

Kamhöfer, Daniel

**Conference Paper**

## The Effect of Early Childhood Language Training Programs on the Contemporary Formation of Grammar Skills

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Education III, No. C18-V1

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Kamhöfer, Daniel (2014) : The Effect of Early Childhood Language Training Programs on the Contemporary Formation of Grammar Skills, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Education III, No. C18-V1, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<http://hdl.handle.net/10419/100374>

**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.*

# The Effect of Early Childhood Language Training Programs on the Contemporary Formation of Grammar Skills\*

Daniel A. Kamhöfer

University of Duisburg-Essen  
and CINCH, Essen

February 26, 2014

PRELIMINARY — do not cite without permission of the author

## **Abstract**

While there is a big literature on the benefits of pre-school education, only little is known why kindergarten attendance improves later-life outcomes. This is partly because most studies analyze the effect of complete 2 years pre-school programs. In order to shed light into the black box of kindergarten education, I am using the German National Educational Panel Study and regress the level of grammar skills – a main intelligence component – on the participation in a nationwide-used language training program and a rich set of conditioning variables. Taking information on mathematical skills into account, this paper also employs matching and differences-in-differences methods as well as a combination of both. The estimated effects of participating in a language training program at the kindergarten on grammar skills range between 11% and 17% of a standard deviation.

**Keywords:** Early Childhood Interventions, Cognitive Skills, DiD Matching

**JEL Classification:** *I21, I28, J24*

---

All correspondence to Daniel A. Kamhöfer, University of Duisburg-Essen and CINCH–Health Economics Research Center, Essen; Mail: Schützenbahn 70, 45127 Essen, Germany; E-mail: daniel.kamhoefer@uni-due.de.

\*For valuable comments I am grateful to my colleagues at the Chair of Health Economics and at the CINCH. Moreover, I would like to thank Martin Fischer, Hendrik Schmitz, Stefanie Schurer, and Matthias Westphal. Financial support from Mercator Research Center Ruhr (Grant number An-2013-022) is gratefully acknowledged.

# 1. Introduction

Since the highly influential contributions by James Heckman at the beginning of the 2000s, the public interest in the effects of early childhood interventions has grown rapidly. Numerous studies analyzed short- and long-term effects of interventions in the education of 2- to 4-year-olds. Usually, the treatment, i.e. the educational intervention, is the participation in a comprehensive 2 or 3 years pre-school program. The benefit of the program participation is measured as the outcome difference between treated and untreated children. Attendance in a pre-school program goes along with substantially higher levels of cognitive skills, educational achievements, and labor market outcomes. However, most studies treat pre-school education as a “black box”, i.e. they do not disaggregate the overall effect into single components. This paper breaks new grounds by shedding light upon the question how pre-school education affects children’s performance. Hence, this study does not only contribute the question “*Does pre-school attendance improve children’s development?*” but it also looks at the question “*How does pre-school education promote performance?*” Because a large share of children already receives facility-based pre-school education, e.g. by visiting a kindergarten, the latter question is of rising importance. In the United States, 78% of all 4-year-olds are enrolled in kindergartens; in the United Kingdom, the share is about 67%; and the average of OECD countries is 82% (OECD, 2013). In Germany nearly all children (96%) visit a kindergarten and since August, 1, 2013 there is a legal right of institutional child care. In order to guarantee universal access to institutional child care, the German Federal Ministry of Family Affairs, Senior Citizens, Women and Youth is going to invest 5.4 billion euros in the pre-school educational infrastructure until the end of 2014.<sup>1</sup> These figures do not only underline the relevance of pre-school education as a whole but they also show how important the analysis of the mechanisms within pre-school education is.

Particularly, I analyze the effect of a Child’s participation in a nationwide-used standardized language training program on its level of verbal intelligence using the German National Educational Panel Study (NEPS). The data set includes a battery of competence tests which measure different kinds of abilities, e.g. grammar and math skills. These skills refer to distinct components of the widely used Thurstone classification of intelligence. To assess the level of verbal skills, a German version of the well-established “Test for Reception of Grammar” which contains 48 items was conducted in the NEPS. The tests take place while the child is at the age of 3 or 4. The dependent variable is the child’s grammar test score (as level and as difference to the

---

<sup>1</sup>For more information see <http://www.bmfsfj.de/BMFSFJ/Kinder-und-Jugend/kinderbetreuung.html>.

math score). Using the contemporary level of skills has the advantage of a lower probability, that third factors compromising the true effects over the time. Moreover, the current level of a child's skills is highly correlated with its level of skills at adolescence and adulthood. The treatment variable is an indicator whether a child's kindergarten has introduced a nationwide-used language training program or not.

To estimate the effect of language training on verbal intelligence, I apply various strategies. The most obvious way to estimate the effect is to compare the grammar scores of treated and untreated children conditioning on observable factors. However, this simple differences approach is likely to mirror the correlation between language training and verbal intelligence, only. This is because neither the kindergarten's decision to introduce a language training program nor the parents' decision for a particular kindergarten can assumed to be independent from the child's level of verbal intelligence. In order to solve this selection problem, I use a differences-in-differences (DiD) strategy. Instead of the level of verbal intelligence, the outcome variable of the DiD approach is the difference between verbal intelligence and numerical intelligence. Thus, the DiD estimator uses the child's math skills to cancel out the effects of individual-specific subject-invariant unobservable factors. This approach identifies a causal effect if unobservable factors affect verbal and numerical skills in the same way. When conditioning on observable factors (like mother tongue), the language training participation is more likely to be independent from the grammar-math score difference than from the level of grammar skills. Moreover, I apply propensity score matching and semi-parametric regression-adjusted differences-in-differences matching in order to ensure the comparability of treated and untreated children and to relax the functional form assumption of the skill formation. The estimated effects of these approaches are similar: children who participate in a language training program exhibit a 11% to 17% of a standard deviation higher verbal intelligence.

This paper contributes to the literature in several ways. First, instead of analyzing the overall effects of a complete 2 or 3 years pre-school education program, it focuses on a particular measure within pre-school education. This rather narrow focus is of special interest since most children already receive pre-school education. Second, this study contributes to the analysis of the determinants of early childhood skill development in Germany. So far, only a few economic studies use German data and they do not overcome a selection process in order to address a causal interpretation. Furthermore, the application of various estimation strategies which rely on distinct identifying assumptions allows a closer investigation of the role of confounding factors.

The paper proceeds as follows. Section 2 briefly summarizes the findings of the previous literature on the effects of pre-school education. Section 3 presents the different estimation strategies and their underlying assumptions. Section 4 introduces the German

National Educational Panel Study, while Section 5 shows the results of participating in language programs. Section 6 concludes.

