

Schmitz, Laura

Working Paper

Heterogeneous effects of afterschool care on child development

DIW Discussion Papers, No. 2006

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Schmitz, Laura (2022) : Heterogeneous effects of afterschool care on child development, DIW Discussion Papers, No. 2006, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/261399>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

2006

Discussion
Papers

Heterogeneous Effects of After-school Care on Child Development

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

IMPRESSUM

DIW Berlin, 2022

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
<https://www.diw.de>

ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
<https://www.diw.de/discussionpapers>

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
<https://ideas.repec.org/s/diw/diwwpp.html>
<https://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html>

Heterogeneous effects of after-school care on child development

Laura Schmitz[†]

DIW Berlin, BSE & FU Berlin

This version: June 2022

Abstract

It is often argued that institutionalized after-school care (ASC) can benefit children lacking adequate homework support at home and, hence, foster equality of opportunity. However, despite considerable policy interest, it is unclear whether these afternoon programs are beneficial for child development and if selection into them is efficient, i.e., whether students benefiting most from the programs choose to attend. In this paper, I examine the effects of ASC on elementary school children's schooling outcomes and non-cognitive skill development. Using a marginal treatment effect framework and regional and temporal variation caused by an extensive reform in Germany, I instrument after-school care attendance with the change in the distance to the next school offering ASC within one district. My findings suggest that children from lower socioeconomic backgrounds, who more often select into treatment, have higher ASC premiums. Concerning schooling outcomes, I find minor positive local average treatment effects but no effect heterogeneity concerning unobserved characteristics. ASC effects on the treated's non-cognitive skills are more sizable than those on the untreated, suggesting that selection into ASC is positive and efficient. Overall, a universal voluntary offer of ASC will likely help reduce educational inequalities.

Keywords: after-school care, marginal treatment effects, inequality

JEL: I21, I24, I26, J08

[†]E-mail: *Email: lschmitz@diw.de*. This paper benefited greatly from comments by Katharina Spieß, Felix Weinhardt, Matthias Westphal, Sönke Matthewes, Jan Stuhler, Jan Marcus, Mara Barschkett, Jonas Jessen and Mathias Huebener, as well as participants of the IWAAE Conference 2022, EffEE Conference 2022, UCL SEHO Conference 2022, LEER Conference 2022, CRC Conference 2021, CIDER Conference 2021, the BeNA summer workshop 2021 and the DIW Graduate Center (GC) summer workshop 2021 for valuable comments on drafts and presentations. I gratefully acknowledge financial support from the DIW GC and BERA through Jan Marcus, as well as technical support by the team of the SOEP infrastructure on using the geo-referenced SOEP data.

1 Introduction

After-school care (ASC) programs are a central element in the attempts of many OECD countries to meet the increased demand for institutionalized child care while simultaneously fostering children’s cognitive and social development (OECD, 2017). A particular hope of ASC lies in its potential equalizing effect: Children with low socioeconomic status (SES) tend to experience lower-quality homework support at home (e.g. Buckingham et al., 2013, van Bergen et al., 2017) and hence are likely to benefit the most from spending additional time in school and afternoon care (e.g., Blau and Currie, 2006, Angrist et al., 2010, Levine and Zimmerman, 2010 Plantenga and Remery, 2015). Following these considerations, the German government heavily subsidized the expansion of ASC in elementary schools after 2003 (BMBF, 2009). Despite immense policy interest, evidence on the causal effects of universal afternoon programs on elementary school children is scarce.

This paper studies the heterogeneous effects of ASC on elementary school child development. My main interest is to understand how the selection into afternoon programs relates to the treatment effect and whether a universal offer of ASC reaches the right children. I use a unique combination of self-collected school-level data from six Western German Federal States between 2003 and 2018 with student-level data from the German Socio-Economic Panel (SOEP). By analyzing effects on grades in math and German, transition to the academic track, the Strengths and Difficulties Score (SDQ), prosocial behavior, and the Big Five personality traits, I cover key aspects of cognitive and non-cognitive development.

I employ a marginal treatment effects (MTE) framework (Björklund and Moffitt, 1987; Heckman and Vytlačil, 1999; Heckman and Vytlačil, 2005; Heckman, Urzua, et al., 2006), which is uniquely suitable for evaluating a policy not only in its efficacy and equity but also in the efficiency of the selection mechanism at play —especially in the presence of unobserved self-selection (e.g. Carneiro et al., 2011; Cornelissen et al., 2018; Felfe and Lalive, 2018). So far, this method has been applied to early and pre-school daycare (Felfe

and Lalive, 2018, Cornelissen et al., 2018) and higher education (e.g., Carneiro et al., 2011, Redmond, 2014, Kaufmann, 2014, Carneiro et al., 2017, Kamhöfer et al., 2019) but not to elementary school children, leaving an "age gap" in our understanding of how selection into educational offers relate to treatment effects.

The elementary school years are a particularly relevant period for several reasons: The effectiveness of (high quality) early childhood and elementary school interventions on improving later-life outcomes is well documented (e.g. Cunha and Heckman, 2007, Kautz et al., 2014). It is also shown that elementary school is a particularly sensitive period for the development of motivations, beliefs, and behaviors (e.g. Kosse et al., 2020). Furthermore, besides marking the beginning of a child's school career, performance during this period builds the basis for tracking into different secondary school types after fourth grade in Germany —hence determining a child's educational and professional trajectory in important ways (e.g. Dustmann, 2004, Dustmann et al., 2017). Thus, the students at the center of my study are in a critical and highly malleable phase of their development.

I make use of arguably exogenous variation in available ASC slots caused by a large reform in Germany that led to an increase in the proportion of primary school children in Germany participating in afternoon care activities at school from roughly ten to close to 70 percent since 2003 (KMK and of the Federal States in the Federal Republic of Germany, 2021). The afternoon care at school consists mostly of homework support and supervised recreational activities; it does not include an increase in instruction time.¹ The country-wide reform was staggered across the federal states and generated regional and time variation in the availability of publicly available ASC slots, hence creating the ideal setting for estimating MTEs. The combination of the SOEP and the administrative school-level data allows me to both observe yearly individual ASC attendance and retrieve the distance to the nearest ASC from the students' home. Controlling for time, district, and cohort fixed effects, the change in the distance over time within the same district

¹In some cases, in the integrated ASC type, instruction time is shifted from the morning to the afternoon. However, this shift does not lead to an increase in the absolute instruction time.

builds the continuous instrument needed to estimate MTEs.

I find substantial heterogeneity in returns to ASC with respect to both observed and unobserved characteristics. Low-SES children² are more likely to attend ASC and experience higher returns in terms of (non-)cognitive development, which points to a positive selection based on observed characteristics. The selection of unobserved characteristics reinforces this finding since children with lower resistance to attending ASC are more likely to benefit from afternoon care. For most of my outcomes in the area of non-cognitive skills, the MTE curve indicates a higher treatment effect for treated individuals than for non-treated. For these outcomes, the average treatment effect (ATE) is either not statistically different from zero or negative, whereas the average treatment effect on the treated (ATT) is positive in all outcomes except for conscientiousness —and statistically significant for prosociality and agreeableness. Hence, while ASC does not benefit everyone, it seems to have beneficial effects on those who select into them, especially in terms of outcomes that broadly categorize as social skills.

My paper contributes to different strands of the literature. First, it adds to our understanding of the effects of afternoon supervision of school-age children (e.g. Aizer, 2004, Blau and Currie, 2006, Felfe and Zierow, 2014, Seidlitz and Zierow, 2020). Most literature in this area focuses either on targeted programs (Blau and Currie, 2006) or on non-specified adult supervision (Aizer, 2004). To my knowledge, Felfe and Zierow (2014) and Seidlitz and Zierow (2020) are the only two economic studies evaluating the impact of universal after-school center-based care on children,³ also using variation caused by the German reform. Their results are inconclusive regarding the global and heterogeneous

²Specifically, I find that children of less-educated parents, social transfer receiving households, lower-income families, single-parent households and with a migration background more often select into ASC. Since all of these characteristics correlate with lower SES, I broadly characterize selection into ASC this way.

³So far, economic studies focus mainly on the effects of the expansion of ASC in Germany on maternal labor supply, e.g. Gambaro et al. (2019), Dehos and Paul (2017), Nemitz (2015)). A number of studies evaluate the reform in terms of their non-causal effects on students, e.g. the StEG study on the development of all-day schools (Konsortium, 2016), Lossen et al. (2016) and Sauerwein et al. (2019)).

effects of ASC.⁴ I add to these findings by revealing important heterogeneity patterns with highly relevant policy implications, and by examining a larger variety of outcomes in cognitive and non-cognitive skill development.

Second, my study adds to our knowledge on personality and non-cognitive skill formation (see, e.g. Heckman, Stixrud, et al., 2006, Brunello and Schlotter, 2011, Fletcher, 2013, Kautz et al., 2014, Deming, 2017) all pointing to (early) childhood as the most critical investment period for developing social skills (e.g. Cunha and Heckman, 2007, Kautz et al., 2014 Kosse et al., 2020). In this sensitive period, the effects of policy interventions can vary substantially depending on the policy design. While (good quality) daycare is generally believed to be beneficial for non-cognitive development (e.g. Heckman et al., 2013), there is no linear or unconditional positive link. For example, full-day care at a daycare center in comparison to half-day care can have detrimental effects on children's socio-emotional well-being, with this result being driven by children from disadvantaged families (Felfe and Zierow, 2018). Baker et al. (2019) find negative long term effects of universal childcare attendance on non-cognitive child outcomes such as self-reported health and life satisfaction among teens in Quebec, with the negative effects presumably being driven by quality issues (Currie and Almond, 2011). Early entry into childcare seems to exhibit no effects in the short term (Kuehnle and Oberfichtner, 2020) but is shown to increase extroversion in adolescence (Bach et al., 2019). I add to this body of literature by examining the effects of full-time care in elementary school on the formation of personality (the Big Five), social skills (prosocial behavior), and socio-emotional development (the SDQ).

Finally, I add to our understanding of how selection into educational offers relate to treatment effects at different educational stages. I contribute to the growing body of literature

⁴Felfe and Zierow (2014), using a value-added approach, find no significant effects on average but positive effects on children of less-educated mothers and low-income families in terms of their socio-emotional development. In contrast, Seidlitz and Zierow (2020), employing an IV approach with treatment defined on the school level, find a positive overall effect on language and math skills as well as on the probability of being recommended for the academic track —but no evidence for an equalizing effect of ASC.

employing MTEs to identify policy-relevant treatment effect heterogeneity in educational interventions. The MTE literature on educational interventions counts notable contributions for both very young children and for young adults, revealing a puzzle on selection patterns into voluntary educational offers: On the one hand, Cornelissen et al. (2018) and Felfe and Lalive (2018), looking at early daycare at the age of 3-6 and below 3, respectively, find that children from disadvantaged backgrounds are less likely to attend but have larger positive treatment effects in terms of school readiness (Cornelissen et al., 2018) and socio-emotional skills (Felfe and Lalive, 2018) —with this pattern of reverse selection being reinforced by selection on unobserved characteristics.⁵ In contrast, MTEs estimated for higher education, where selection is mainly driven by the preferences of the individuals themselves rather than their agents, suggest a positive selection pattern, implying that individuals with a higher gain from treatment are more likely to attend and have larger gains from treatment (e.g. Carneiro et al. (2011), Redmond (2014), Kaufmann (2014), Kamhöfer et al. (2019), Westphal et al. (2020) for tertiary education Carneiro et al. (2017) for higher secondary education).⁶ These contrasting findings raise the question of when the negative selection observed for early daycare turns into the positive selection found for higher education. This paper helps to fill this research gap by estimating MTEs for the close-to universal and mostly voluntary ASC offer for elementary school children. In contrast to the literature on SES gaps in early daycare (Jessen et al., 2020), research on the selection into ASC in Germany (e.g. Marcus et al., 2016, Gambaro et al., 2020) suggests a less straightforward selection pattern.⁷ Hence, it is not clear whether children with a higher propensity to enroll in afternoon care in elementary school are more or less

⁵The two papers, however, come to different conclusions regarding the effects of child care on motor skills: while Cornelissen et al. (2018) find no statistically significant heterogeneity pattern with respect to this outcome, Felfe and Lalive (2018) finds a positive selection pattern here, i.e. children who more readily select into child care benefit more. This divergence could stem from the differences in the observed age groups.

⁶Nybom (2017) also finds significant self-selection into college but primarily on observed characteristics.

⁷On the one hand, disadvantaged groups, like children of migrants and single-parent households, as well as, on the other hand, children whose mothers have an academic degree have a higher probability to attend ASC.

likely to benefit from the extra hours than their peers who go home after the regular classes are over.

My findings have two important policy implications: First, spending more time in school does indeed seem to benefit low-SES students and, hence, the universal offer of ASC slots can serve as a tool to increase equality of opportunity. Second, with different organizational types of ASC currently co-existing in Germany and beyond, these results make a strong case for organizing ASC in a non-integrated way, i.e. offering afternoon slots on a voluntary basis instead of making participation mandatory for all pupils. However, the current dynamic in Germany —primarily low-SES students taking up these offers—bears the risk of increased segregation, where afternoon programs might become increasingly less attractive for high-SES students. Given the importance of peer effects on students' achievement (e.g. Sacerdote, 2001, Lavy et al., 2012, Feld and Zölitz, 2017, Golsteyn et al., 2021), group composition should be of concern for schools and policymakers.

My study design offers several advantages over most other studies estimating MTEs to evaluate educational interventions. Most importantly, I use variation in the instrument not only cross-sectionally but using panel data over 16 years, hence using variation in the access to ASC across time and space, which allows me to control for time-constant unobserved district characteristics. In addition, using survey data representative of more than 70 percent of the German elementary school-age population⁸ rather than focusing on a narrow geographical area, my results likely have strong external validity and are informative for educational policies in Germany and beyond. Finally, I observe a variety of outcome and control, allowing me to (i) draw a comprehensive picture of the effects of afternoon programs on child development, and (ii) to narrow down better what is captured by the error term of the selection equation, which is important for interpreting the heterogeneity based on unobserved characteristics.

⁸I cover the six Western German federal states of North Rhine-Westphalia, Baden-Wuerttemberg, Bavaria, Hesse, Rheinland Palatinate and Lower Saxony, which together make up 73.8 percent of the German population. Therefore, the 8,274 students in my SOEP sample are roughly representative of 73 percent of the German population of elementary school children

2 Institutional background and mechanisms

2.1 Elementary schools in Germany and the IZBB reform

Since education policy in Germany is decentralized and regulated by the federal states, some aspects of the primary education system vary across states. Compulsory primary education starts when children are around six years old and, in most cases, lasts for four years.⁹ Hence I focus on children between the age of six/seven and ten/eleven. Based on their performance in fourth grade, children are then divided into three secondary school tracks: basic track (five years), middle track (six years), and higher track (eight to nine years), with only the latter granting access to universities. The recommendation is given by the headteacher but, in most cases, this recommendation is not strictly binding. The rules again differ by the federal states.¹⁰ Once assigned to a track, mobility across tracks is rare, with upward mobility, i.e. moving from the lower to the higher track, being especially difficult (e.g. Dustmann, 2004, Bellenberg, 2012, Dustmann et al., 2017). Hence, performance in primary schools greatly affects the children's future educational and professional paths.

Primary schools in Germany were traditionally designed as half-day schools, i.e., they started at 8 a.m. and ended before 1 p.m. While institutionalized afternoon care existed in the form of *Horts*, their offer and take-up significantly differed across rural and urban areas as well as Eastern and Western Germany. In Western Germany, merely six percent of 6- to 9-year-old children attended after-school care in 2002, compared to around 60 percent in Eastern German states (DJI, 2008). My analysis shows that 30 percent were cared for by family members or family friends, and two-thirds of the children were taken

⁹This is the case for all six federal states that I focus on in my analysis: North Rhine-Westphalia, Bavaria, Baden-Wuerttemberg, Hesse, Lower Saxony, and Rhineland-Palatinate.

