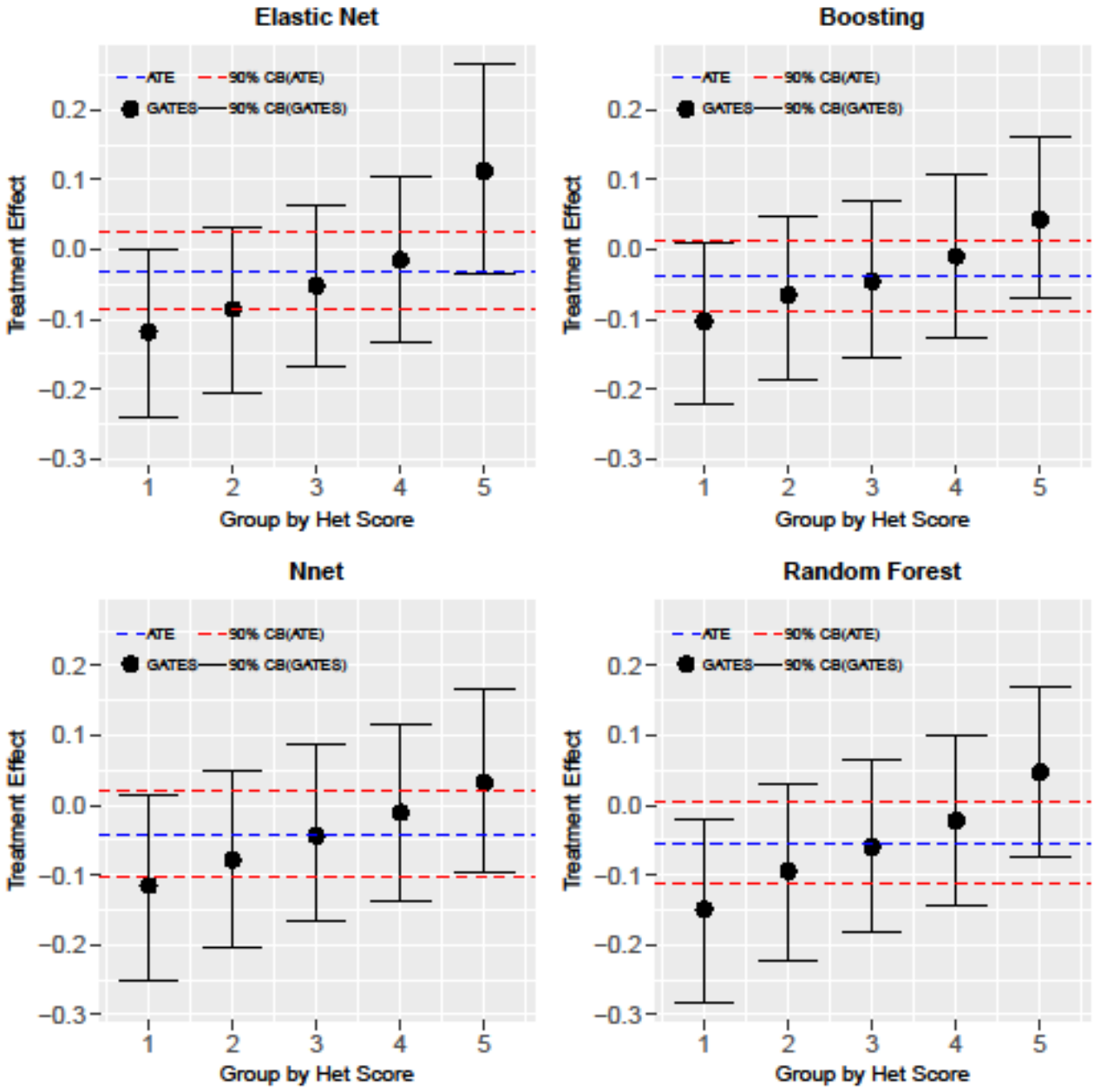
Table B3: Sorted Group Average Treatment Effects and difference

	Most affected	Least affected	Difference
Social wellbeing	0.112	-0.099	0.214
	(-0.012,0.239)	(-0.203,0.002)	(0.051,0.374)
	[0.155]	[0.111]	[0.021]

Note: The table shows sorted group average treatment effects for the 20 percent most and least affected pupils using the elastic net method as machine learning proxy. The last column shows the difference between the two groups. 90% confidence intervals are in parentheses and p-values from testing against the null hypothesis are in square brackets.

Finally, we summarize the results from all four machine learning proxies in Figure B1 with the Elastic Net as the preferred method in the top left corner. The figure presents the GATES for the five heterogeneity groups (black dots) along with average treatment effect for the entire sample (blue dashed line). The "Group 1" category is the 20% of students who are affected most negatively by PERSPEKT 2.0, while "Group 5" contains the top 20% most positively affected. The general picture is that we cannot reject the hypothesis that PERSPEKT 2.0 has no impact on pupil social wellbeing.

Figure B1: GATES from all four machine learning methods



As we do not find significant impact of PERSPEKT 2.0 for any of the heterogeneity groups using the best machine learning proxy, we do not present CLAN results comparing characteristics of the most and least affected groups. These results are available upon request.