## 2. Previous Literature

### 2.1. Psychology

The investigation and classification of skills is subject to psychological research for over a century. Since a couple of decades the biological transmission and development of skills is also analyzed by neuroscientists. A widely used classification of distinct types of skills is Thurstone's multiple factor approach of intelligence. The overall intelligence consists seven "primary mental abilities": verbal relations, word fluency, numerical skills, perceptual speed, reasoning, space, and memory (Anderson, 2005). Using monozygotic and dizygotic twins as well as an environmental variation due to adoption, it has been shown that about half of the intelligence is determined by heritability (Plomin and DeFries, 1998). The other half of one's intelligence is caused by the socio-economic environment with a prominent role of early childhood conditions. An important insight of the psychological research is that "it seems that genes are responsible for most of the overlap between cognitive skills" (Plomin and DeFries, 1998, p.66). I.e. the genetic endowment varies over individuals (or, to a lower degree, over families) but not over specific skills. When analyzing early childhood interventions, psychological studies often focus on "high-priced model programs operated by universities" which may not comparable to large-scale public programs (Barnett, 1998, p.204). Hence, besides the psychological research, there is still a need for an economic and social science perspective of the development of skills.

### 2.2. Economics

The first studies with an economic focus based on randomized trials, so-called early childhood intervention programs. The Head Start intervention programs ran by the U.S. Department of Health and Human Services in the 1960s are most influential in the economic literature. These programs cover a wide range of treatment measures, e.g. pre-school education, nutrition services, health care, and support of parents for disadvantaged children (HHS, 2010). In one of the largest Head Start programs, the Perry Preschool Program for African-American children, treated participants receive 2 years of pre-school education in the morning and teacher visits in the afternoon (Schweinhart et al., 2005). More than 35 years after the participation, Heckman et al. (2010a)

find longer educational attendance, higher labor market outcomes, and fewer records of criminal activities due to the program participation. Heckman et al. (2013) show that cognitive skills and personality traits are drivers of the long-lasting economic outcomes. The estimated yearly average returns to each Dollar invested range from 7% (Heckman et al., 2010b) to 17% (Rolnick and Grunewald, 2003). The estimated overall returns are up to \$13 per \$1 invested in early childhood intervention (Belfield et al., 2006). Moreover, it has been shown that early childhood investments exhibit the largest returns when followed by later investments in education (Heckman, 2006).

Experimental evidence for early childhood intervention programs is limited to the US because large randomized trials were not conducted elsewhere. Furthermore, the US intervention programs had a focus on children from disadvantaged families. The rewards of participating in language training may differ for those children. Due to this rather narrow focus, an increasing number of studies uses field data and quasi-experimental variations to investigate the effect of early childhood interventions. The main scope of those studies is the effect of center-based preschool attendance (i.e. kindergarten care) on children's development. Because of the limited time horizon in the data at hand, most studies only estimate short- and medium-term effects. Sylva et al. (2004) take data from the UK Effective Pre-school and Primary Education project and identify the gains of low and high quality pre-school education (measured as qualification of teachers) on math and reading skills. Using a value-added approach over several time periods and socio-economic background controls, the immediate effects are positive and significant for both pre-school qualities. However, the effect is only persistent for high-quality pre-schools. Fitzpatrick (2008) uses US state-year variations in the availability of kindergartens and employs differences-in-differences estimation in order to identify the effects of pre-school attendance on grade four abilities. Results suggest significant effects on math and reading test scores. Using propensity score matching, Warren and Haisken-DeNew (2013) show for Australian data that intermediate numeracy, reading, spelling, and writing performance increases due to preschool attendance. For Argentina, Berlinski et al. (2009) use a time-state differences-in-differences estimator and find a positive effect on third grade math and language scores. Analyzing Uruguayan data, Berlinski et al. (2008) provide evidence for an increased educational attainment due to pre-school attendance. For Germany, only two studies analyze the consequences of early childhood interventions. Spiess et al. (2003) and Landvoigt et al. (2007) estimate educational returns to kindergarten attendance. However, using a Probit model they "are not able to distinguish entirely between a selection and a treatment effect" (Spiess et al., 2003, p.261) and "do not interpret our estimates as causal" (Landvoigt et al., 2007, p.9).

### 3. Empirical Strategy

#### 3.1. Simple Differences

In order to identify the effect of language training on verbal intelligence, I employ simple differences and differences-in-differences estimation as well as (regression-adjusted) propensity score (PS) matching and combination of these techniques. Starting point is the linear-in-parameters version of a cognitive abilities production function for grammar skills, e.g. as in [Todd and Wolpin \(2003\)](#):

$$S_{ik}^g = \beta_0^g + \beta_1^g T_k^g + \beta_2^g \mathbf{K}_k + \beta_3^g \mathbf{B}_i + \beta_4^g A_i + \varepsilon_{ik}^g. \quad (1)$$

$S_{ik}^g$  denotes the skills of child  $i$  in kindergarten  $k$ . The superscript  $g$  marks the ability component, here grammatical skills. The intercept is given by  $\beta_0^g$ . The treatment indicator  $T_k^g$  is 1 if the intervention, i.e. the language training, takes place and 0 otherwise. Since the treatment was assigned on kindergarten level,  $T$  has the subscript  $k$ . The treatment effect is  $\beta_1^g$ . Other observable factors which influence grammar skills are kindergarten characteristics  $\mathbf{K}_k$  and the family background  $\mathbf{B}_i$ , their effects are denoted with  $\beta_2^g$  and  $\beta_3^g$  (the bold font marks matrix and vector notation). Beside observable factors, the level of skills might also be influenced by an unobservable component  $A_i$  with the effect  $\beta_4^g$ . This factor may depict innate abilities or the parent's emphasis concerning their children's development. The error term is  $\varepsilon_{ik}^g$ .

The simple differences approach regresses the level of grammar skills on the treatment and the observable confounders. This yields to the estimation of the treatment effect  $\hat{\beta}_1^g = \beta_1^g + \beta_4^g \sigma_{A,T}$ , where  $\sigma_{A,T}$  is the covariance between the language training participation and the unobserved factor. The estimator of  $\beta_1^g$  would be unbiased if either the treatment participation is independent from unobservable factors ( $\sigma_{A,T} = 0$ ) or if the unobservable factor does not affect the grammar skills ( $\beta_4^g = 0$ ). In other words,  $\beta_1^g$  identifies the causal effect if the conditional independence assumption (CIA)

$$S_{ik}^g(T_k^g = 1), S_{ik}^g(T_k^g = 0) \perp\!\!\!\perp T_k^g | \mathbf{K}_k, \mathbf{B}_i, A_i \quad (\text{A-1})$$

holds.<sup>2</sup> The CIA states that, given the covariates, the level of skills either with or without the language training –  $S_{ik}^g(T_k^g = 1)$  and  $S_{ik}^g(T_k^g = 0)$ , respectively – is independent (denoted with “ $\perp\!\!\!\perp$ ”) from the parents' choice for a kindergarten and the kindergarten's decision to offer a language training program. This assumption is violated if there is a selection due to the unobservable factor.  $\hat{\beta}_1^g$  overestimates the true effect of the language training e.g. if high-skilled parents with more intelligent offsprings are more in-

<sup>2</sup>This condition is sometimes also referred to as “unconfoundedness”.



terested in their child’s development and choose a kindergarten because of the higher educational quality – including the language program. On the other hand,  $\hat{\beta}_1^g$  may underestimate the true effect. This is the case if parents choose a kindergarten with additional language training in order to compensate for the child’s low innate skills. I.e. if Assumption (A-1) fails, it is not even clear whether the simple-differences estimation of  $\beta_1^g$  overestimates or underestimates the true effect of the language training.