¹⁰In North Rhine-Westphalia between 2006 and 2010 and in Bavaria for the whole observation period, the teacher's recommendation was binding. Children whose parents disagreed with the recommendation had the opportunity to attend three-day *trial lessons* after which they had to pass exams in German and mathematics with certain grades (Bavarian Ministry for Education and Cultural Affairs, 2010, Ministry of Education North Rhine-Westphalia, 2012). While, in most states in my sample, the teacher's recommendation is not binding, it is usually the case that parents follow the teacher's recommendation (Bos, 2003).

care of by their parents, typically mothers.¹¹ The stark differences in afternoon care for school-age children pre-2003 questions a joint analysis of Western and Eastern Germany. Therefore, this paper focuses only on Western German federal states.

Germany began reforming the half-day elementary school system in the early 2000s, reacting to both an increased demand for afternoon care due to higher maternal labor market participation and calls for improving student outcomes and equality of opportunity in the German school system.¹² The "Investition Zukunft Bildung und Betreuung" (future of education and care) investment program (IZBB) subsidized the expansion of ASC with 4 billion euros between 2003 and 2009. More than half of these funds were spent on elementary schools (BMBF, 2009).¹³¹⁴ As a consequence of the IZBB program, the number of primary schools running afternoon programs grew significantly. In the 2018/19 school year, 67.5 percent of all primary schools in Germany offered ASC (KMK, 2018). The majority (65.8 percent) of primary ASC programs are organized in a non-integrated way (*offene Ganztagschule*), where participation in lunch and afternoon programs is voluntary. The strictest form of ASC is the integrated type (*gebundene Ganztagschule*), where every student attending the respective elementary school is obliged to participate in these afternoon activities.¹⁵ In my sample, 75 percent of ASC programs are non-integrated, 19 percent are partly integrated, and six percent are integrated. Parents can apply for the afternoon program for their children each year, paying a moderate fee that covers the lunch costs. Including the afternoon program, the median time of supervision amounts

¹¹Numbers based on own calculations using the Socio-economic Panel v35.

¹²The so-called *PISA Shock* ranked German schools below the OECD average in the 2000 Program for International Student Assessment (PISA) study. Besides the overall disappointing performance, Germany also stood out for a high level of educational inequality measured by a strong link between achievement and family background.

¹³The invested funds, however, were to be spent on construction purposes only. Hence, the states had to cover running costs such as personnel costs, which, varying with the intensity of the program, are estimated to range between 992 and 1981 Euros per child per year (Klemm and Zorn, 2017).

¹⁴The expansion of the ASC sector continued beyond the end of the IZBB investment program, albeit at a slower pace. In spring 2021, the Federal Government announced it would invest a further 3.5 billion to grant a legal entitlement to all-day care for children of primary school age from 2026 onwards (BMBF, 2021).

¹⁵A third type is the semi-integrated ASC program (*teilgebundene Ganztagschule*), where attendance in the afternoon is obligatory only on certain days of the week.

to 8.5 hours per weekday (Fischer et al., 2013), compared to the usual average school day of 4.5 hours. Figure A.1 maps the geographical expansion of ASC in my sample.

ASC in Germany consists to a large extent of lunch, homework support, and supervised recreational activities. In some cases, the integrated ASC type includes instruction time in the afternoon. However, in these cases, instruction time is shifted from the morning to the afternoon rather than added to the standard curriculum. Hence students attending ASC do not receive more instruction than their peers. Quality standards are defined by the federal states and are not uniform for all of (West-)Germany. Around three-quarters of the staff in afternoon care have a non-university pedagogical formation over three years or hold a university diploma in social pedagogy. In Bavaria and Lower Saxony, the share of staff with only a short-term formation in child care (less than three years) is relatively high, at 26 and 15 percent, respectively. The average children-per-staff ratio is similar in all six federal states in my sample, ranging between 9 and 10 (Felfe and Zierow, 2014).

2.2 Mechanisms

The nexus between selection into ASC and its heterogeneous effects on student's (non-)cognitive development is mainly determined by two factors: (I) Who benefits from the substitution of home care with institutionalized care, and (II) who is more likely to participate in terms of observed and unobserved characteristics. If the first and second group coincide, selection into ASC is positive and efficient. If further, this group consists predominantly of disadvantaged students, ASC can be said to have an equalizing effect.

Regarding (I), the estimated effects of the ASC expansion should be evaluated as the result of substituting home-based care with institutionalized care. The former differs from the latter in three main aspects: spending more time with a relatively large group of peers, the presence of at least one adult with pedagogical training, and a predetermined program of supervised recreational activities (Felfe and Zierow, 2014). Increased peer interactions could be beneficial for children's socio-emotional development.¹⁶ Due to the vital link be-

¹⁶For example, prosocial behavior is found to be highly malleable at elementary school age, with social

tween non-cognitive and cognitive skills development (e.g., Cunha and Heckman, 2007), a positive effect on schooling outcomes is also plausible. In addition, more peer time could be beneficial for the language skills of children of immigrant ancestry (Washington-Nortey et al., 2020). On the other hand, considering self-control as a scarce resource (Baumeister and Alquist, 2009), adverse effects of ASC attendance on non-cognitive development are possible, especially for children with a predisposition for conduct or attention difficulties. Because of the large group size and the often short training period of the ASC staff, it is unclear whether the second point, the presence of a trained pedagogue, has a significant impact on an average child attending an afternoon program, especially on children with special needs. However, for children who do not receive (high-quality) homework support at home, the offered support could make a substantive difference (Felfe and Zierow, 2014). Hence, heterogeneity in the effects of ASC is likely driven by differences in the counterfactual care mode. In the same fashion, the offer of regular supervised sports, cultural events, and other educational activities plausibly makes a larger difference for low-SES children since they have fewer resources and less often take up these offers outside of school (e.g. Hille and Schupp, 2015, Whitaker et al., 2019).

Regarding (II), as shown by Marcus et al. (2016), selection into ASC is complex —albeit, for Western Germany, it is more clearly low-SES children that are selecting into ASC. In addition to selection based on levels, the treatment decision is likely affected by the expected idiosyncratic return to treatment, causing selection based on gains: children who choose treatment because they have a low resistance to it might have different gains than those with high resistance. For example, more motivated children may voluntarily stay in school in the afternoon. This would likely result in a positive selection based on the latent propensity to attend (similar to the MTE literature on higher education, e.g., Carneiro et al., 2017). On the other hand, selection based on gains could be induced by persons other than the children themselves, e.g., their parents and teachers. Parents could decide to send their kids to ASC due to their unobserved preference for working full

interactions being one of the main drivers (Kosse et al., 2020).

time. Suppose this mechanism is the dominant one at a place. In that case, there could be a reverse selection similar to that identified for early daycare by Cornelissen et al. (2018) and Felfe and Lalive (2018) since full employment is associated with a high SES and these children might benefit less from ASC because of the high-quality support they would otherwise receive at home. Further, it is possible that teachers specifically target low-SES students for ASC by convincing the parents to register because they think that these children would benefit most from the extra homework support. This would add to a positive selection pattern if low-SES students benefit more from the services offered by ASC, as hypothesized in (I).

While ASC quality differs and is not ideal, the environment encountered at ASC is much more homogeneous than the counterfactual environment at home. Hence, ASC likely has an equalizing effect, with differences in effect size driven predominantly by differences in the counterfactual care mode (Felfe and Zierow, 2014). In addition, since low-SES children in Western Germany have a higher probability of selecting into ASC, a positive selection mechanism seems to be in place. Whether this pattern is reinforced by a positive selection based on unobserved characteristics is to be determined.

3 Research design

3.1 Model

I study the effect of after-school care on elementary school students' skill development. In a simple OLS setting, the model would look like this:

$$Y_i = ASC_i\beta_1 + X_i\beta_2 + \epsilon_i \tag{1}$$

where Y_{ji} is the outcome of interest for individual i , ASC_i is the treatment dummy, which takes value one if the individual attended an afternoon program for most of their

elementary schooling,¹⁷ X_i is a vector of individual and household covariates, and ϵ_{ij} is the error term.

The selection mechanisms described in the last chapter pose a critical problem for identifying β_1 , my parameter of interest since individuals who select themselves into treatment differ from the control group in systematic ways that affect the outcome Y_{ji} . Instrumental variable (IV) techniques can solve selection on levels by estimating local average treatment effects (LATE) for instrument compliers. However, in most settings, selection mechanisms will not solely be based on observed characteristics but also on what the individuals expect to be their gain from treatment, i.e., on a certain degree of "resistance to treatment" unobserved by the researcher (Cornelissen et al., 2016, Andresen, 2018, Zhou and Xie, 2019). The Marginal Treatment Effect (MTE) framework developed by Heckman and Vytlacil (1999, 2001a, 2005, 2007b) allows for studying heterogeneous treatment effects in the presence of self-selection. MTEs identify the average treatment effects (ATE) for people with particular resistance to treatment, allowing to recover economically relevant parameters, like the average treatment effects on the treated (ATT) and the average treatment effect on the untreated (ATUT), given full instrument support (Andresen, 2018).

In the baseline model, let Y_{1i} denote the outcome of student i in the case of treatment, i.e., attending afternoon programs for most of elementary school ($D_i = 1$), Y_{0i} the outcome of student i if he or she attends elementary school only half-day ($D_i = 0$), X_i a vector of observed child and household characteristics, while R_i , T_i and C_i are districts, survey year and birth year fixed effects, respectively:

$$Y_{ji} = X_i\beta_j + R_i\alpha + T_i\delta + C_i\gamma + U_{ji}, \quad j = 0, 1 \quad (2)$$

¹⁷This variable is derived by: $D_i = mode_i(ASC_{it})$ and is equivalent to at least two years for most students in my sample, and to at least one year when I observe a child for only two years during elementary school, as is the case for a part of my sample in the latest years (2017-2018)

I use the following latent index model for selection into treatment D_i :

$$D_i^* = Z_i\beta_d - V_i \tag{3}$$

$$D_i = \begin{cases} 1 & \text{if } D_i^* \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

where $Z = (X, R, T, C, Z^*)$ is equivalent to X in equation (2) but additionally contains an instrument Z^* . I rely on spatial and temporal variation in the offer of ASC slots caused by the IZBB reform by using the change in the distance to the next ASC offering school within one district as an instrument for ASC attendance (see section 4.2). The term V_i represents the “unobserved resistance” to the treatment of individual i , capturing all unobserved characteristics that lower the probability of attending ASC (Cornelissen et al., 2016).

The treatment effect $Y_{1i}-Y_{0i}$ may vary among students with different observed characteristics X as well as among those with different values of the unobserved components U_1 and U_0 —who may have the same characteristics X . In order to trace the dependence between the treatment effect and the unobserved component of the treatment choice, I rely on the quantiles of the distribution of V , as is common in the MTE literature:

$$Z\beta_d - V \geq 0 \Leftrightarrow Z\beta_d \geq V \Leftrightarrow \Phi(Z\beta_d) \geq \Phi(V) \tag{4}$$

with $\Phi(V)$ denoting the c.d.f. of V . $\Phi(Z\beta_d)=P(Z)$ is the propensity score of attending ASC based on observed characteristics. The MTE as a function of these quantiles can then be expressed as:

$$MTE(X = x, U_D = u_D = p(Z)) = E(Y_i - Y_0 | X = x, U_D = u_D = p(Z)) \tag{5}$$

where the MTE is the return to treatment for an individual with observed characteristics $X = x$, who is in the u_D th quantile of the V distribution, which is equal to their propensity to treatment. The treatment effect at low values of u_D is the effect for students who have a low unobserved resistance, i.e., are eager to attend ASC. A weighted average of the MTEs then yields estimates of the ATE, ATT, and ATUT.¹⁸

3.2 Assumptions

The following assumptions are necessary for identification: First, there needs to be a first stage in which the instrument Z^* (the change in the Euclidean straight-line distance to the nearest ASC in km) causes variation in the probability of treatment after controlling for (X, R, T, C) . Section 4.2 and Table 1 present evidence that the distance to the nearest ASC indeed has a strong and significant effect on ASC enrollment after controlling for individual characteristics, district, year, and cohort fixed effects. Second, Z^* must be independent of the unobserved component of the outcome and selection equation. That is, $Z^* \perp\!\!\!\perp (U_0, U_1, V) | (X, R, T, C)$. This assumption requires that the instrument is assigned as good as randomly, depending on (X, R, T, C) . In addition, this implies the exclusion restriction that the distance to the nearest ASC must not directly affect the outcome conditional on D_i and (X, R, T, C) . It further implies that how U_1 and U_0 relate to V (i.e., The MTE curve) must not depend on Z^* . These two first assumptions, along with the monotonicity assumption, are virtually equivalent to the standard assumptions necessary to interpret an IV as LATE (Andresen, 2018, Angrist and Imbens, 1995). Section 4.2 is dedicated to defending these assumptions.

In the ideal case of full support of the propensity score in both treated and untreated samples for all values of X , it is possible to estimate MTEs with no further assumptions (Carneiro et al., 2011, Cornelissen et al., 2016). In practice, however, this is rarely achieved, especially if X is a high-dimensional vector (e.g. Cornelissen et al., 2018, Carneiro et al., 2011, Carneiro et al., 2017, Felfe and Lalive, 2018). Hence, in the MTE

¹⁸Consult Cornelissen et al. (2016) for a derivation of the respective weights

literature, it is common to make a further assumption: $E(U_j|V, X) = E(U_j|V)$ This assumption, called the *separability assumption*, allows for identifying the MTE over the unconditional support of the propensity score, jointly generated by the instrument and the covariates, as opposed to the support of the propensity score conditional on $X=x$. This assumption has two implications: first, that the shape of the MTE - the manner in which U_0 and U_1 depend on V - is independent of X . Second, that the MTEs are additively separable into an observed component X and unobserved component U_D :

$$Y_{1i} - Y_0 = X_i(\beta_1 - \beta_0) + U_{1i} - U_{0i} \quad (6)$$

Unfortunately, to my knowledge, no testing procedure is available to provide evidence for the additive structure of the MTE (Su et al., 2015). Therefore, I proceed by inspecting how much my data deviates from the ideal case described above and by relying on the separability assumption common in the applied literature.

3.3 Estimation

Using $p = P(Z)$, the separability assumption and taking expectations, I deduce the following equation:

$$\begin{aligned} & \mathbb{E}[Y|X = x, R = r, T = t, C = c, P(Z) = p] \\ &= X\beta_0 + R\alpha + T\delta + C\gamma + X(\beta_1 - \beta_0)p + \underbrace{p \mathbb{E}(U_1 - U_0|U_D \leq p)}_{K(p)} \end{aligned} \quad (7)$$

where $K(p)$ is a nonlinear function of p capturing heterogeneity across U_D . Taking the derivative of this expression with respect to the propensity score p yields the MTE:

$$\frac{\partial \mathbb{E}[Y|X = x, R = r, T = t, C = c, P(Z) = p]}{\partial p} = X(\beta_1 - \beta_0) + \frac{\partial K(p)}{\partial p} \quad (8)$$

I start by identifying the selection equation employing a probit model to obtain estimates of the propensity score $\hat{p} = \Phi(Z\beta_d)$. In a second step, I need to assume the unknown shape of $K(p)$ by choosing a polynomial in p of degree k to estimate the outcome equation:

$$Y = X\beta_0 + R\alpha + T\tau + C\gamma + X(\beta_1 - \beta_0)p + \sum_{k=2}^K \alpha_k \hat{p}^k + \epsilon \quad (9)$$

I assume a second-order polynomial in \hat{p} in my baseline specification but find similar results for $K=3$, $K=4$, a joint normal and a semiparametric specification of $K(p)$ (see section 5.4).

4 Data and instrument

4.1 Data set description

The empirical analysis combines two individual-level data sets, the German Socio-Economic Panel (SOEP, see the Goebel et al., 2019) and a self-collected school data set, as well as information on the district and municipality level.¹⁹

Student-level data

My estimates use data from the SOEP, a nationally representative survey that started in 1986. Nearly 13,000 households and more than 30,000 individuals are surveyed each year, gathering information about respondents' demographics, household composition, educational outcomes, and labor market characteristics (Goebel et al., 2019). The SOEP

¹⁹I use the *INKAR* database for SES data on the district and municipality level, which mainly serves as control variables and for the balancing test in 4.2.1. Districts correspond to the NUTS 3 definition and there are currently 400 districts in Germany. With currently 11,130 items, municipalities represent an even smaller territorial unit.

is particularly suitable for my research questions because it has data on both attendances of ASC alongside detailed information on the individual and family background of the children. It also comprises a large set of interesting outcome parameters that allow for comprehensively assessing the effects of ASC, covering variables in the domain of (non-)cognitive and social skills. In addition, the SOEP contains several geographically referenced indicators and detailed regional information such as the community type²⁰ and community size classification, which serve as control variables.²¹

The 2003-2018 SOEP data contains several special surveys, such as the M1 and M2 Migration Sample and the M3 Refugee sample.²² All estimates include individual weights to avoid the oversampling of these particular groups (Kroh et al., 2017).

School-level data

I complement this student-level data set with self-collected administrative data on the location of elementary schools offering ASC slots between 2003 and 2018 from the six most populous Western German federal states. Specifically, my sample includes school-level data from the following federal states and years, respectively: Bavaria 2003-2018, North Rhine-Westphalia 2005-2018, Baden-Wuerttemberg 2012-2018, Hesse 2005-2018, Rhineland-Palatinate 2005-2018, and Lower Saxony 2010-2018.²³ All these states have experienced a significant expansion in offered ASC slots during the observed period.²⁴

Overall, the combination of these two data sources is uniquely suitable to estimate the MTE of ASC in Western Germany for several reasons: First, I observe individual yearly

²⁰the community type groups regions into categories according to the number of inhabitants of the specified socio-economic region, like peripheral regions or agglomerations

²¹I make use of the on-site access at DIW Berlin to obtain street-block geo-coordinates and district and municipality keys. The location of the street-block coordinates allows me to calculate the Euclidean straight-line distance from the respondent household's location to the next ASC. The district keys control for district fixed effects, and the municipality keys control for additional socio-economic municipality characteristics. I thank the team of the SOEP infrastructure for their technical support in the geo-matching process.

²²See SOEP (2021) for a detailed overview of the different SOEP samples.

²³I received data from these federal states in response to a request to all Western German federal states posed in spring 2020. I thank the DIW Graduate Center and Jan Marcus for generous funding support.

²⁴This area covers almost all of Western Germany except for the city-states Berlin, Bremen, and Hamburg, as well as the small federal states of Schleswig Holstein and Saarland.

ASC attendance. Second, I can link information on the exact location of the respondents' street block to the address data of ASC provided by the federal statistic offices of the six federal states in my sample. Specifically, I calculate the Euclidean straight-line distance to the nearest school offering ASC and the distance to the nearest elementary school irrespective of ASC offers. The former forms the continuous instrument I need to implement the MTE framework, and the latter serves as a control variable. Third, the data set has strong external validity. There are 8,274 students in my sample²⁵, representing 73 percent of the German population and 85 percent of the Western German population of elementary school children.²⁶ Fourth, using the survey data also means that I observe a rich set of outcome and control variables, which are discussed in the following sections.

4.1.1 Outcome variables

As mentioned before, the afternoon programs offered by German ASC consist of care, homework support, and supervised recreational activities, targeting cognitive and non-cognitive skill formation. I explore a variety of outcomes related to these areas, drawing a comprehensive picture of the effects of afternoon care on elementary school-age children's skill formation. I use the mother-child questionnaire surveying the mothers of children aged nine to ten, i.e. when they are still in primary school, for most of the main outcome variables.²⁷

Schooling outcomes

I use *grades in German and mathematics* as the primary measure of short-term schooling outcomes and as proxies for —but going beyond —cognitive skills.²⁸

In the German school system, grades range from 1 (best) to 6 (worst), with 4 (sufficient)

²⁵The initial sample counts 10,376 children, of which I have complete information (no missings in the control variables) for 8,274 individuals.

²⁶The population in these states amounted to 60.5 million inhabitants in 2020, which corresponded to 72.8 percent of the 83.2 million inhabitants in Germany and to 85.7 percent of 70.6 million in Western Germany the same year (Destatis, 2021).

²⁷The full questionnaires for each year is found here: https://www.diw.de/en/diw_02.c.222729.en/questionnaires.html

²⁸While grades are not a perfect substitute for IQ, they are, in fact, a better predictor of a variety of life outcomes like wages and health since they additionally capture personality aspects that have independent predictive power (e.g., Borghans et al., 2016).

being the grade at which a class counts as passed. I reverse the scale for better interpretability; thus, 6 is the best possible grade in the outcome variable. The third outcome variable is *transition to the academic track* and takes value one if the student qualifies for and chooses to attend a *Gymnasium* after grade four.²⁹ The variable is generated by tracking the parent’s indication of their child’s school type in the SOEP Household questionnaire.

Non-cognitive skills

I observe a number of outcomes broadly classified as non-cognitive skills (NCS),³⁰ which I divide into two sub-categories: socio-emotional skills and personality. For socio-emotional skills, I use the *Strengths and Difficulties Questionnaire (SDQ)*, (Goodman et al., 1998), which captures children’s behavior on five scales: hyperactivity, emotional problems, peer problems, conduct problems, and prosocial behavior. The *SDQ* questionnaire has been surveyed since 2010 (Richter et al., 2013). The formation of *prosociality*, which does not form a part of the SDQ Difficulty Scale but rather is assessed separately, is receiving special attention in the economic context for playing a critical role for later life outcomes, including educational success, labor market success, health, well-being, and social capital (Algan et al., 2014, Deming, 2017, Peter et al., 2016, Kosse et al., 2020).³¹ I measure personality development using the *Big Five* personality traits (McCrae and Costa Jr, 2008, Borghans et al., 2008, Lang et al., 2011). Table A1 defines these personality traits in more detail. The SOEP adopts a slightly shorter scale that can nevertheless reflect the basic structure of the Big-Five model in a reliable way (Richter et al., 2017). The mother-child survey features two questions per factor, with answers ranging from 1 (not

²⁹This is the case for all federal states in my sample. In some other federal states, e.g., Berlin, tracking occurs two years later.

³⁰For a discussion on the term *non-cognitive skills* and an overview of the commonly used concepts, see Borghans et al. (2008).

³¹The prosocial scale includes five items reading as follows: ‘considerate of other people’s feelings,’ ‘shares readily with other children,’ ‘helpful if someone is hurt, upset, or feeling ill,’ ‘kind to younger children,’ and ‘often volunteers to help others.’ Responses were given on a seven-point Likert scale ranging from ‘does not apply at all’ (1) to ‘applies completely’ (7). In addition, as is common practice, I construct an average score based on the four development dimensions to measure children’s socio-emotional development.

at all) to 10 (applies fully). I standardize all development indicators to have a zero mean and a standard deviation of one. In my results on non-cognitive and social skills, I do not compare children of immigrant ancestry to those who were born and raised in Germany.³² I further include a rich set of control variables.³³

4.2 Instrument assessment

The expansion of the ASC sector in Western Germany, described in Section 2.1, offers a source of variation in ASC attendance that does not depend on child or family characteristics. Proximity to relevant educational institutions is widely used in the economics discipline as an instrument for attending these institutions (e.g. Card, 1993, Rouse, 1995, Neal, 1997). The basic rationale behind using distance as an instrument for participation in (voluntary) educational offers is intuitive: individuals, weighing their costs and potential benefits, are more willing to take up an educational offer when commuting time and costs are reduced. This section reviews the validity criteria of this instrument in my setting.

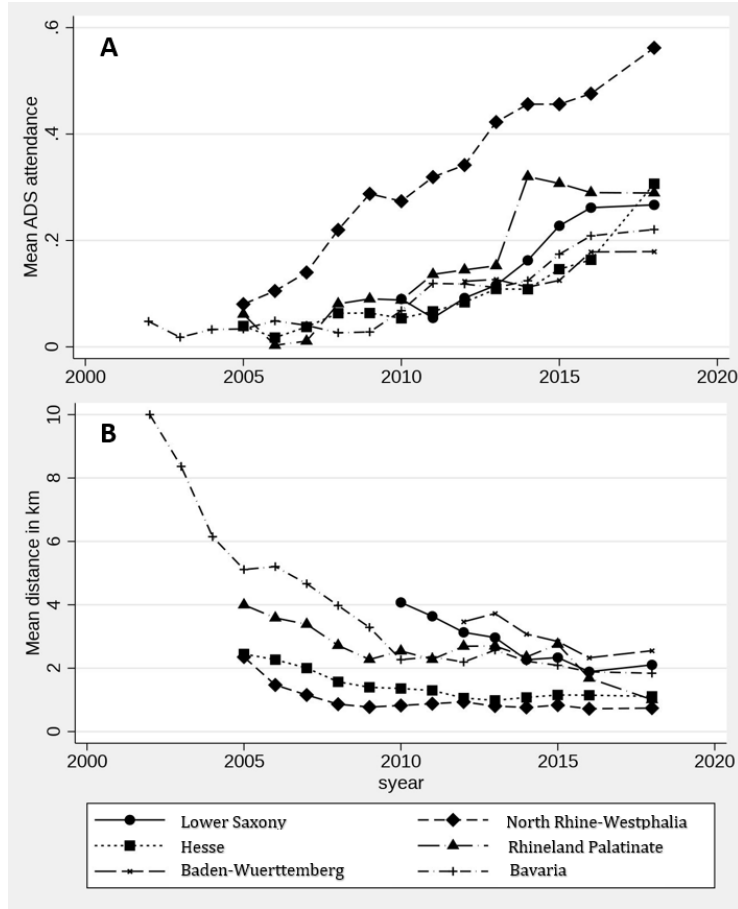
Relevance

First, the change in the average distance to the nearest ASC within one district must be a strong predictor of ASC attendance. For most federal states in Germany, distance is automatically a predictor of public elementary school attendance since children are usually

³²Immigrant parents might be rooted in the culture of their country of origin, which might affect the way they regard socially desirable behavior in children and, hence, how they report about their children (Runge and Soellner, 2019).

³³These control variables are: age and sex of the child, whether the child resides in a single-parent household, highest parental education (no degree vs. apprenticeship vs. university degree), log household net income adjusted by the OECD modified equivalence scale, employment status of the mother before the child entered primary school, age of the mother, the number of children in the household, whether there are younger children in the household, a dummy that takes value one if the household receives a social transfer (unemployment benefits I or II, benefits from the educational package, asylum seeker allowance, and/or subsistence allowance), migration background (taking value one when both parents were not born in Germany), as well as the legal status (public vs. private) and type (integrated, semi-integrated, non-integrated) of the nearest ASC. To account for systematic differences in the expansion of the ASC sector, I employ district, year, age, and municipality size (≤ 500 , 500-5000, 5000-50,000, 50,000-500,000, and $\leq 500,000$ inhabitants) fixed effects. Finally, to further account for differences in school infrastructure, I include the distance to the nearest primary school.

Figure 1: Evolution of ASC attendance and distance to ASC over time



Notes: Part A shows the evolution of the share of elementary school students attending afternoon programs for most of elementary school (the treatment). Part B displays the evolution of the distance to the nearest elementary school offering ASC (the instrument) by survey year. Source: Own calculations based on the SOEP v35, 2003-2018, as the main data source.

allocated to schools closest to their homes.³⁴ 97 percent of the schools in my sample are public; this implies that, in most cases, children attend the school that is closest to their home. Hence, the first stage in my setting is determined by a child starting to attend ASC because their school started offering these programs as a result of the IZBB reform. Figure 1 displays the correlation between the ASC extension (panel A) and the change in the distance to the next ASC in kilometers (panel B) for an average student in the six federal states in my sample over the respective observation period. Table 1 shows the selection equation results and includes the parameter estimates for the first-stage probit selection. To allow for a non-linear relationship between distance to the next ASC and afternoon program attendance, e.g., at very low or high values of Z^* , I additionally include the square of the distance in km. A reduction in the distance to the nearest ASC facility by one kilometer is associated with a 4.3 percentage point increase in attending afternoon programs, a result that is significant at the 1 percent level. The first stage estimated by OLS yields a similar coefficient of -0.045 with a t-value of -5.39. The Kleibergen-Paap Wald F statistic is 35.4.³⁵

Independence assumption

While the relevance of distances as instruments is usually uncontested, its independence has been called into question. The main concern is that, in many cases, educational institutions are not distributed randomly.³⁶ To address this concern, I residualize out district fixed effects in addition to birth and survey year fixed effects, the municipality size,

³⁴In most federal states, allocation to elementary schools is organized in catchment areas, with students attending the school closest to their home. North Rhine-Westphalia (NRW) is the only federal state that abolished this system for elementary schools in 2008, switching to an admission regime based on parental choice. However, even after the reform, incentives for staying within the former catchment area remained since travel costs are reimbursed only when the nearest school is attended (Breuing, 2014).

³⁵The heterogeneity in the first stage across selected sub-groups is displayed in Figure A.3, showing that children with fully employed mothers, social transfer receivers, single-parent households, and with a migration background are more likely to attend ASC and to react more strongly to a reduction in the distance to the next ASC.

³⁶For instance, highly educated parents might select into urban areas with a higher density of educational institutions. By including a large set of individual, family, and county-level characteristics, later studies (e.g. Dee, 2004) can plausibly defend the independence assumption of distance to educational institutions.

and the distance to the nearest elementary school to account for urban-rural differences in the geographical density of educational institutions, as well as individual and household characteristics. Therefore, the independence assumption in my case transmutes to the weaker assumption that the change of the average distance within one district conditional on the controls X_{ij} must not co-vary systematically with the child outcomes Y_{ij} .

While this assumption cannot be tested directly, I argue it is likely to hold. The pace and magnitude of the expansion of the ASC sector in a given region depended on different factors, the most important one being the allocation of IZBB investments. The latter was based on a two-step decision-making process. First, schools had to apply for investments with a school concept developed by the school's director and the school committee. In a second step, schools were selected by the federal states on a first-come, first-served basis. Beforehand, the federal states had declared a particular investment focus. While some states like Hamburg and Saxony-Anhalt focused on schools located in areas with low baseline socio-economic status or a high share of immigrants, the six states in my sample did not state any such priorities (BKJ, 2005). Instead, Bavaria, Baden-Wuerttemberg, and Lower Saxony announced that they would initially focus on secondary schools for the ASC expansion, initially slowing the expansion of all-day primary schools. On the other hand, North Rhine-Westphalia pursued a clear focus on primary schools. (BKJ, 2005).³⁷

I run two balancing tests to strengthen the independence argument. First, I test the assumption of the absence of targeting based on differences in socio-economic status by regressing the change in the distance to the next ASC offering school on socio-economic district characteristics at the beginning of my observation period.³⁸ The results of this first balancing test are shown in Table A3. As expected, the change in average distance in a district correlates highly and positively with the initial distance in the same district,

³⁷Previous studies exploiting the IZBB reform (e.g., Seidlitz and Zierow, 2020) show that pre-expansion municipality characteristics do not predict whether a municipality received IZBB funding or not, causing an exogenous variation in the number of ASC slots available in a given municipality.