### 3.2. Differences-in-Differences Estimation

The rich National Educational Panel Study at hand allows me to estimate a differences-in-differences (DiD) approach in order to relax the identifying Assumption (A-1). For this propose, I take the children’s mathematical skills into account, too. The production function of a child’s mathematical abilities is – analogous to Equation (1) – given by:

$$S_{ik}^m = \beta_0^m + \beta_1^m T_k^m + \beta_2^m K_k + \beta_3^m B_i + \beta_4^m A_i + \varepsilon_{ik}^m, \quad (2)$$

where the superscript  $m$  denotes the mathematical component of cognitive abilities instead of grammar. While the effects of the right-hand-side variables on skills may vary between grammar and math, the factors which influence the skill level can assumed to be the same. Hence, only the superscript of the coefficients changes compared to Equation (1). Since the language training is the only intervention and mathematical skills are supposed to be independent from it,  $T_k^m$  equals 0.<sup>3</sup> Subtracting Equation (2) from Equation (1) leads to grammar-math difference:<sup>4</sup>

$$\begin{aligned} \Delta S_{ik} &= S_{ik}^g - S_{ik}^m \\ &= (\beta_0^g - \beta_0^m) + \beta_1^g T_k^g + (\beta_2^g - \beta_2^m) K_k + (\beta_3^g - \beta_3^m) B_i + (\beta_4^g - \beta_4^m) A_i + (\varepsilon_{ik}^g - \varepsilon_{ik}^m) \\ &= \beta_0^d + \beta_1^m T_k^m + \beta_2^d K_k + \beta_3^d B_i + (\beta_4^g - \beta_4^m) A_i + \varepsilon_{ik}^d. \end{aligned} \quad (3)$$

The coefficients, now marked with a  $d$ , do no longer measure the effect of the respective factor on either grammar or math skills but give the effect on the difference between both levels of skills. If one factor, e.g. mother tongue does affect grammar skills more (less) than math skills, the coefficient has a positive (negative) sign.  $\beta_1^m$  drops in Equation (3) because of  $T_k^m = 0$ , thus the treatment effect is again given by  $\beta_1^g$ . This is the DiD estimator because it states the difference of the average differences between grammar and math skills of treated and untreated children. Since the child’s math skills are

<sup>3</sup>This is not an assumption but rather the result of a literature review and careful screening of the data at hand.

<sup>4</sup>Where  $\beta_0^d \equiv \beta_0^g - \beta_0^m$ ,  $\beta_2^d \equiv \beta_2^g - \beta_2^m$ ,  $\beta_3^d \equiv \beta_3^g - \beta_3^m$ , and  $\varepsilon_{ik}^d \equiv \varepsilon_{ik}^g - \varepsilon_{ik}^m$  in the last line.



used to control for innate factors that affect the grammar skills, the DiD approach could also be seen as a child-fixed effects estimation.<sup>5</sup>

The simple-differences Assumption (A-1) implies that the unobservable factor does not affect the *level* of grammar skills and the language training participation simultaneously. On the contrary, to achieve a causal estimate of  $\beta_1^g$  in Equation (3) one only needs to assume that the unobservable factor does not affect the *difference* between the grammar and math skills and the training participation simultaneously:

$$\Delta S_{ik}^g(T_k^g = 1), \Delta S_{ik}^g(T_k^g = 0) \perp T_k^g | \mathbf{K}_k, \mathbf{B}_i, A_i. \quad (\text{A-2})$$

The estimated DiD effect of the language training is  $\hat{\beta}_1^g = \beta_1^g + (\beta_4^g - \beta_4^m)\sigma_{A,T}$ . The estimator is unbiased if either  $(\beta_4^g - \beta_4^m) = 0$  or  $\sigma_{A,T} = 0$ . The latter is, again, not observable and every unlikely to be the case. The assumption that  $(\beta_4^g - \beta_4^m) = 0$  or after conditioning on observable factors

$$E(\beta_4^g - \beta_4^m | \mathbf{K}_k, \mathbf{B}_i) = 0 \quad (\text{A-3})$$

is the common trend assumption in terms of the DiD literature. This assumption would be violated if a child's innate abilities *and* its parents' behavior were not the same toward math and grammar skills. E.g. if parents learn that their child is more talented in languages and they decide to promote its talent. However, as mentioned in Section 2.1, psychometric studies provide evidence that innate abilities affect all components of intelligence in the same way (Plomin and DeFries, 1998). Moreover, even if innate factors affect grammar and math skills in different ways, parents need to observe their child's talent. Since most children were enrolled into kindergarten before the age of 3, it is unlikely that parents were able to observe the child's talent. Additionally, it seems not plausible that the parents' reaction would only be reflected in the choice of the kindergarten and not in other observable factors like number library visits.

### 3.3. Propensity Score Matching and Regression-Adjustment

Because the plausibility of Assumptions (A-1) and (A-2) is based on the quality of the conditioning factors and their implementation, I also employ a propensity score (PS) matching approach. PS matching compares the outcome variable of treated and untreated children with the same likelihood of being treated, i.e. participating in the

---

<sup>5</sup>The strategy to establish a DiD approach by taking the differences between subjects of skills is also employed in Jürges et al. (2005) and Jürges and Schneider (2010) as the authors analyze the effect of central exit exams. By converting the grammar and math skills for each child into separate observation and adding an indicator for the ability component, one can easily show that Equation (3) is equal to the "standard" DiD equation where the treatment effect is given by an interaction term.

language training. As algorithm to assign the best match, I use Epanechnikov kernel matching. Compared to the regression method, PS matching seems to be a more appropriate way to control for conditioning variables for two reasons. First, children outside the common support area are not taken into account. Hence, information on children who are very likely or very unlikely to participate in the language training because of extreme value in the covariates are neglected. Second, the regression versions of the simple differences and the DiD approaches presented above rely on the linear-in-parameter-specification of the skills production functions. PS matching avoids the assumption of a specific functional form in which the covariates enter the estimation. This feature makes the PS matching results more robust (Dehejia and Wahba, 2002).

A disadvantage of matching is that the validity is rather sensitive to specification of the PS. In the application here, it is a priori not clear if the PS only depends on kindergarten characteristics – since the decision to offer the language training is on kindergarten level – or if the child’s family background should also be considered when estimating the PS. To avoid misleading results because of a misspecification of the PS, I use a three-step procedure called regression-adjusted PS matching. In the first step, I run a Probit regression of the language training indicator on the full set of covariates.<sup>6</sup> I keep only those untreated observations with a PS equal or greater than the lowest PS of the treatment group (common support). In a second step, the Epanechnikov kernel algorithm is used to calculate weights according on how close the estimated PS of treated and untreated children is. The third step is a weighted regression of the outcome variable on the treatment and the observable factors  $K_k$  and  $B_i$  using the weights of step two. The regression-adjustment has the advantage that it does no longer solely rely on the specification of the PS estimation. The results are robust as long as either the PS estimation or the regression-adjustment is correctly specified. Bang and Robins (2005) refer this property to be “doubly robust”.

I apply (regression-adjusted) PS matching for both outcome variables – the level of grammar skills and the grammar-math difference. The method which combines DiD estimation and PS matching – that is semi-parametric regression-adjusted differences-in-differences matching – is developed in Heckman et al. (1997, 1998). The method is semi-parametric due to the use of the kernel algorithm. Recent applications are by Marcus (2012) and Schmitz and Westphal (2013) in the context of unemployment and health behavior as well as care giving and health.

All estimators provide the average treatment effect on the treated. That is the average benefit from participating in the language training program for those how do par-

---

<sup>6</sup>The full set of covariates includes kindergarten and teacher characteristics as well as home environment and the child’s activities. As a robustness check, I vary the specification of the PS. The results do not differ.

ticipate in language training. Because all children within a kindergarten receive the additional language training, I do not only estimate an intention to treat parameter.

## 4. Data

### 4.1. National Education Panel Study

The German National Education Panel Study (NEPS) provides information on the educational and occupational histories, socio-economic conditions, as well as non-cognitive and cognitive skills of 83,500 persons of all ages and at all educational stages. Six so-called starting cohorts cover infants, kindergarten-aged children, children in grades 5 and 9, university students, and adults (birth cohorts 1944 to 1990). For each starting cohort the data contain information on competence development, learning environments, educational decisions, migrational background, and returns to education ([Blossfeld et al., 2011](#)). The data used here are taken from the first two waves of the kindergarten starting cohort and cover the birth cohorts 2005 (702 children) and 2006 (2,027 children). The data were gathered in 2010/2011 (wave 1) and 2011/2012 (wave 2) and are available for research proposes since 2012 and 2013, respectively.<sup>7</sup> For the kindergarten starting cohort, the NEPS comprises several questionnaires. General information and information on the family's socio-economic status are given by either the mother or the father but concerning both parents. The kindergarten teacher's questionnaire asks about the qualification and the work experience of the teacher. A third questionnaire is answered by the head of the kindergarten. It includes detailed information on the size of the kindergarten, the share of children from families with a low socio-economic status, the share of children with migrational background, and the supply of other child care facilities in the area. The children only answer competence tests which assess various skills under supervision and instruction of the teacher. Questions regarding the child's personality traits were answered by the parent and the kindergarten teacher. For more information, see [Blossfeld et al. \(2011\)](#).

The final estimation sample contains 1,793 children. Starting with 2,729 observations, I drop children with missing parent, teacher, or head of the kindergarten questionnaire. Furthermore, children with missing information on the language training participation (or with more than one participation) and children who refuse to do the competence tests are dropped, too. Because my analysis is based on two (competence) test procedures and three interviews (parent, teacher, and head of kindergarten) the number of missing information is relatively high. Thus, I follow the approach of [Stuart \(2010\)](#)

---

<sup>7</sup>The NEPS release used here is from October 23, 2013 (doi:10.5157/NEPS:SC2:2.0.0).

and replace a missing value of a variable with 0 and add an indicator that the specific information is missing for the child. This procedure allows me to use the large number of control variables that is necessary to justify the plausibility of the identifying assumptions presented in Section 3.

## 4.2. Standardized Language Training Programs

The NEPS focuses on the children's language usage and their need for language training. A set of questions in the questionnaire answered by the head of the kindergarten deals with the existence of nationwide-established language training programs:<sup>8</sup>

- Does your facility offer a special nationwide-used language program? (yes/no)
- How many times was such a procedure offered? (number)
- Which procedure was offered? (name of the procedure)

214 kindergartens (out of 237) are in the final sample, 82 kindergartens offer one nationwide-used training procedure and 7 kindergartens offer more than one program. The questionnaire contains the four most common language training programs and the possible answers "other" and "I don't know". I define children as "treated" if the kindergarten conducts one of the programs. I drop children in kindergartens with more than one program in order to avoid that they drive the results. Table 1 provides an overview over the programs and their frequency. An example for such a program is "Kon-Lab", a software package which teaches the right usage of grammar by carrying out a picture comparison game. Beside these programs, the NEPS also includes various other information on educational measures and education improving activities within the kindergarten and at home, see Section 4.4. Using those information it is possible to determine whether other educational measure exist which may also affect children's abilities. By controlling for all other measures, I am able to isolate the effect of the nationwide-employed language programs. However, due to the rather low number of observations for each program, I am not able to run separate estimations.

## 4.3. Skill Measures

The only part of the NEPS that directly involves participation of the 3- and 4-year-old children are the competence tests. The kindergarten teacher undertakes tests in the fields of grammar, vocabulary, and science (wave 1), as well as math and logic

---

<sup>8</sup>For the exact questions and the possible answers, see [NEPS \(2013\)](#).

Table 1: Overview over the language training programs

Intervention	Number of kindergartens	Number of children affected
No language program	132	1,127
Participation in one nationwide used program	82	668
<i>where of</i>		
DELFIN 4	19	126
Würzburger Sprachprogramm	12	83
Deutsch 240	2	15
Kon-Lab	10	109
Other	31	236
Answer: "Don't know"	8	99
Participation share (in %)	62.1	59.3
Total	214	1,795

*Notes:* Own calculation based on NEPS–Starting Cohort 2. Information based on the final sample used for the baseline regressions.