³⁸Because of data limitations, the observation period differs across the different federal states in my sample and starts in 2003 for Bavaria, 2005 for North Rhine-Westphalia, Hesse, Lower Saxony, and Rhineland-Palatinate, and 2012 for Baden Wuerttemberg.

which shows that areas with low baseline density of ASC —hence a higher baseline distance —more rapidly expanded the offer of afternoon programs during the observation period. There is also a small and marginally significant correlation between the pace of ASC expansion with GDP per inhabitant. Surprisingly, districts with a high baseline share of women in the labor force experienced a slightly lower expansion of ASC over the observation period. Apart from that, the ASC expansion does not correlate with initial district characteristics. Second, I test whether the average change in distance to ASC correlates with individual characteristics in my sample. Table A4 in the appendix shows that a higher rate of ASC expansion correlates positively with having a migration background, stressing the need to include it in X_{ij} . Notably, no statistically significant correlation exists with any other critical individual SES variables.

Common support

Full common support implies that for each value of $P(Z)$, I should observe treated and non-treated individuals. To test this assumption, I estimate the propensity score using a probit regression and plot the histogram of common support. Figure A.4 graphs the unconditional support jointly generated by variation of both the instrument and the covariates (X, R, T, C), showing that the first stage generates full common support for the propensity score $P(Z)$, albeit with relatively few observations for the non-treated starting at $P(Z)=.9$.³⁹ To account for the scarcity in observations at very high levels of $P(Z)$, I limit my analysis in my baseline specification to $0 \geq P(Z) \geq .9$. As discussed in Section 3, I also test how much variation my instrument creates in each covariate cell of X , i.e., conditional on $X_i = x$, to inspect how much my data deviates from the ideal case imposing the minimal assumption of (U_0, U_1, V) being independent of Z given X . Figure A.2 reveals relatively small support of $P(Z)$ for each value of X , as is also the case in other applications (e.g., Carneiro et al., 2011). However, under the separability assumption,

³⁹Full common support is a condition rarely achieved in practice in the MTE literature (e.g. Carneiro et al., 2011, Cornelissen et al., 2018), although it is critical for computing the ATT and the ATUT, which heavily weigh individuals at the extremes of the propensity score distribution.

MTEs are identified over the marginal support of $P(Z)$ (Figure A.4).

Monotonicity

With heterogeneous treatment effects, an additional necessary assumption to identify causal effects is monotonicity. This assumption requires that students who attend an afternoon program would also do so if they lived closer to a school offering them, hence ruling out the existence of defiers in the sample. This assumption is intuitively plausible in my context since it is difficult to think of why students would stop attending afternoon programs once their catchment area school starts offering them. To strengthen this argument, Table A2 shows that the first stage is positive and statistically significant in all subsamples of the data.⁴⁰

5 Results

5.1 First stage and descriptives

Table 1 displays the results of the selection equation, showing that, on the one hand, children from single-parent households and children of immigrant ancestry are more likely to attend afternoon programs. On the other hand, children with fewer siblings and whose parents are both gainfully employed are more likely to be found in ASC. Hence, while it is generally low-SES children who take up ASC offers more often, there is also a fraction of high-SES students with a high propensity to attend ASC. Given this selection mechanism, it is not surprising that children who participate in ASC differ in the mean outcome variables from those who did not (Table A5). Children in the treatment group have lower outcomes in all outcome variables but prosociality. On average, children attending ASC have lower grades, qualify less often for the academic track, have more socio-emotional difficulties measured by the SDQ scale, are less open, and are more introverted (Table A5).

⁴⁰This is equivalent to testing a weaker form of the monotonicity assumption —average monotonicity—which is sufficient to interpret 2SLS estimates as causal effects (De Chaisemartin, 2017).

Table 1: Selection equation

| | ASC attendance |
|-------------------------------------|----------------------|
| Distance to ASC in km | -0.045*** (0.010) |
| Distance squared | 0.004*** (0.001) |
| Female | -0.007 (0.013) |
| Social transfer receiving household | 0.030 (0.022) |
| Single parent household | 0.115*** (0.021) |
| Number of children in household | -0.027** (0.010) |
| Both parents working | 0.040** (0.017) |
| Migration background | 0.065*** (0.019) |
| At least technical degree | -0.027 (0.044) |
| Academic degree | -0.007 (0.047) |
| Age mother | -0.001 (0.001) |
| Urban area | 0.031 (0.025) |
| Number of observations | 8,274 |

Notes: The reported estimated represent the average marginal effects from the probit selection model in which the dependent variable is equal to one if the child attends ASC for most of elementary school. The probit model includes the full set of individual control variables. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: SOEP v35, own calculations.

5.2 Treatment Effect Heterogeneity in Observed Student Characteristics

Based on equation 7, Table 2 reports estimates for the effects of ASC on cognitive skills measured by schooling outcomes and NCS. The results for math and German grades (Panel A) point to an equalizing effect of ASC attendance. As can be seen from comparing the coefficient of the non-interacted child characteristics with the coefficient of the characteristics interacted with the propensity score, children with lower baseline schooling outcomes tend to benefit less from ASC attendance than their peers, and vice versa. This equalizing pattern is statistically significant for children with a migration background. The latter is an important result since the (lack of) reading skills of low SES, and migrant students has been a particular concern since the 2018 PISA study, where low SES children scored significantly lower in reading literacy than high SES children (Weis et al., 2018). In contrast, children with highly educated parents, who score significantly better in the baseline, benefit significantly less from ASC attendance. However, these positive effects on low-SES children’s grades do not seem to translate into a positive effect on qualifying for the academic track.⁴¹

This equalizing pattern is also observed for the effects of ASC attendance on the SDQ and, to a slightly lesser extent, on the Big Five personality traits (Table 2, Panel B). Regarding the SDQ, children from single-parent households, who have more difficulties in the baseline, enjoy larger benefits from attending afternoon care. Children from social transfer receiving households —also with lower baseline scores —benefit disproportionately from attending ASC in terms of the Big Five, and significantly so concerning extroversion and emotional stability. The pattern of observed effect heterogeneity is more complicated for

⁴¹Note that it is not contradictory to find that participation in afternoon programs benefits migrant students more with respect to grades but not concerning their transition to secondary school. Due to biased expectations on the part of teachers and different ways for parents to intervene in the placement of their children (e.g., Jackson, 2013), there is evidence of systematic biases against low-SES students. Studies in Germany (e.g., Kristen and Dollmann, 2010) show that students with a migration background are less likely to transition to the academic track even after controlling for the grade point average.

Table 2: Outcome heterogeneity based on observed characteristics

| Panel A | Math grade | German grade | Academic track | Prosociality | SDQ |
|------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Female | -0.171** (0.067) | 0.293*** (0.071) | -0.025 (0.032) | 0.038 (0.075) | -0.194** (0.083) |
| Migration background | -0.187 (0.118) | -0.268* (0.139) | 0.001 (0.050) | | |
| Social transfer | -0.232** (0.112) | -0.350*** (0.116) | -0.090** (0.040) | -0.245** (0.118) | -0.555*** (0.141) |
| Single parent | -0.131 (0.092) | -0.133 (0.089) | -0.062 (0.051) | -0.214* (0.124) | -0.374*** (0.110) |
| Academic degree | 0.216** (0.094) | 0.380*** (0.105) | 0.326*** (0.045) | -0.273*** (0.105) | 0.278** (0.109) |
| Propensity score (PS) | 0.960* (0.564) | 0.270 (0.545) | 0.647** (0.285) | 1.416** (0.655) | 0.210 (0.676) |
| PS squared | -1.367** (0.543) | -0.514 (0.549) | -0.304 (0.316) | -0.916 (0.657) | -1.348* (0.720) |
| Female x PS | -0.118 (0.199) | -0.020 (0.195) | 0.206* (0.111) | 0.014 (0.248) | 0.517** (0.244) |
| Migrant x PS | 0.836*** (0.281) | 0.727** (0.293) | -0.151 (0.154) | | |
| Social transfer x PS | -0.002 (0.242) | -0.006 (0.230) | -0.074 (0.123) | 0.387 (0.259) | 0.313 (0.350) |
| Single parent x PS | 0.105 (0.236) | 0.157 (0.236) | -0.178 (0.157) | 0.229 (0.278) | 0.805** (0.326) |
| Academic degree x PS | -0.413 (0.272) | -0.609** (0.290) | -0.238 (0.159) | 1.042*** (0.337) | -0.546 (0.373) |
| Number of observations | 6,548 | 6,548 | 7,684 | 7,031 | 7,025 |
| Panel B | Open | Conscientious | Extroverted | Agreeable | Emotional stability |
| Female | -0.147** (0.073) | -0.134* (0.072) | -0.188** (0.083) | -0.167* (0.089) | -0.056 (0.081) |
| Social transfer | -0.270** (0.119) | -0.292** (0.124) | -0.302*** (0.113) | -0.406*** (0.145) | -0.407*** (0.132) |
| Single parent | -0.213** (0.107) | -0.262** (0.106) | 0.113 (0.094) | -0.001 (0.115) | -0.064 (0.106) |
| Academic degree | 0.084 (0.099) | 0.134 (0.096) | 0.108 (0.119) | -0.034 (0.114) | 0.014 (0.109) |
| Propensity score (PS) | 1.438** (0.633) | 1.543** (0.630) | -0.245 (0.752) | 0.627 (0.630) | 0.765 (0.595) |
| PS squared | -2.076*** (0.638) | -2.003*** (0.636) | -0.736 (0.709) | -1.084* (0.642) | -0.424 (0.586) |
| Female x PS | 0.597*** (0.229) | 0.562** (0.232) | 0.519** (0.236) | 0.323 (0.224) | 0.091 (0.221) |
| Social transfer x PS | 0.223 (0.290) | 0.209 (0.293) | 0.490* (0.265) | 0.319 (0.292) | 0.632** (0.275) |
| Single parent x PS | 0.514* (0.295) | 0.467 (0.298) | -0.033 (0.273) | 0.097 (0.259) | -0.007 (0.252) |
| Academic degree x PS | -0.234 (0.331) | -0.219 (0.332) | -0.567 (0.377) | -0.088 (0.307) | -0.327 (0.333) |
| Number of observations | 7,010 | 7,010 | 7,007 | 6,996 | 7,009 |

Notes: The table displays estimates from the first part of the outcome equation (eq. [9]). The grades and the SDQ are reversed (higher score = better outcome) to ease interpretability. Coefficients of the independent variables not interacted with the propensity score in the first part of the table measure effects on the outcome in the untreated state (i.e., β_{zero} in eq. [9]), whereas coefficients of the same regressors interacted with the propensity score measure the difference of the effects between the treated and the untreated state ($\beta_0 - \beta_1$ in eq. [9]). The regression takes into account individual variation in the instrument, i.e., individuals may enter multiple times if the distance to an ASC offering school changes while they are in elementary school. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. *Source:* SOEP v35, own calculations.

the outcome of prosociality. On the one hand, children from social transfer receiving and single-parent households score lower in the baseline and have higher treatment effects, albeit not statistically different from zero. On the other hand, children with highly educated parents show a similar but statistically significant pattern thus, they are significantly more likely than their peers to develop a more prosocial behavior in response to ASC attendance.⁴²

5.3 Essential heterogeneity

In addition to heterogeneity in terms of observed characteristics documented in section 5.2, I find substantial heterogeneity in the treatment effect based on unobserved characteristics, also referred to as essential heterogeneity. Similar to the positive selection based on levels (observed characteristics), I find a positive selection pattern concerning gains (unobserved resistance to treatment) for most outcomes. Figure 2 displays the MTE curves described by equation 9 for mean values of X in my sample. The MTE curve relates the unobserved parts of the ASC treatment effect ($U_1 - U_0$) to the unobserved parts of the choice for ASC participation (U_D) —the resistance to treatment. For most non-cognitive outcomes, except conscientiousness and the SDQ, I observe a falling curve. This shape of the MTE curve implies that for these outcomes, the treatment effect decreases as the resistance to treatment increases, meaning that students who are more eager to sign up for afternoon programs appear to benefit the most from them in terms of NCS. For the three schooling outcomes, German grade, maths grade, and transition to the academic track, the flat MTE curve signifies that students with low and high resistances for treatment on average do not react differently to ASC attendance with respect to these outcomes. This can also be seen in the p-value for the test of essential heterogeneity for these three outcomes in Table 3. Figure A.5 plots the MTE curves when applying a joint normal

⁴²As explained in section 4.1.1, when using parent-reported survey data on non-cognitive skills, comparisons of children with and without migration background are problematic. Hence I only distinguish between other SES characteristics.

approach⁴³. In this model, the MTE curves are more clearly downward sloping.

Table 3: Summary treatment statistics

| Ind. variable | Math grade | German grade | Academic track | Prosociality | SDQ |
|---------------------|----------------------|---------------------|-------------------|--------------------|---------------------|
| ATE | 0.011 (0.818) | -1.173 (0.806) | 0.254 (0.520) | 0.784 (0.932) | 0.101 (1.175) |
| ATT | 0.416 (0.512) | -0.197 (0.502) | 0.043 (0.197) | 1.006** (0.478) | 0.718 (0.536) |
| ATUT | -0.163 (1.171) | -1.593 (1.142) | 0.310 (0.635) | 0.689 (1.329) | -0.172 (1.671) |
| LATE | 0.039*** (0.011) | 0.032*** (0.005) | -0.001 (0.002) | 0.017** (0.007) | -0.006 (0.010) |
| p (observable het.) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| p (essential het.) | 0.630 | 0.825 | 0.781 | 0.177 | 0.642 |
| Ind. variable | Openness | Conscient. | Extroversion | Agreeable | Emotional stability |
| ATE | -0.812 (0.900) | 1.050 (0.880) | -0.725 (0.780) | -0.529 (0.946) | -1.282* (0.762) |
| ATT | 0.639 (0.537) | -0.186 (0.508) | 0.055 (0.559) | 0.938* (0.531) | 0.384 (0.472) |
| ATUT | -1.447 (1.244) | 1.593 (1.218) | -1.065 (1.093) | -1.174 (1.332) | -2.012* (1.128) |
| LATE | -0.022*** (0.006) | -0.011 (0.008) | -0.002 (0.005) | 0.010 (0.007) | -0.026** (0.012) |
| p (observable het.) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| p (essential het.) | 0.035 | 0.670 | 0.448 | 0.096 | 0.108 |

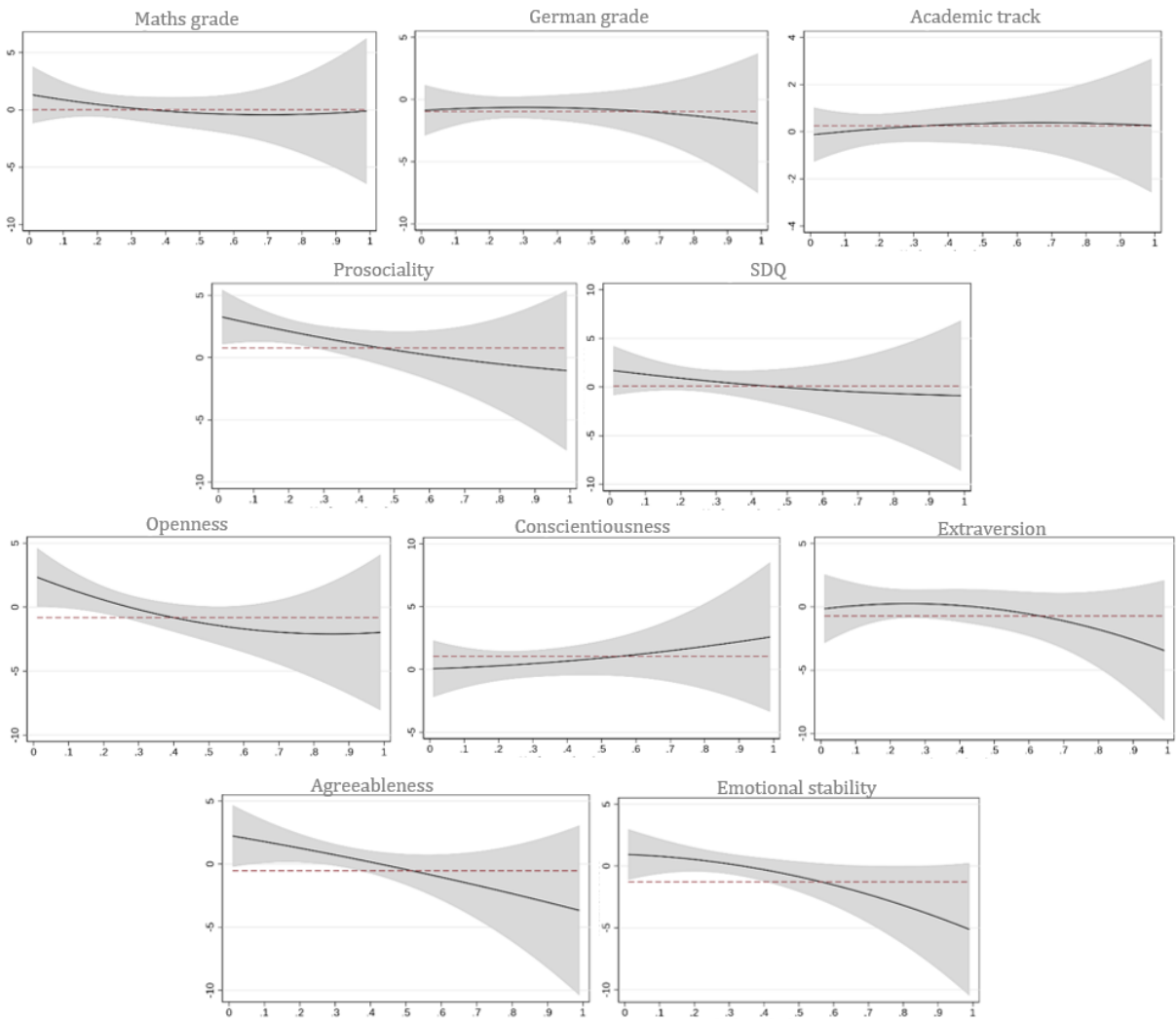
Notes: The table reports the average treatment effect (ATE), the treatment effect on the treated (TT), the treatment effect on the untreated (TUT) for all outcomes and the p-value for a test of observable and essential heterogeneity. The regression takes into account individual variation in the instrument, i.e., individuals may enter multiple times if the distance to an ASC offering school changes while they are in elementary school. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: SOEP v35, own calculations.

These patterns are also reflected in the summary treatment parameters ATE (average treatment effect), ATT (average effect of treatment on the treated), and ATUT (average treatment effect on the untreated) in Table 3. The overall pattern mirrors the falling curves in Figure 2 for NCS: The ATUT and even the ATE are negative for four of the Big Five personality traits (openness, extroversion, agreeableness, and emotional stability), with a large and statistically significant negative effect on emotional stability. In contrast,

⁴³In this model, one needs to make the stronger assumption of joint normality of (U_0, U_1, V) instead of the separability assumption (Cornelissen et al., 2016, Andresen, 2018).

Figure 2: MTE curve



Notes: The figure plots the MTE curves (expression $\sum_{k=2}^K \alpha_k \hat{p}^k$ in eq. [9] for my main outcome variables evaluated at mean values of the covariates. The 90 percent confidence interval is based on standard errors clustered at the district level. *Source:* Own calculations based on the SOEP v35, 2003-2018, as the main data source.

the ATT in these cases is positive and statistically significant in the case of agreeableness. This finding suggests a positive selection into treatment in terms of personality development: the average student who selects into ASC benefits from it, while students who do not are, on average, better off without it. Conscientiousness stands out as the outcome with a reverse outcome, with both ATUT and ATE being positive, albeit not statistically significant. This general pattern of the ATT being larger than the ATUT can also be observed for the outcomes of prosociality and the SDQ, with a sizable and significant positive ATT on prosociality.

While there is no essential heterogeneity in the treatment effects on schooling outcomes, the local average treatment effect (LATE) estimates in table 3 show that ASC had a positive and significant effect on the maths and German grades of instrument compliers. The LATE is also statistically significant and positive for the outcome of prosociality and negative for openness and emotional stability.

Notably, the ATE, derived by equally weighting over the MTE curves in figure 2 and evaluating at average values of X^{44} , is negative for half of the outcomes and statistically significant for emotional stability. This result implies that, concerning these outcomes, the average student does not benefit from ASC. Hence, if the afternoon programs were mandatory for all elementary school students, this would likely result in adverse effects on the average child regarding socio-emotional skills as measured by prosociality, SDQ, openness, extroversion, and emotional stability. In contrast, with the current regulation of most afternoon programs being voluntary, children selecting into them have either positive or zero effects, with the positive effect on prosociality and agreeableness being statistically significant and sizeable (close to 1 SD in both cases). The pattern of a positive essential heterogeneity, indicating that selection into treatment is efficient, is statistically significant at the 5 and 10 percent level for openness and agreeableness, respectively.

⁴⁴Cornelissen et al. (2016) and Andresen (2018) for a derivation of the weights

5.4 Sensitivity to alternative specifications

I conduct several sensibility checks to validate the robustness of my results. Table A6 shows the main summary parameters for the outcomes for which I observe either a statistically significant pattern of essential heterogeneity or statistically significant treatment effects in my baseline specification.⁴⁵ While for some of the specifications in Table A6 the magnitudes of the treatment effect parameters change, the overall pattern remains stable, and all robustness checks confirm a positive selection pattern, where the ATT is positive, and the ATUT is more negligible or negative. Figure A.6 additionally shows the MTE plots for the same outcome variables comparing alternative functional forms: the parametric joint normal model, third polynomial, and semiparametric estimation. The latter is an important test to check for potential misspecification in the estimated propensity score (Cornelissen et al., 2018). I also run a placebo test by estimating the MTE on the SDQ Score at age five, i.e., shortly before school entry. As shown in Figure A.7, the MTE curve for this placebo outcome is flat as expected.

5.5 Discussion

The finding of a positive selection into ASC based on levels and gains for most of my main outcome variables stands in stark contrast to the adverse selection found by Cornelissen et al. (2018) for pre-school daycare. This result may at first glance be surprising because of the apparent similarities of the institutional setting care facilities in Germany and the target age group (3-6 vs. 6-10). Still, the selection pattern differs from the study by Cornelissen et al. (2018) in important ways: First, ASC in Germany was created to increase equality of opportunity (BMBF, 2009); hence it is likely that teachers specifically target low-SES children and encourage their parents to register for the afternoon care. In

⁴⁵First, I make amendments to the way I define the treatment variable: I first include students in the treatment group who indicate that they are attending *Hort* instead of including it as a control variable (column 2). The rationale is that the afternoon care offered by *Horts* is comparable to that of ASC. Second, I modify the instrument by using the log distance to the next ASC instead of the total distance in km and its square (column 3). Third, I refrain from restricting the $P(Z)$ to $P(Z) \leq .9$ and include all observations (column 4). Forth, I cluster the standard errors on the individual level instead of on the district level (column 5).

contrast, child care centers in the 1990s and early 2000s admitted children based on their mothers' employment status and time on the waiting list, which gave high-SES families an advantage (Cornelissen et al., 2018). Second, Cornelissen et al. (2018) use administrative data on a specific region in Lower Saxony with relatively few control variables. In contrast, I use survey data representative for most of Western Germany and employ more control variables. Third, the cultural differences that prevent(ed) families of immigrant ancestry from sending their children to early daycare may no longer be relevant for my age group. Elementary school is compulsory virtually all over the world, with many countries offering ASC schemes (European Commission, 2018). Hence there may be fewer reservations for mothers with different cultural backgrounds about sending their children to ASC. Finally, attitudes on female labor market participation and child care have changed in the last decades.

Overall, my findings suggest more substantial effects of ASC on NCS than on schooling outcomes. With a sizeable average group size and staff with partly limited pedagogical training, it seems that it is more the social aspects of ASC than the homework support that make a difference for the children. This result is in line with the findings of qualitative evaluation studies (Radisch, 2009, Fischer et al., 2011, Konsortium, 2016), which also highlight the special role of ASC for social skill development of migrant children. It also strengthens the argument that during early childhood and through elementary school, children's social skills and personalities are remarkably malleable and reactive to social interactions and daycare activities (e.g., Kosse et al., 2020, Bach et al., 2019). This link is intuitive since becoming "social," "open," and "agreeable" is more easily practiced when regularly surrounded by peers. When interpreting the results on NCS, it is vital to remember that the outcomes are based on information given by the mothers. A potential limitation of these results is that I cannot rule out that the impression mothers get of their children attending ASC is affected by the shorter time window they spend together and the activities they share compared to families in the control group.⁴⁶

⁴⁶For example, children who are regularly in school in the afternoon may be more tired when they get

My analysis also demonstrates that the LATE effects estimated by conventional IV methods differ from MTE estimates and disguise important heterogeneity patterns. Interestingly, in contrast to the MTE treatment parameters, the LATE estimates on math and German grade are positive and significant, suggesting that it is primarily the group of instrument compliers who benefit from ASC in terms of math and German grades. This finding is in line with the results of the IV analysis of Seidlitz and Zierow (2020), which also find positive ASC effects on children’s achievements in math and German as assessed by their teacher.

Generally, it is vital to stress that my results do not suggest global positive effects of ASC. Indeed, the ATE is negative for most of my outcomes and statistically significant for emotional stability. In contrast, the ATT is positive for all primary outcomes except for German grade, transition to the academic track, and conscientiousness, and statistically significant for prosociality and agreeableness. Taken together, these results make a critical case not for an overall positive effect of afternoon programs but for the voluntariness of their offer since selection into ASC is positive and efficient. This means that federal states in the course of expanding the offer of ASC toward the legal entitlement of an ASC slot until 2026 should opt for the non-integrated type. Urging all elementary school students to participate in after-school programs would likely negatively affect their emotional stability and potentially other NCS.⁴⁷

However, an important point to consider is that with most afternoon programs being voluntary and low-SES children selecting into ASC, a concentration of low-SES children in ASC will likely lead to a further decline in acceptance of ASC by high-SES families (Steiner, 2009). In the long run, this dynamic could increase social segregation and

home, which can systematically impact how they interact with their parents. Furthermore, since they less often have to do homework at home, there may be fewer conflicts at home, leaving the parents under the impression that the child is more agreeable.

⁴⁷On the other hand, making ASC mandatory for all children would likely result in a more effective lobby for better care quality on the part of the parents. The role of peer and rank effects on student achievement and non-cognitive development is contested in the literature (Burke and Sass, 2013; Denning et al., 2018; Elsner et al., 2021)

lead to low-SES children benefiting less from ASC attendance because of adverse peer effects. Indeed, I find that the selection of low-SES children into ASC has increased since 2012 (Table A7). Hence, it should be a priority to keep ASC attractive for all students, e.g., by employing more and better trained pedagogical staff and offering more engaging recreational activities.

6 Conclusion

I examine the heterogeneous effects of after-school care (ASC) in Western Germany on children’s cognitive and non-cognitive skill development. Employing the Marginal Treatment Effect (MTE) framework, I estimate how the effect differs along observed child characteristics and their latent propensity to attend ASC. Understanding this heterogeneity is essential to determine whether selection into ASC is efficient and, i.e., if a universal roll-out is likely to result in positive effects on those who select into treatment, and whether the offer should be compulsory for all elementary school children or not. My estimation strategy relies on spatial and time variation caused by the rapid expansion of ASC slots following an extensive investment program in Germany in 2003, instrumented by the change in the distance to the next ASC from the children’s home.

I find that low-SES children —those of single-parent households and of immigrant ancestry —are more likely to attend afternoon programs during elementary school. For most of my outcomes in the area of cognitive skills (as measured by grades in math and German and transition to the academic track), socio-emotional development (prosociality and the Strengths-and-Difficulties Score), and the Big Five personality traits (openness, extroversion, and emotional stability), ASC fosters equality of opportunity. Low-SES children, who on average have worse baseline outcomes in these categories, seem to benefit more from treatment than their peers. Heterogeneity in unobserved characteristics reinforces this pattern of positive selection into ASC: children with a low resistance to treatment, i.e., who are more likely to enroll in afternoon programs, have either zero or positive treatment effects. For most of these outcomes, the average treatment effect on the treated (ATT) is

positive —and statistically significant in the case of prosociality, and agreeableness. On average, children who select into ASC are close to one SD more social and agreeable as a result of attending regular afternoon programs. In contrast, the average treatment effect on the untreated (ATUT) and the average treatment effect (ATE) tend to be negative and are statistically significant for emotional stability. This result implies that the average elementary school student would not benefit from ASC if it were mandatory.

My findings have two imperative policy implications: First, ASC does indeed seem to benefit low-SES students and can serve as a tool to increase equality of opportunity. Second, with different organizational types of ASC currently co-existing in Germany and beyond, these results make a strong case for organizing ASC in a non-integrated way, where participation in the afternoon programs is voluntary. However, this dynamic bears the risk of increased segregation, where afternoon programs become even less attractive for high-SES students. Therefore, group composition and ASC quality should be a concern for schools and policy makers.

References

- Aizer, A. (2004). Home alone: Supervision after school and child behavior. *Journal of public economics*, 88(9-10), 1835–1848.
- Algan, Y., Beasley, E., Vitaro, F., Tremblay, R. E., et al. (2014). The impact of non-cognitive skills training on academic and non-academic trajectories: From childhood to early adulthood. *Sciences Po Working Paper*.
- Andresen, M. E. (2018). Exploring marginal treatment effects: Flexible estimation using stata. *The Stata Journal*, 18(1), 118–158.
- Angrist, J., Dynarski, S., Kane, T., Pathak, P., & Walters, C. (2010). Inputs and impacts in charter schools: Kipp lynn. *American Economic Review*, 100(2), 239–43.
- Angrist, J., & Imbens, G. (1995). Identification and estimation of local average treatment effects.
- Bach, M., Koebe, J., & Peter, F. (2019). Long run effects of universal childcare on personality traits.
- Baker, M., Gruber, J., & Milligan, K. (2019). The long-run impacts of a universal child care program. *American Economic Journal: Economic Policy*, 11(3), 1–26.
- Baumeister, R., & Alquist, J. (2009). Self-regulation as a limited resource: Strength model of control and depletion. *Psychology of self-regulation: Cognitive, affective, and motivational processes*, 11, 21–33.
- Bavarian Ministry for Education and Cultural Affairs. (2010). *Der übertritt auf einen blick*. Retrieved May 6, 2021, from <https://web.archive.org/web/20100615105141/http://www.km.bayern.de/km/schule/schularten/allgemein/uebertritt/>
- Bellenberg, G. (2012). *Schulformwechsel in deutschland: Durchlässigkeit und selektion in den 16 schulsystemen der bundesländer innerhalb der sekundarstufe i*. Bertelsmann-Stiftung.
- Björklund, A., & Moffitt, R. (1987). The estimation of wage gains and welfare gains in self-selection models. *The Review of Economics and Statistics*, 42–49.
- BKJ. (2005). Kulturelle bildung und ganztagschulen: Rahmenbedingungen und umsetzung von kooperationen in den ländern. eine ländersynopse. stand 15.11. 2005.
- Blau, D., & Currie, J. (2006). Pre-school, day care, and after-school care: Who’s minding the kids? *Handbook of the Economics of Education*, 2, 1163–1278.
- BMBF. (2009). *Investitionsprogramm âzukunft bildung und betreuungâ ganztagschulen. zeit fâ¼r mehr*. Retrieved August 12, 2019, from http://ganzttag.nepomucenum.de/offen/presentation/imagbroschuere_bmbf.pdf
- BMBF. (2021). *Gesetz zur ganztägigen förderung von kindern im grundschulalter (ganztagsförderungsgesetz - gafög)*. Retrieved August 12, 2021, from <https://www.bmfsfj.de/bmfsfj/service/gesetze/gesetz-rechtsanspruch-ganztagsbetreuung-grundschulen-178966>
- Borghans, L., Duckworth, A. L., Heckman, J., & Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of human Resources*, 43(4), 972–1059.
- Borghans, L., Golsteyn, B. H., Heckman, J., & Humphries, J. E. (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences*, 113(47), 13354–13359.
- Bos, W. (2003). *Erste ergebnisse aus iglu*. Waxmann Verlag.
- Breuing, K. (2014). *Schulwahl und schulwettbewerb im dualen system: Zur aufhebung der berufsschulbezirke in nordrhein-westfalen*. Springer-Verlag.