(wave 2). The duration of the tests varies between 10 and 30 minutes and they were conducted in an atmosphere which is suitable for children at this age. Moreover, the NEPS includes also a meta test each competence. This meta test asked about the self-assessed performance of the children using five different smilies from sad to cheerful. For more information, see [Weinert et al. \(2011\)](#). As measure for verbal intelligence I use the grammar competence test. It contains 48 items which based on [Fox's \(2006\)](#) German version of the "Test for Reception of Grammar" by [Bishop \(1989\)](#). In this internationally well-established test procedure, children assign sentences to corresponding pictures. The procedure takes 10 minutes and tests "semantic, syntactic or morphological aspects of understanding grammatical structural forms [...] in a playfully arranged individual test situation" ([NEPS, 2011](#), p.7). The possible test score ranges between 0 and 48 points. In order to simplify the interpretation, I standardize the mean of the test score to 0 and the standard deviation to 1. The grammar test score is used as measure for verbal intelligence for two reasons. First, grammar skills do not directly affect math skills, while the vocabulary is more likely to affect the math performance. The latter is because child-appropriate math questions deal with examples and not with solving equations. The second reason for using the grammar test score is that grammar directly correspondences to verbal relation. Vocabulary, on the other hand, correspondences to world fluency – another intelligence component. The second competence used here, is the math test score. It refers to the mental abilities space and numerical knowledge.

The math test procedure contains 26 items dealing with quantity, space and shape, change in relationships, as well as data and chance (NEPS, 2011). The duration of the test is 20 minutes and the possible score ranges between 0 and 26 points. Again, the test score was standardized with mean 0 and standard deviation 1.

#### 4.4. Conditioning Variables

As shown in Section 3, the identification relies on the quality of the conditioning variables. Table A1 in the appendix gives an overview over the variables. To avoid a bad control problem – that arises when the conditioning variables were affected by the language program – only pre-treatment variables should be chosen. Since the pre- and post-treatment difference is achieved by a comparison of the subjects, pre-treatment variables could be defined as independent from the child’s participation in the language training.

This paper distinguishes between factors which affecting the child at the kindergarten and those which affect the child at home. The kindergarten variables include information on the facility as well as information on the child’s kindergarten teacher. These factors are likely to be crucial because the decision to conduct a language training program takes place at the kindergarten. I control whether the kindergarten is public organized or belongs to a nonprofit private organization (e.g. the church). The head of the kindergarten was asked to estimate the share of children within the kindergarten who belong to families with a low, medium, and high socio-economic status (SES). A medium SES being the comparison characteristic, it is controlled for the share of children with a low and a high socio-economic background. Moreover, the square of both shares is included. The head of the kindergarten was also asked about the share of children with migrational background. This share and the squared share are included, too. I also control for the monthly kindergarten fee and dummy variables indicating whether the kindergarten has a focus on foreign languages or math. Regarding the child’s kindergarten teacher, I take indicators for the gender, migrational background, and a high-school degree into account. Additionally, the number of participation in further education programs of the teacher is included.

Besides those factors which affect the child at the kindergarten, individual characteristics can assumed to be important as well. A major driver of the cognitive development is the child’s age. In spite of the fact that all children are aged 3 or 4, there may be even a variation between the age in month. Therefore, a full set of dummy variables indicating the age in month is included. Furthermore, it is controlled for the child’s gender and whether it is born in East Germany. Regarding the home learning envi-

ronment, the number of siblings, the parent's years of education (also squared), an employment indicator, and the monthly family income is considered. A crucial factor is the child's mother tongue. This is because the usage and knowledge of language is likely to affect only the grammar skills but not the math skills. Hence, the effect of the mother tongue is not canceled out in the DiD framework. The NEPS does not only include an indicator whether the child's native language is German but it also particularly asked for the mother tongue. About 85% of all children are German native speaker and for a relative large share of children Turkish or the Russian language is the mother tongue. However, more than 50 different native languages are covered in the NEPS. In order to make the languages comparable, I summarize them in 3 categories: the family of Slavic languages, the family of Turk languages, and other (mostly Germanic) languages. The share of children in the three groups is approximately the same. Instead of only a mother tongue indicator, I include dummy variables for the three groups of foreign languages – German is the reference group – in order to control the language distance. This procedure e.g. covers that Germanic languages are more related to German than e.g. Turk languages. I apply the same procedure also for the mother's and the father's mother tongue. Furthermore, an interaction term indicating that both parents are non-native speakers is considered. Since the mother tongue is highly correlated with migrational background but possibly more important, the paper does not include an indicator for born abroad. The NEPS has also a rich set of information on the children's activities. I control for reading books and fairy tales, writing letters, playing games including numbers, painting, and library visits (a dummy variable for each factor which is 1 when the activity is done at least daily). These factors cover the learning environment at home.

## 5. Results

### 5.1. Matching Quality

Matching aims at balancing the values of the conditioning variables between treated and untreated observations. Therefore, the success of the matching can be evaluated comparing the means of the covariates between the treatment and control group. The difference is standardized by the percentage of the square root of the average variances in the treatment and control group. [Rosenbaum and Rubin \(1985\)](#) call this the stan-



standardized bias. With respect to a successful balancing, it should not exceed 5%.<sup>9</sup> Table A2 in the appendix gives the means as well as the standardized bias for each conditioning variable (but the dummy variables for the age in month) before and after the Epanechnikov kernel matching with bandwidth 0.06. The median standardized bias was 5.95 before matching and has diminished to 1.80 after matching was conducted. This suggests that the kernel matching algorithm was successful in balancing treated and untreated conditioning variables over the groups. Another criterion to evaluate the matching procedure is to compare the distributions of the estimated PS. In an ideal scenario, for each treated child there would be a untreated child with the same probability of receiving the language training. Hence, the treated and untreated distributions of the PS would be equal. Figure A1 in the appendix shows the share of observations for every 2% interval of the PS for both groups. Since most PS intervals have a suitable number of observations in both groups, the usage of PS matching seems justified.

## 5.2. Baseline Results

Table 2 shows the estimation results for the baseline specifications. Each coefficient is taken from a different estimation and gives the estimated effect of the participation in the language training program on skills. Columns with an odd number give the effect on grammar skills (simple differences approach) while columns with an even number report the effect of the grammar-math difference (DiD approach). The specification reported in columns (1) to (4) give the results of ordinary least square regression. The first two columns do not consider control variables, thus the coefficients only reflect the raw correlation. In column (1) the participation in a language training program is negatively correlated with the level of grammar skills. This finding underlines the conjecture of a selection of children with low grammar skills into language training. Using the child's math skills as individual-fixed effect, the negative relationship vanishes in the specification of column (2). The participation in language training is positively correlated with the difference of grammar and math skills. Children who participate in the language training score on average 11.3% of a standard deviation (SD) higher in grammar than in math.<sup>10</sup> However, one can think of factors with affect the gram-

---

<sup>9</sup>The exact formula for the calculation of the standardized bias is

$$\text{Standardized bias}(X) = 100 \times \frac{\bar{X}_t - \bar{X}_c}{\sqrt{\frac{1}{2}(\sigma_{X_t}^2 + \sigma_{X_c}^2)'}}$$

where  $\bar{X}_t$  is the mean of variable  $X$  for the treatment group and  $\bar{X}_c$  the mean of the control group.  $\sigma_{X_t}^2$  and  $\sigma_{X_c}^2$  are the corresponding variances.