- Brunello, G., & Schlotter, M. (2011). Non-cognitive skills and personality traits: Labour market relevance and their development in education & training systems.
- Buckingham, J., Wheldall, K., & Beaman-Wheldall, R. (2013). Why poor children are more likely to become poor readers: The school years. *Australian Journal of Education*, 57(3), 190–213.
- Burke, M. A., & Sass, T. R. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics*, 31(1), 51–82.
- Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling.
- Carneiro, P., Heckman, J., & Vytlacil, E. J. (2011). Estimating marginal returns to education. *American Economic Review*, 101(6), 2754–81.
- Carneiro, P., Lokshin, M., & Umapathi, N. (2017). Average and marginal returns to upper secondary schooling in indonesia. *Journal of Applied Econometrics*, 32(1), 16–36.
- Cornelissen, T., Dustmann, C., Raute, A., & Schönberg, U. (2016). From late to mte: Alternative methods for the evaluation of policy interventions. *Labour Economics*, 41, 47–60.
- Cornelissen, T., Dustmann, C., Raute, A., & Schönberg, U. (2018). Who benefits from universal child care? estimating marginal returns to early child care attendance. *Journal of Political Economy*, 126(6), 2356–2409.
- Costa Jr, P. T., & McCrae, R. R. (1988). From catalog to classification: Murray’s needs and the five-factor model. *Journal of personality and social psychology*, 55(2), 258.
- Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review*, 97(2), 31–47.
- Currie, J., & Almond, D. (2011). Human capital development before age five. *Handbook of labor economics* (pp. 1315–1486). Elsevier.
- De Chaisemartin, C. (2017). Tolerating defiance? local average treatment effects without monotonicity. *Quantitative Economics*, 8(2), 367–396.
- Dee, T. S. (2004). Are there civic returns to education? *Journal of public economics*, 88(9-10), 1697–1720.
- Dehos, F., & Paul, M. (2017). The effects of after-school programs on maternal employment.
- Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, 132(4), 1593–1640.
- Denning, J. T., Murphy, R., & Weinhardt, F. (2018). Class rank and long-run outcomes. *The Review of Economics and Statistics*, 1–45.
- Destatis. (2021). *Bevölkerung nach nationalität und bundesländern*. Retrieved August 12, 2021, from <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand/Tabellen/bevoelkerung-nichtdeutsch-laender.html>
- DJI. (2008). Zahlenspiegel 2007–kindertagesbetreuung im spiegel der statistik.
- Dustmann, C. (2004). Parental background, secondary school track choice, and wages. *Oxford Economic Papers*, 56(2), 209–230.
- Dustmann, C., Puhani, P. A., & Schönberg, U. (2017). The long-term effects of early track choice. *The Economic Journal*, 127(603), 1348–1380.
- Elsner, B., Isphording, I. E., & Zölitz, U. (2021). Achievement rank affects performance and major choices in college. *The Economic Journal*, 131(640), 3182–3206.

- European Commission. (2018). *The organisation of school time in europe. primary and general secondary education â 2018/19*. (tech. rep.). Publications Office of the European Union.
- Feld, J., & Zölitz, U. (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics*, *35*(2), 387–428.
- Felfe, C., & Lalive, R. (2018). Does early child care affect children’s development? *Journal of Public Economics*, *159*, 33–53.
- Felfe, C., & Zierow, L. (2014). After-school center-based care and children’s development. *The BE Journal of Economic Analysis & Policy*, *4*(4), 1299–1336.
- Felfe, C., & Zierow, L. (2018). From dawn till dusk: Implications of full-day care for children’s development. *Labour Economics*, *55*, 259–281.
- Fischer, N., Kuhn, H. P., & Züchner, I. (2011). Entwicklung von sozialverhalten in der ganztagschule. wirkungen der ganztagesteilnahme und der angebotsqualität. *Ganztagschule: Entwicklung, Qualität, Wirkungen: Längsschnittliche Befunde der Studie zur Entwicklung von Ganztagschulen (StEG)*, 246–266.
- Fletcher, J. M. (2013). The effects of personality traits on adult labor market outcomes: Evidence from siblings. *Journal of Economic Behavior & Organization*, *89*, 122–135.
- Gambaro, L., Kemptner, D., Pagel, L., Schmitz, L., & Spieß, C. K. (2020). Integration of refugee children and adolescents in and out of school: Evidence of success but still room for improvement. *DIW Weekly Report*, *10*(34), 345–354.
- Gambaro, L., Marcus, J., & Peter, F. (2019). School entry, afternoon care, and mothers’s labour supply. *Empirical Economics*, *57*(3), 769–803.
- Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., & Schupp, J. (2019). The german socio-economic panel (soep). *Jahrbücher für Nationalökonomie und Statistik*, *239*(2), 345–360.
- Golsteyn, B. H., Non, A., & Zölitz, U. (2021). The impact of peer personality on academic achievement. *Journal of Political Economy*, *129*(4), 1052–1099.
- 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.
- Gough, H. G. (1979). A creative personality scale for the adjective check list. *Journal of personality and social psychology*, *37*(8), 1398.
- Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, *103*(6), 2052–86.
- Heckman, 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.
- Heckman, J., Urzua, S., & Vytlacil, E. (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics*, *88*(3), 389–432.
- Heckman, J., & Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica*, *73*(3), 669–738.
- Heckman, J., & Vytlacil, E. J. (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the national Academy of Sciences*, *96*(8), 4730–4734.

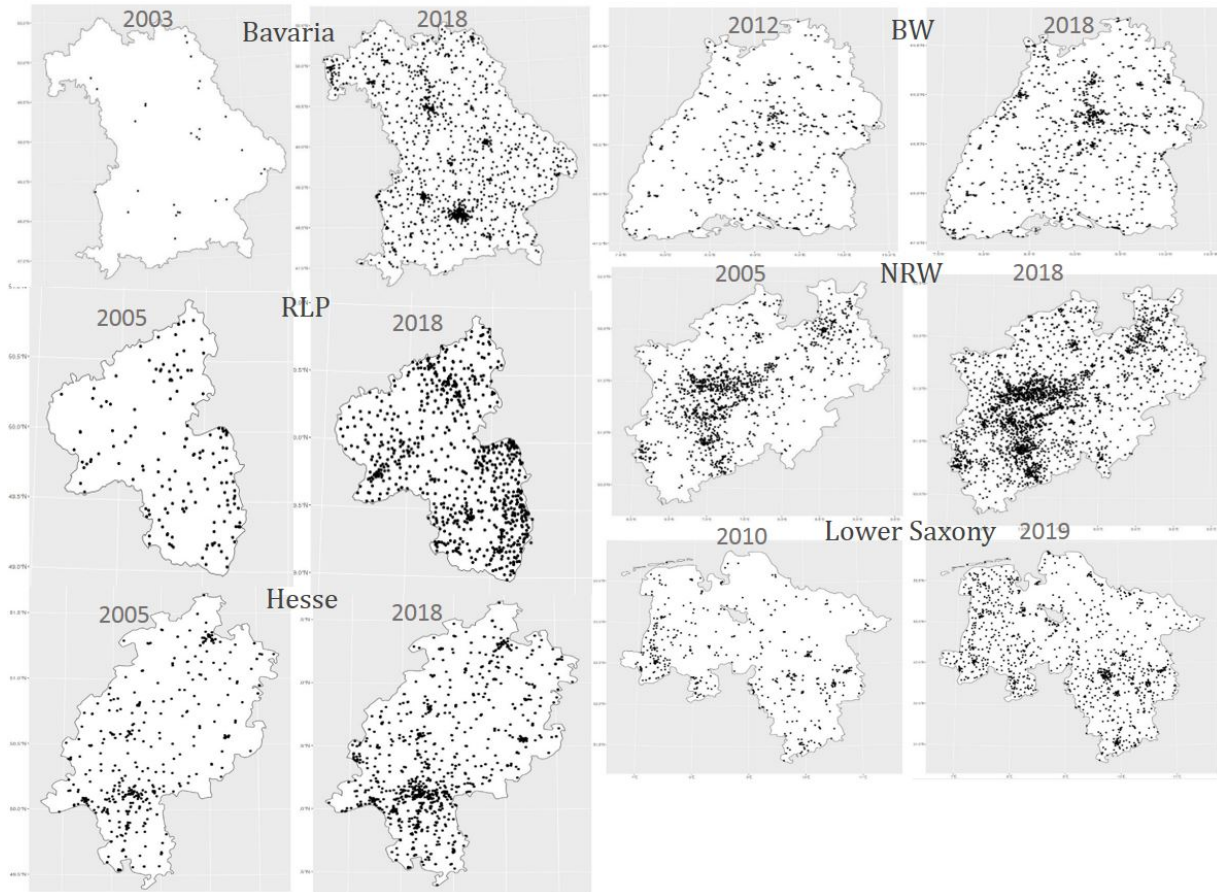
- Hille, A., & Schupp, J. (2015). How learning a musical instrument affects the development of skills. *Economics of Education Review*, 44, 56–82.
- Hogan, R. (2002). The hogan personality inventory. *Big five assessment*, 329–351.
- Jackson, M. (2013). *Determined to succeed?: Performance versus choice in educational attainment*. Stanford University Press.
- Jessen, J., Schmitz, S., & Waights, S. (2020). Understanding day care enrolment gaps. *Journal of Public Economics*, 190, 104252.
- Kamhöfer, D. A., Schmitz, H., & Westphal, M. (2019). Heterogeneity in marginal non-monetary returns to higher education. *Journal of the European Economic Association*, 17(1), 205–244.
- Kaufmann, K. M. (2014). Understanding the income gradient in college attendance in mexico: The role of heterogeneity in expected returns. *Quantitative Economics*, 5(3), 583–630.
- Kautz, T., Heckman, J., Diris, R., Ter Weel, B., & Borghans, L. (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success.
- Klemm, K., & Zorn, D. (2017). *Gute ganztagschule für alle: Kosten für den ausbau eines qualitätsvollen ganztags-schulsystems in deutschland bis 2030*. Bertelsmann Stiftung.
- KMK, S. o. t. S. C. o. t. M. o. E., & of the Federal States in the Federal Republic of Germany, C. A. (2021). *Allgemeinbildende schulen in ganztagsform in den ländern in der bundesrepublik deutschland â statistik 2011 bis 2019*. Retrieved August 12, 2021, from <https://www.kmk.org/dokumentation-statistik/statistik/schulstatistik/allgemeinbildende-schulen-in-ganztagsform.html>
- Konsortium, S. (2016). Ganztagschule: Bildungsqualität und wirkungen außerunterrichtlicher angebote. ergebnisse der studie zur entwicklung von ganztagschulen 2012–2015.
- 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.
- Kristen, C., & Dollmann, J. (2010). Sekundäre effekte der ethnischen herkunft: Kinder aus türkischen familien am ersten bildungsübergang. *Vom kindergarten bis zur hochschule* (pp. 117–144). Springer.
- Kroh, M., Kühne, S., Jacobsen, J., Siegert, M., & Siegers, R. (2017). *Sampling, non-response, and integrated weighting of the 2016 iab-bamf-soep survey of refugees (m3/m4)* (tech. rep.). SOEP Survey Papers.
- Kuehnle, D., & Oberfichtner, M. (2020). Does starting universal childcare earlier influence children’s skill development? *Demography*, 57(1), 61–98.
- Lang, F. R., John, D., Lüdtke, O., Schupp, J., & Wagner, G. G. (2011). Short assessment of the big five: Robust across survey methods except telephone interviewing. *Behavior research methods*, 43(2), 548–567.
- Lavy, V., Silva, O., & Weinhardt, F. (2012). The good, the bad, and the average: Evidence on ability peer effects in schools. *Journal of Labor Economics*, 30(2), 367–414.
- Levine, P. B., & Zimmerman, D. J. (2010). *Targeting investments in children: Fighting poverty when resources are limited*. University of Chicago Press.
- Lossen, K., Tillmann, K., Holtappels, H. G., Rollett, W., & Hannemann, J. (2016). Entwicklung der naturwissenschaftlichen kompetenzen und des sachunterrichts-bezogenen selbstkonzepts bei schüler/-innen in ganztagsgrundschulen. ergebnisse

- der längsschnittstudie steg-p zu effekten der schülerbeteiligung und der angebotssqualität. *Zeitschrift für Pädagogik*, 62(6), 760–779.
- Marcus, J., Nemitz, J., & Spieß, C. K. (2016). Veränderungen in der gruppenspezifischen nutzung von ganztägigen schulangeboten-längsschnittanalysen für den primarbereich. *Zeitschrift für Erziehungswissenschaft*, 19(2), 415–442.
- McCrae, R. R., & Costa Jr, P. T. (2008). The five-factor theory of personality.
- Ministry of Education North Rhine-Westphalia. (2012). *Grundschulempfehlung*. Retrieved May 6, 2021, from <https://web.archive.org/web/20120603224849/http://www.schulministerium.nrw.de/BP/Eltern/Grundschule/Grundschulempfehlung/>
- Neal, D. (1997). The effects of catholic secondary schooling on educational achievement. *Journal of Labor Economics*, 15(1, Part 1), 98–123.
- Nemitz, J. (2015). The effect of all-day primary school programs on maternal labor supply. *Available at SSRN 2706505*.
- Nybom, M. (2017). The distribution of lifetime earnings returns to college. *Journal of Labor Economics*, 35(4), 903–952.
- OECD. (2017). Starting strong 2017: Key oecd indicators. Retrieved August 12, 2021, from <https://www.oecd-ilibrary.org/content/publication/9789264276116-en>
- Peter, F. H., Schober, P. S., & Spiess, K. C. (2016). Early birds in day care: The social gradient in starting day care and children’s non-cognitive skills. *CESifo Economic Studies*, 62(4), 725–751.
- Plantenga, J., & Remery, C. (2015). Provision of childcare services: A comparative review of eu member states. *CESifo DICE Report*, 13(1), 20–24.
- Radisch, F. (2009). *Qualität und wirkung ganztägiger schulorganisation: Theoretische und empirische befunde*. Beltz Juventa.
- Redmond, C. (2014). Heterogeneity, selection and advantage in the graduate and non-graduate labour market.
- Richter, D., Metzing, M., Weinhardt, M., & Schupp, J. (2013). *Soep scales manual* (tech. rep.). SOEP Survey Papers.
- Richter, D., Rohrer, J., Metzing, M., Nestler, W., Weinhardt, M., & Schupp, J. (2017). *Soep scales manual (updated for soep-core v32. 1)* (tech. rep.). SOEP Survey Papers.
- Rouse, C. E. (1995). Democratization or diversion? the effect of community colleges on educational attainment. *Journal of Business & Economic Statistics*, 13(2), 217–224.
- Runge, R. A., & Soellner, R. (2019). Measuring children’s emotional and behavioural problems: Are sdq parent reports from native and immigrant parents comparable? *Child and adolescent psychiatry and mental health*, 13(1), 1–15.
- Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly journal of economics*, 116(2), 681–704.
- Sauerwein, M. N., Thieme, N., & Chiapparini, E. (2019). Wie steht es mit der ganztagschule? ein forschungsreview mit sozialpädagogischer kommentierung. *Soziale Passagen*, 1–17.
- Seidlitz, A., & Zierow, L. (2020). *The impact of all-day schools on student achievement-evidence from extending school days in german primary schools* (tech. rep.). CESifo Working Paper.
- SOEP. (2021). *The soep samples in detail*. Retrieved August 12, 2019, from <http://companion.soep.de/Target%20Population%20in%20Detail.html>