<sup>10</sup>The coefficients can be interpreted as percentage change of a SD because both test scores are standardized with mean 0 and SD 1.

mar but not the math skills and the language training participation simultaneously, e.g. mother tongue. Hence, the correlations in specifications (1) and (2) do not give the causal effect.

The regressions in columns (3) and (4) are analogous to (1) and (2) but include kindergarten and teacher characteristics as well as family background and activities at home as control variables. When the grammar skills are conditioned on these characteristics the negative relationship between grammar skills and language training can no longer be found in column (3). The change in the sign suggests that one can at least partly control for a selection in the language training because of a higher need by using the observable factors. Regarding the magnitude, the participation in language training is associated with an on average 11.1% of a SD higher grammar test score. Due to the rather large standard error, the coefficient is only significant at the 10% level. When individual-specific but subject-invariant unobservable effects were canceled out using the DiD specification in column (4), the effect is 13.7% of a SD and significant at 5%. The findings of specifications (3) and (4) give a hint that unobservable subject-invariant factors do not play a role after controlling for observable characteristics.

To overcome the rather strict specification of the abilities production function in the regression models and to make sure that treated and untreated children are comparable in the covariates at all, I apply Epanechnikov kernel matching with a bandwidth of 0.06 in columns (5) and (6). Only observations within the common support are compared. While the coefficient in specification (5) remains unchanged, the standard error increases which lead to insignificance.<sup>11</sup> The DiD coefficient of specification (6) is with a magnitude of 17.4% higher than the simple differences coefficient of specification (5) and significant at 1%. The coefficients in columns (7) and (8) state the results of regression-adjusted matching as described in Section 3.3. The simple differences estimator of the effect of language training on grammar skills is in line with the matching estimator of specification (5). However, the standard error is smaller, thus the effect is significant at 5%. With a point estimator of 16.1% of a SD the semi-parametric regression-adjusted DiD matching estimator is relatively high and significant at 1%.

All in all, the estimated coefficients in specifications (2) to (8) are in the same range. The magnitude of the DiD coefficient exceed the simple differences coefficients, however, the variation is rather small. I would interpret the similarity of the DiD estimate of specification (2) and the conditioned effect of the simple differences estimate of specification (3) as a hint, that unobserved factors – captured through the math skills in specification (3) – do not play a role once it is controlled for observable factors. Moreover, the small variation between regression and matching estimates suggests that the

---

<sup>11</sup>This is driven by the choice of clustered standard errors. In the case of robust – however, less appropriate – standard errors the coefficient would be significant at 5%.

Table 2: Baseline estimation results

	Raw difference		Regression		Matching		Matching+Regression	
	Simple diff. (1)	DiD (2)	Simple diff. (3)	DiD (4)	Simple diff. (5)	DiD (6)	Simple diff. (7)	DiD (8)
<b>Estimation results</b>								
Coefficient	-0.014	0.113*	0.111*	0.139**	0.122	0.174***	0.112**	0.161***
S.E.	(0.080)	(0.064)	(0.058)	(0.058)	(0.090)	(0.068)	(0.057)	(0.056)
<b>Conditioning variables</b>								
Kindergarten	--	--	✓	✓	✓	✓	✓	✓
Teacher	--	--	✓	✓	✓	✓	✓	✓
Family	--	--	✓	✓	✓	✓	✓	✓
Activities	--	--	✓	✓	✓	✓	✓	✓
<b>Observations</b>								
Total	1,793	1,793	1,793	1,793	1,793	1,793	1,793	1,793
Treated	668	668	668	668	668	668	668	668

Notes: Own calculations based on NEPS-Starting Cohort 2. Every cell states the estimated effect of the participation in a nationwide-used language training program on the grammar skills (odd columns), respectively, the grammar-math skill difference (even columns), obtained by a separate estimation. Outcome variables are standard normalized with mean 0 and standard deviation (SD) 1. Thus coefficients can be interpreted as percentage change of a SD of the outcome variable. Columns (1) and (2): ordinary least square regression without conditioning variables; columns (3) and (4): ordinary least square regression with conditioning variables; columns (5) and (6): propensity score (PS) matching with Epanechnikov kernel on the common support with bandwidth 0.06; columns (7) and (8): regression-adjusted matching using the PS as regression weights. Given sets of conditioning variables were used for Probit estimation of the PS and for regression-adjustment. Missing values in the conditioning variables were replaced with 0. The replacement is indicated by a dummy for each variable with missing values. Coefficients of the conditioning variables are not reported but available on request. Kindergarten-level clustered standard errors (S.E.) in parentheses. Significance: \*  $p \leq 0.1$ , \*\*  $p \leq 0.05$ , \*\*\*  $p \leq 0.01$ .

functional form of the skills production function is of minor importance. Combining matching and regression does not change the magnitude of the coefficients but their significance. An explanation for this finding could be that the additional regression reduces variance and, thus, lead to smaller standard errors. The baseline results of Table 2 provide clear evidence that children who participate in language training at the kindergarten exhibit a higher level of grammar skills in the range of 11% to 17% of a SD. Given the assumption that unobserved factors affect grammar and math in the same way, one could interpret the effect of language training causally.

### 5.3. Robustness Checks

The first robustness check picks up the bad control problem mentioned in Section 4. The child's activities at home are may not independent from its participation in the language training. An increased frequency of reading books could e.g. be caused by the language training program. To make sure that such a connection does not bias the results, Table A3 presents the results for the same specifications as in Table 2 but does not take the child's activities into account. Because the raw differences are unaffected by control variables, Table A3 leaves them out. The results are more or less the same. This indicates that a bad control problem does not occur in the specifications of Table 2.

As pointed out in Section 5.1, the matching results strongly depend on the quality of balancing of the conditioning variables. To avoid that the results are driven by the matching algorithm, I employ 5-to-1 nearest neighbor (NN) matching and radius matching, additionally. While kernel matching uses all untreated observation but assigns a different weight to each member of the control group according to the distance of the PS, 5-to-1 NN matching assigns to each treated observation only those 5 untreated observations with the closest PS.<sup>12</sup> Column (1) to (4) of Table A4 show the results of the NN matching. Since NN matching does not use all untreated children, the number of observations decreases. Using NN instead of kernel matching barely changes the results, the NN coefficient are even somewhat higher than the kernel coefficients. In specification (5) to (8) of Table A4, I apply radius matching. Instead of a fixed number of observations, radius matching uses all untreated observations within a given distance. I use a caliper of 0.5% of the PS. This is compromise between a high matching quality and a sufficient number of observations. Again, the coefficients underline the results of kernel matching.

---

<sup>12</sup>Furthermore, the observation is only used if distance between the PS of the treated and untreated observation is in range of 0.25 SD of the PS.

Since the choice of the matching algorithms seems to have no effect on the matching quality, I concentrate on kernel matching in order to keep the analysis as simple as possible. Table A5 gives the results of further robustness checks. Panels (A) to (C) change the bandwidth of the kernel matching. A higher bandwidth leads to a better fit of the estimated and the true density function and, therefore, a smaller variance. However, the higher bandwidth may smooth underlying features away which increases the risk of biased estimates as [Caliendo and Kopeinig \(2008\)](#) argue. In application here, the results are rather robust to the bandwidth choice as Table A5 shows. In panels (D) only significant variables are used to estimate the PS. For the regression-adjustment, I still condition on all covariates. The results do not change noteworthy. As argued before, the child's mother tongue is very likely to be an important factor. In panel (E), all observations with missing information on the mother tongue were dropped (instead of replaced with 0 and indicated by a dummy variable). 200 observation were lost, the coefficients are, however, in the same range. The analysis in panel (F) compares only treated and untreated children with the same value for binary mother tongue indicator. In other words, the PS matching algorithm compares only treated and untreated children within the same German/foreign language cells. For the simple differences specifications the coefficients drop sharply. The DiD coefficients are smaller but in the range of the baseline specification of Table 2. This provides evidence that there is a selection of non-native speaker into the language training. This finding is in line the negative correlation between language training and grammar skills in specification (1) of Table 2. Panel (G), the matching procedure matches exact on a German mother tongue indicator, the gender, and the year of birth. The results do not differ from panel (F). That underlines the prominent role of the mother tongue.

To reduce the risk of a self-selection bias, I limit the sample to children who visit the kindergarten for more than 2 years or who were enrolled aged younger than 2.5 years. The coefficients are in the same range as before; however, the standard errors decrease due to the smaller sample. The results are also robust to different duration and age thresholds. The results are available on request.

## 6. Conclusions

This paper estimates the effect of the participation in a nationwide-used language training program at the kindergarten on the level of contemporary grammar skills. While previous studies mostly focus on the effect of complete 2 or 3 year pre-school programs, this paper sheds light into the question how pre-school education works. Using the rich German National Educational Panel Study which includes information

on grammar and math skills, several estimation strategies are applied. The most natural way to assess the benefits of the language training is to compare the level of grammar skills between treated and untreated children. This simple differences effect is, however, likely to be biased because of a selection process into the treatment. Such a selection process could be driven by child-specific but subject-invariant unobservable factors, e.g. a general innate intelligence. To overcome the resulting bias of the simple difference estimation, the differences between grammar and math skills is used as outcome variable instead of the level of grammar skills. The resulting differences-in-differences estimates indicating a positive and highly significant effect of language training participation on grammar skills. The estimated effects range between 11% and 17% of a standard deviation. Therefore, I conclude, that the content of pre-school education is highly relevant for given children a good start. Since most children in OECD countries already receive kindergarten education, one should focus more on the content of this education in order to guarantee the efficiency of educational measures.

## References

- Anderson, J. (2005). *Cognitive Psychology and its Implications* (6 ed.). Basingstoke, UK: Palgrave Macmillan.
- Bang, H. and J. Robins (2005). Doubly Robust Estimation in Missing Data and Causal Inference Models. *Biometrics* 61(4), 962–972.
- Barnett, S. (1998). Long-Term Cognitive and Academic Effects of Early Childhood Education on Children in Poverty. *Preventive Medicine* 27(2), 204 – 207.
- Belfield, C., M. Nores, S. Barnett, and L. Schweinhart (2006). The High/Scope Perry Preschool Program: Cost-Benefit Analysis Using Data from the Age-40 Followup. *Journal of Human Resources* 41(1).
- Berlinski, S., S. Galiani, and P. Gertler (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics* 93(1-2), 219–234.
- Berlinski, S., S. Galiani, and M. Manacorda (2008). Giving children a better start: Preschool attendance and school-age profiles. *Journal of Public Economics* 92(5-6), 1416–1440.
- Bishop, D. (1989). *TROG – Test for Reception of Grammar*. Medical Research Council: Chapel Press.
- Blossfeld, H.-P., H.-G. Roßbach, and J. von Maurice (2011). Education as a Lifelong Process – The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft* 14, Special Issue.
- Caliendo, M. and S. Kopeinig (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys* 22(1), 31–72.
- Dehejia, R. and S. Wahba (2002). Propensity Score-Matching Methods For Nonexperimental Causal Studies. *The Review of Economics and Statistics* 84(1), 151–161.
- Fitzpatrick, M. (2008). Starting school at four: The effect of universal pre-kindergarten on children’s academic achievement. *The B.E. Journal of Economic Analysis & Policy* 8(1), 1–40.
- Fox, A. (2006). *TROG-D Test zur Überprüfung des Grammatikverständnisses*. Idstein: Schulz-Kirchner Verlag.
- Heckman, J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science* 312(5782), 1900–1902.