- Steiner, C. (2009). *Mehr chancengleichheit durch die ganztagschule?*
- Su, L., Tu, Y., & Ullah, A. (2015). Testing additive separability of error term in nonparametric structural models. *Econometric Reviews*, 34(6-10), 1057–1088.
- van Bergen, E., van Zuijen, T., Bishop, D., & de Jong, P. F. (2017). Why are home literacy environment and children’s reading skills associated? what parental skills reveal. *Reading Research Quarterly*, 52(2), 147–160.
- Washington-Nortey, P.-M., Zhang, F., Xu, Y., Ruiz, A. B., Chen, C.-C., & Spence, C. (2020). The impact of peer interactions on language development among preschool english language learners: A systematic review. *Early Childhood Education Journal*, 1–11.
- Weis, M., Müller, K., Mang, J., Heine, J.-H., Mahler, N., & Reiss, K. (2018). Soziale herkunft, zuwanderungshintergrund und lesekompetenz. *Pisa*, 129–162.
- Westphal, M., Kamhöfer, D., & Schmitz, H. (2020). Marginal college wage premiums under selection into employment.
- Whitaker, A. A., Baker, G., Matthews, L. J., McCombs, J. S., & Barrett, M. (2019). *Who plays, who pays?: Funding for and access to youth sports*. RAND.
- Zhou, X., & Xie, Y. (2019). Marginal treatment effects from a propensity score perspective. *Journal of Political Economy*, 127(6), 3070–3084.

Annex

Figure A.1: ASC expansion in Western Germany



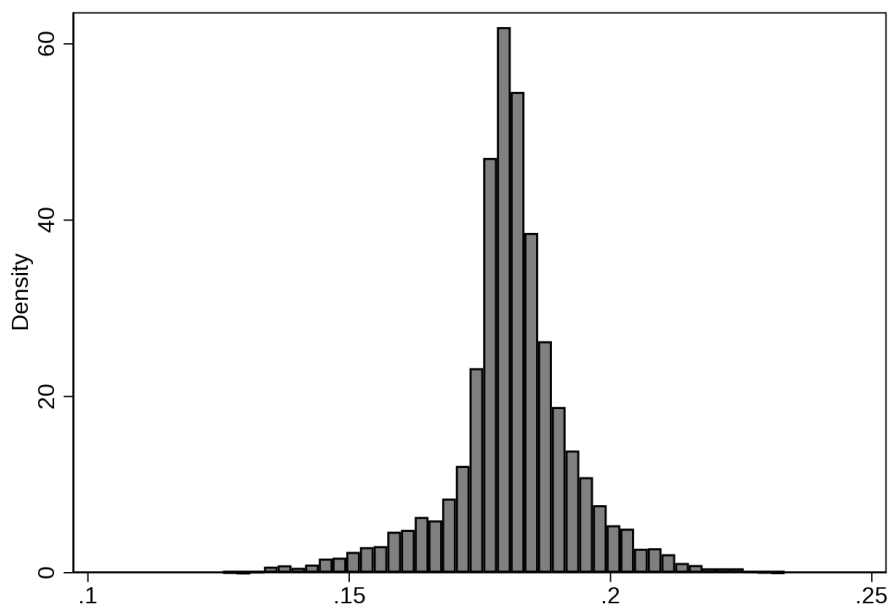
Notes: The figure plots the ASC expansion in the six federal states in my sample —Bavaria, Rhineland-Palatinate (RLP), Hesse, Baden-Wuerttemberg (BW), North Rhine-Westphalia (NRW) and Lower Saxony —in the first and last observed year, respectively. Source: self-collected school data from federal statistical offices.

Table A1: The Big Five Domains

| Factor | Facets | Definition of Factor | ACL ¹ Marker Items for Factor |
|---------------------------------------|----------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| I. Openness to Experience (Intellect) | Fantasy, Aesthetics, Feelings, Actions, Ideas, Values | The degree to which a person needs intellectual stimulation, change, and variety. | Commonplace, Narrow-interest, Simple- vs. Wide-interest, Imaginative, Intelligent |
| II. Conscientiousness | Competence, Order, Dutifulness, Achievement striving, Self-discipline, Deliberation | The degree to which a person is willing to comply with conventional rules, norms, and standards. | Careless, Disorderly, Frivolous vs. Organized, Thorough, Precise |
| III. Extraversion | Warmth, Gregariousness, Assertiveness, Activity, Excitement seeking, Positive emotions | The degree to which a person needs attention and social interaction. | Quiet, Reserved, Shy vs. Talkative, Assertive, Active |
| IV. Agreeableness | Trust, Straight-forwardness, Altruism, Compliance, Modesty, Tender-mindedness | The degree to which a person needs pleasant and harmonious relations with others. | Fault-finding, Cold, Unfriendly vs. Sympathetic, Kind, Friendly |
| V. Neuroticism (Emotional Stability) | Anxiety, Angry hostility, Depression, Self-consciousness, Impulsiveness, Vulnerability | The degree to which a person experiences the world as threatening and beyond his/her control. | Tense, Anxious, Nervous vs. Stable, Calm, Contented |

Notes: Source: Borghans et al. (2008), Costa Jr and McCrae (1988) and Hogan (2002). 1: ACL = Adjective Check List (Gough, 1979).

Figure A.2: Test for minimal assumptions



Notes: The figure plots the predicted propensity scores from a probit regression of ASC attendance on the residualized values of Z^* , i.e., the error term of a regression of Z^* on $[X, R, T, C]$.

Source: Own calculations based on the SOEP v35, 2003-2018, as the main data source.

Table A2: Monotonicity of the instrument

| | Migration background | | social transfer receiving | | Single parent | |
|-------------|----------------------|--------------------|---------------------------|--------------------|--------------------|--------------------|
| | yes | no | yes | no | yes | no |
| Distance | -0.06*** (0.02) | -0.04*** (0.01) | -0.04** (0.02) | -0.05*** (0.01) | -0.04*** (0.01) | -0.04*** (0.01) |
| Sample mean | 0.32 | 0.17 | 0.39 | 0.17 | 0.28 | 0.20 |
| | Academic degree | | Younger siblings | | Both parents work | |
| | yes | no | yes | no | yes | no |
| Distance | -0.05*** (0.01) | -0.04*** (0.01) | -0.03** (0.01) | -0.05*** (0.01) | -0.05*** (0.01) | -0.03*** (0.01) |
| Sample mean | 0.21 | 0.22 | 0.24 | 0.21 | 0.19 | 0.24 |
| | Household income | | Urban/rural setting | | School performance | |
| | low | high | urban | rural | above | below |
| Distance | -0.04*** (0.01) | -0.05*** (0.01) | -0.05*** (0.01) | -0.02** (0.01) | -0.05*** (0.01) | -0.03*** (0.01) |
| Sample mean | 0.23 | 0.18 | 0.24 | 0.13 | 0.18 | 0.21 |

Notes: This table reports first stage results by subsamples based on household characteristics. All specifications control for time and district fixed effects. Robust standard errors are clustered at the district level and reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: SOEP v35, own calculations.

Table A3: Balancing Test: Avg. distance change and initial district characteristics

| | Distance change in district |
|------------------------------------------------|-----------------------------|
| Distance in first year | 0.853*** (0.035) |
| Share of unemployed | -0.005 (0.035) |
| Share of inhabitants with migration background | -0.002 (0.044) |
| Median household income | 0.003 (0.539) |
| GDP per inhabitant | 0.012* (0.007) |
| Share of migrants in age group 6-9 | 0.019 (0.026) |
| Share of school dropouts | -0.007 (0.036) |
| Share of children age 6-9 in population | 0.242 (0.199) |
| Share of academic track amlumni | -0.036 (0.025) |
| Share of lower track alumni | 0.011 (0.013) |
| Share of women in the labor force | -0.056** (0.025) |
| Number of observations | 8,274 |

Notes: The table displays the determinants of the expansion in ASC slots in the different districts in my sample, by regressing the change in the distance to the nearest ASC between the first and last observed years on the initial average distance and baseline district characteristics. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: SOEP v35, own calculations.

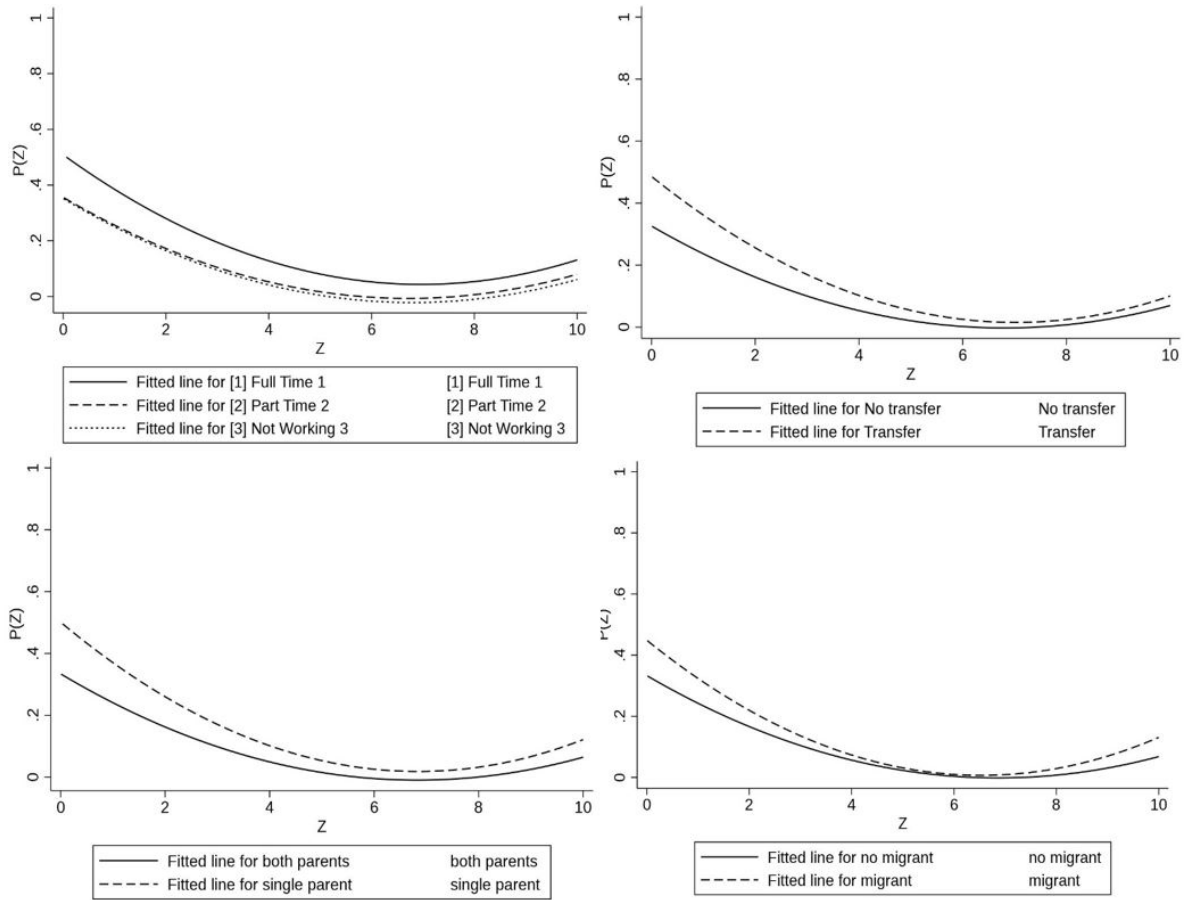
Table A4: Balancing Test: Individual determinants of ASC expansion

| | Individual characteristics |
|-------------------------------------------|----------------------------|
| Distance in first year | 0.858*** (0.039) |
| Social transfer | 0.025 (0.046) |
| Single parent | -0.061 (0.043) |
| Children in the household | -0.006 (0.018) |
| At least one parent with technical degree | -0.093 (0.070) |
| At least one parent with academic degree | -0.120 (0.079) |
| Migration background | 0.099** (0.047) |
| Household income group | 0.014 (0.042) |
| Both parents employed | 0.059 (0.053) |
| Number of observations | 8,274 |

Notes: The reported estimates are derived by regressing the change in the average distance to the nearest ASC in the child's district between the first and last observed year. The OLS regression includes the full set of individual control variables * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: SOEP v35, own calculations.

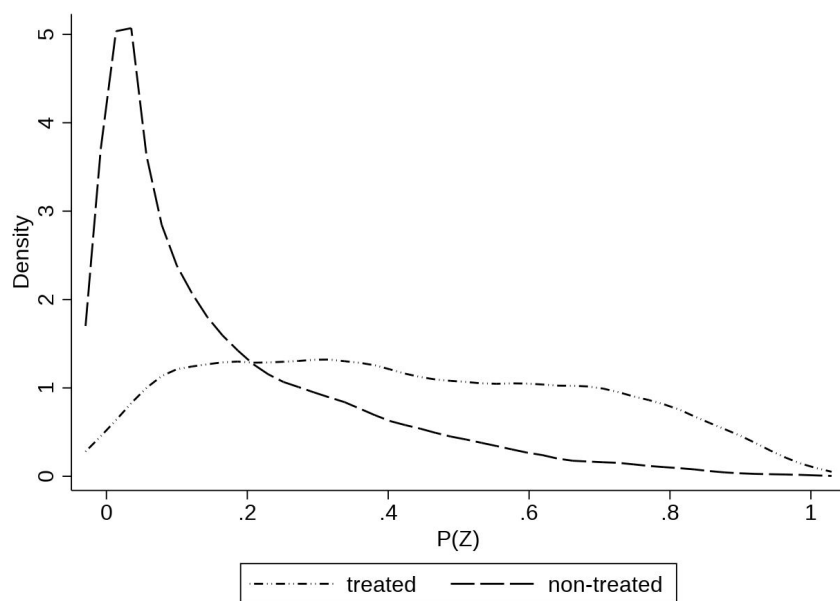
Figure A.3: Heterogeneity in the first stage



Notes: The graph displays MTE curves for the outcome of math grade, evaluated at mean values of the covariates. The solid MTE curve refers to my baseline specification (implying a linear MTE curve). The figure also shows three additional MTE curves that allow for richer patterns of the MTE curve: one curve obtained from a semiparametric approach and two curves based on a square and a cubic of the propensity score in equation 10.

Source: Own calculations based on the SOEP v35, 2003-2018, as the main data source.

Figure A.4: Common support graph



Notes: The figure plots the frequency distribution of the propensity score by treatment status. The propensity score is predicted from the baseline first-stage regression.

Source: Own calculations based on the SOEP v35, 2003-2018, as the main data source.

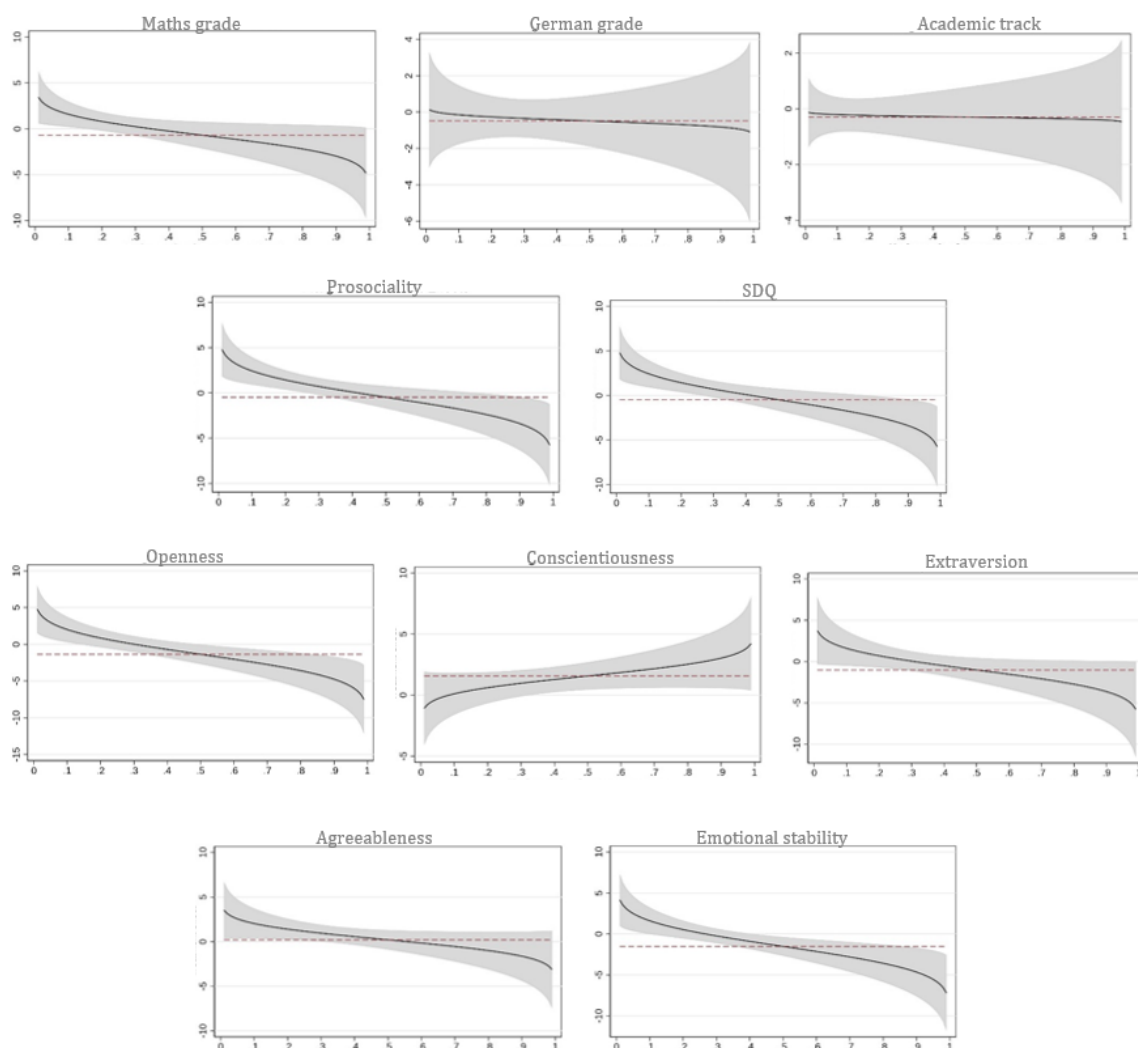
Table A5: T-test of differences in outcome variables between children attending ASC and children who do not

| Outcome variables | Treatment | Control | Δ | t |
|-------------------------|-----------|---------|----------|---------|
| Math grade (reversed) | 3.58 | 3.68 | -0.10*** | (-4.26) |
| German grade (reversed) | 3.52 | 3.62 | -0.11*** | (-4.95) |
| Academic track | 0.35 | 0.42 | -0.07*** | (-6.29) |
| Prosociality | 8.24 | 8.22 | 0.01 | (0.33) |
| SDQ Scale (reversed) | 28.93 | 30.18 | -1.25*** | (-7.70) |
| Openness | 14.52 | 14.96 | -0.44*** | (-4.32) |
| Conscientiousness | 10.45 | 10.64 | -0.19 | (-1.62) |
| Extraversion | 14.75 | 14.95 | -0.20* | (-2.07) |
| Agreeableness | 12.71 | 12.79 | -0.08 | (-0.83) |
| Emotional stability | 12.88 | 12.92 | -0.03 | (-0.30) |

Notes: T-test of background characteristics between treatment and control group for the full sample. Definition of Treatment: Having attended afternoon programs for most of elementary school in elementary school. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: SOEP v35, own calculations.

Figure A.5: MTE curves: joint normal model



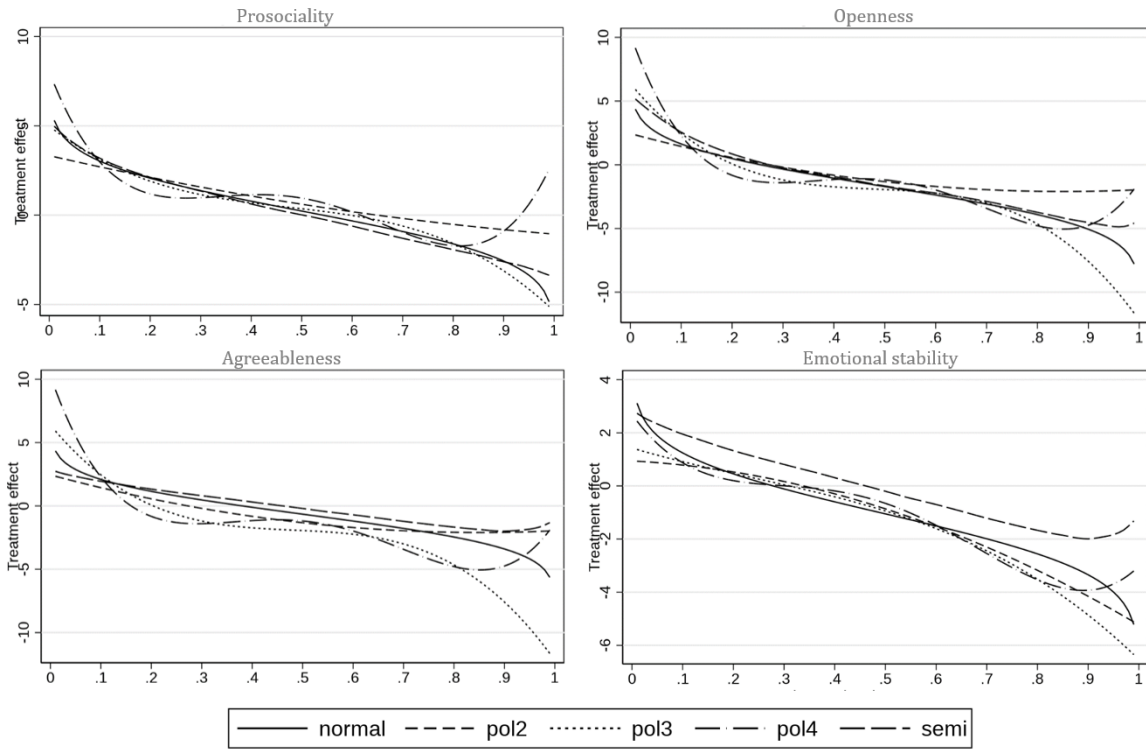
Notes: The figure plots the MTE curves (expression $\sum_{k=2}^K \alpha_k \hat{p}^k$ in eq. [9] for my outcome variables evaluated at mean values of the covariates. The underlying functional form is the joint normal model. The 90 percent confidence interval is based on standard errors clustered at the district level.
Source: Own calculations based on the SOEP v35, 2003-2018, as the main data source.

Table A6: Robustness check: ATE, TT, and TUT for main outcomes

| Prosociality | Baseline | <i>Hort</i> | log distance | P(Z) | clustering |
|---------------------|----------------------|--------------------|----------------------|----------------------|----------------------|
| ATE | 0.784 (0.932) | 0.575 (0.810) | 1.039 (0.909) | 0.352 (0.656) | 0.784 (0.914) |
| ATT | 1.006** (0.478) | 1.075** (0.473) | 0.866 (0.551) | 0.975** (0.475) | 1.006** (0.502) |
| ATUT | 0.689 (1.329) | 0.364 (1.069) | 1.119 (1.301) | 0.075 (0.924) | 0.689 (1.282) |
| LATE | 0.017** (0.007) | -0.014 (0.009) | 0.016** (0.006) | 0.016** (0.007) | 0.017** (0.012) |
| p (essential het.) | 0.177 | 0.027 | 0.452 | 0.016 | 0.150 |
| Openness | Baseline | <i>Hort</i> | log distance | P(Z) | clustering |
| ATE | -0.812 (0.900) | -0.104 (0.772) | -1.940** (0.859) | -0.491 (0.628) | -0.812 (0.879) |
| ATT | 0.639 (0.537) | 0.395 (0.528) | 0.332 (0.620) | 0.716 (0.545) | 0.639 (0.467) |
| ATUT | -1.447 (1.244) | -0.312 (0.982) | -2.935** (1.177) | -1.029 (0.854) | -1.447 (1.239) |
| LATE | -0.022*** (0.006) | -0.015 (0.018) | -0.019*** (0.005) | -0.020*** (0.006) | -0.022*** (0.007) |
| p (essential het.) | 0.035 | 0.188 | 0.033 | 0.018 | 0.031 |
| Agreeableness | Baseline | <i>Hort</i> | log distance | P(Z) | clustering |
| ATE | -0.529 (0.946) | -0.370 (0.933) | -0.414 (0.919) | -0.025 (0.653) | -0.529 (0.852) |
| ATT | 0.938* (0.531) | 1.049* (0.563) | 0.718 (0.596) | 0.737 (0.533) | 0.938* (0.479) |
| ATUT | -1.174 (1.332) | -0.983 (1.311) | -0.911 (1.306) | -0.365 (0.896) | -1.174 (1.205) |
| LATE | 0.010 (0.007) | 0.006 (0.005) | 0.008 (0.006) | 0.008 (0.007) | 0.010 (0.009) |
| p (essential het.) | 0.096 | 0.065 | 0.335 | 0.159 | 0.052 |
| Emotional Stability | Baseline | <i>Hort</i> | log distance | P(Z) | clustering |
| ATE | -1.282* (0.762) | -1.236 (0.862) | -1.217* (0.688) | -0.916* (0.535) | -1.282 (0.837) |
| ATT | 0.384 (0.472) | 0.358 (0.513) | 0.425 (0.589) | 0.430 (0.468) | 0.384 (0.462) |
| ATUT | -2.012* (1.128) | -1.920 (1.264) | -1.937* (0.983) | -1.517* (0.781) | -2.012* (1.164) |
| LATE | -0.026** (0.012) | -0.007 (0.006) | -0.021 (0.013) | -0.024** (0.012) | -0.026* (0.015) |
| p (essential het.) | 0.108 | 0.087 | 0.070 | 0.059 | 0.049 |

Notes: The table reports the average treatment effect (ATE), the treatment effect on the treated (TT), the treatment effect on the untreated (TUT), the local average treatment effect (LATE) and the p-value for the test of essential heterogeneity for alternative specifications of the outcomes with statistically significant results in the baseline specification. Column 2 presents the results when the treatment includes *Hort* programs, column 3 corresponds to the results when log of distance is used as an alternative instrument, column 4 shows the results of a model that includes the whole range of P(Z), and column 5 displays the estimates when clustering on the individual level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. *Source:* SOEP v35, own calculations.

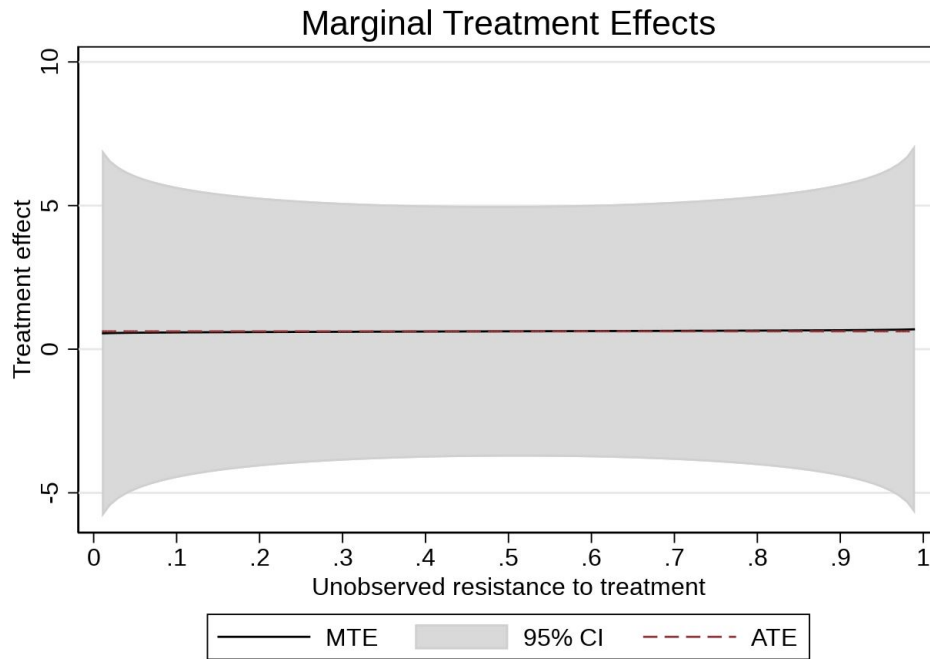
Figure A.6: MTE curves: alternative functional forms



Notes: The figure displays MTE curves for the outcomes for which my baseline specification —assuming a joint normal distribution —found statistically significant essential heterogeneity, i.e. prosociality, openness, agreeableness and emotional stability. The solid MTE curve refers to the joint normal model specification, the finely dashed, dotted and widely dashed lines show the pattern of the MTE curves obtained by using a square, cubic and a quartic of the propensity score, respectively, and the larger dashed-dotted line corresponds to the MTE curve resulting from a semiparametric approach.

Source: Own calculations based on the SOEP v35, 2003-2018, as the main data source.

Figure A.7: MTE curve of SDQ prior to school entry



Notes: The figure displays the MTE curve for the placebo outcome SDQ score at age five. Source: Own calculations based on the SOEP v35, 2003-2018, as the main data source.

Table A7: Selection into ASC pre vs. post 2012

| | 2003-2012 | 2012-2018 |
|---------------------------------|----------------------|----------------------|
| Distance to ADS in km | -0.033*** (0.012) | -0.057*** (0.018) |
| Distance squared | 0.002** (0.001) | 0.005*** (0.002) |
| Social transfer | -0.006 (0.028) | 0.056* (0.033) |
| Single parent | 0.124*** (0.025) | 0.128*** (0.037) |
| Number of children in household | -0.022** (0.011) | -0.040** (0.016) |
| Both parents working | 0.040* (0.021) | 0.050* (0.029) |
| Migration background | 0.033 (0.027) | 0.098*** (0.029) |
| At least technical degree | -0.070 (0.055) | 0.012 (0.061) |
| Academic degree | -0.046 (0.056) | 0.028 (0.065) |
| Urban area | -0.018 (0.026) | 0.122*** (0.042) |
| Number of observations | 4,130 | 4,144 |

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: SOEP v35, own calculations.