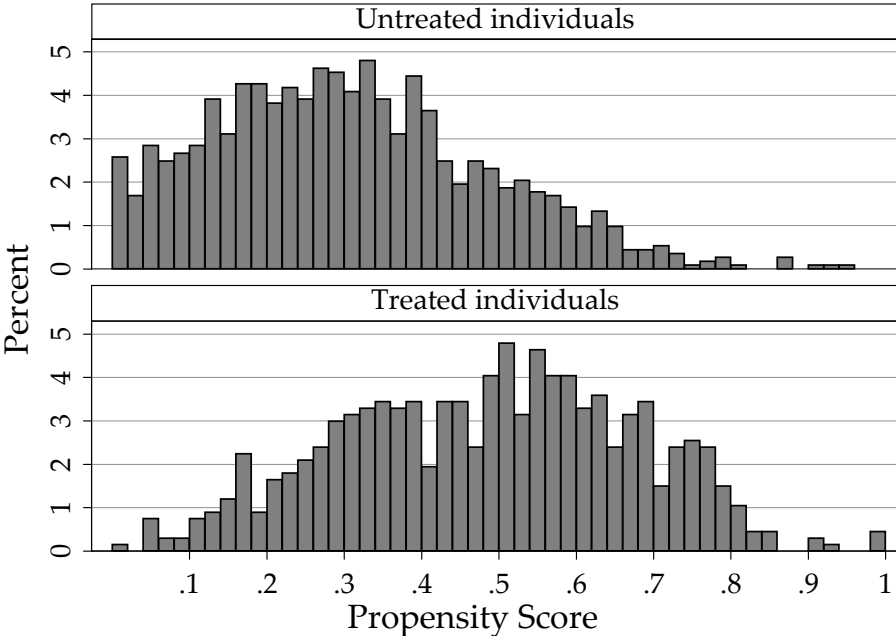


- Heckman, J., H. Ichimura, and P. Todd (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies* 64(4), 605–654.
- Heckman, J., H. Ichimura, and P. Todd (1998). Matching as an Econometric Evaluation Estimator. *Review of Economic Studies* 65(2), 261–294.
- Heckman, J., S. Moon, R. Pinto, P. Savelyev, and A. Yavitz (2010a). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics* 1(1), 1–46.
- Heckman, J., S. H. Moon, R. Pinto, P. Savelyev, and A. Yavitz (2010b). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics* 94(1-2), 114–128.
- Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review* 103(6), 2052–2086.
- HHS (2010). Head Start Impact Study. Final report, U.S. Department of Health and Human Services (HHS), Administration for Children and Families, Washington, DC.
- Jürges, H., Kerstin, and F. Büchel (2005). The Effect Of Central Exit Examinations On Student Achievement: Quasi-Experimental Evidence From TIMSS Germany. *Journal of the European Economic Association* 3(5), 1134–1155.
- Jürges, H. and K. Schneider (2010). Central exit examinations increase performance... but take the fun out of mathematics. *Journal of Population Economics* 23(2), 497–517.
- Landvoigt, T., G. Mühler, and F. Pfeiffer (2007). Duration and Intensity of Kindergarten Attendance and Secondary School Track Choice. ZEW Discussion Papers 07-051, ZEW–Zentrum für Europäische Wirtschaftsforschung/Center for European Economic Research, Mannheim.
- Marcus, J. (2012). Does Job Loss Make You Smoke and Gain Weight? SOEPpapers on Multidisciplinary Panel Data Research 432, DIW Berlin, The German Socio-Economic Panel (SOEP).
- NEPS (2011). Starting Cohort 2, Main Study 2010/11 (A12) Children in Kindergarten – Information on the Competence Test. Data documentaiton, National Educational Panel Study (NEPS) Research Data Center, Bamberg.
- NEPS (2013). Starting Cohort 2: Kindergarten (SC2), Wave 1 – Questionnaires (SUF Version 1.0.0). Data documentaiton, National Educational Panel Study (NEPS) Research Data Center, Bamberg.

- OECD (2013). Education at a Glance 2013: OECD Indicators. Report, Organisation for Economic Co-operation and Development (OECD), Paris.
- Plomin, R. and J. DeFries (1998). The Genetics of Cognitive Abilities and Disabilities. *Scientific American*, May, 62–69.
- Rolnick, A. and R. Grunewald (2003). Early Childhood Development: Economic Development with a High Public Return. *The Region* (2003-12), 6–12.
- Rosenbaum, P. and D. Rubin (1985). The Bias Due to Incomplete Matching. *Biometrics* 41(1), 103–116.
- Schmitz, H. and M. Westphal (2013). Short- and Medium-term Effects of Informal Care Provision on Health. Ruhr Economic Papers #426, Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Ruhr-Universität Bochum, Universität Dortmund, Universität Duisburg-Essen.
- Schweinhart, L., J. Montie, Z. Xiang, S. Barnett, C. Belfield, and M. Nores (2005). Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40. Report, High/Scope Educational Research Foundation, Ypsilanti, MI.
- Spiess, K., F. Büchel, and G. Wagner (2003). Children’s school placement in Germany: Does Kindergarten attendance matter? *Early Childhood Research Quarterly* 18(2), 255 – 270.
- Stuart, E. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science* 25(1), 1–21.
- Sylva, K., E. Melhuish, P. Sammons, I. Siraj-Blatchford, and B. Taggart (2004). The Final Report: Effective Pre-School Education. Report, The Effective Provision of Pre-School Education [EPPE] Project, London.
- Todd, P. and K. Wolpin (2003). On The Specification and Estimation of The Production Function for Cognitive Achievement. *The Economic Journal* 113(485), F3–F33.
- Warren, D. and J. P. Haisken-DeNew (2013). Early Bird Catches the Worm: The Causal Impact of Pre-school Participation and Teacher Qualifications on Year 3 National NAPLAN Cognitive Tests. Melbourne Institute Working Paper Series 2013-34, Melbourne Institute of Applied Economic and Social Research, Melbourne.
- Weinert, S., C. Artelt, M. Prenzel, M. Senkbeil, T. Ehmke, and C. Carstensen (2011). Development of competencies across the life span. *Zeitschrift für Erziehungswissenschaft* 14, 67–86.

# A. Appendix

Figure A1: Distribution of the propensity score by treatment status



Notes: Own calculations based on NEPS-Starting Cohort 2. Observations: 1,793. Epanechnikov kernel, common support, bandwidth 0.06.

Table A1: Variables and definitions

Variable	Definition
<b>Kindergarten characteristics</b>	
Public	=1 if public organized
Share low SES	Share of children with low socio-economic status (SES)
Share high SES	Share of children with high socio-economic status (SES)
Share migration	Share of children with migrational background
Monthly fee	Kindergarten's monthly fee in €
Focus: foreign language	Head of kindergarten reports focus on foreign language
Focus: math	Head of kindergarten reports focus on math
<b>Kindergarten teacher</b>	
Female	=1 if kindergarten teacher is female
High-school degree	=1 if kindergarten teacher has at least high-school degree
Migrational background	=1 if kindergarten teacher has migrational background
Number further educ	Number of further education programs
<b>Family background</b>	
Female	=1 if child is female
East-German	=1 if child lives in East Germany
Child: mother tongue Slavic	=1 if child's mother tongue is Slavic
Child: mother tongue Turkic	=1 if child's mother tongue is Turkic
Child: other mother tongue	=1 if mother tongue is neither German nor Slavic nor Turkic
Parent: mother tongue Slavic	=1 if parent's mother tongue is Slavic
Parent: mother tongue Turkic	=1 if parent's mother tongue is Turkic
Parent: other mother tongue	=1 if mother tongue is neither German nor Slavic nor Turkic
Partner: mother tongue Slavic	=1 if partner's mother tongue is Slavic
Partner: mother tongue Turkic	=1 if partner's mother tongue is Turkic
Partner: other mother tongue	=1 if mother tongue is neither German nor Slavic nor Turkic
Both parents foreign language	=1 if both parents do not speak German as mother tongue
Household income	Monthly household income in € 1000
Siblings	Number of siblings
Parent: years educ	Parent's years of education
Partner: years educ	Partner's years of education
Parent's × Partner's educ	interaction term of education years
Parent employed	Parent works full-time or half-time
Partner employed	Partner works full-time or half-time
<b>Child's activities at home</b>	
Reading (picture) books	=1 if child reads (picture) books at least once a day
Writing letters	=1 if child writes letters at least once a day
Games including numbers	=1 if child plays games including numbers at least once a day
Reading fairy tales	=1 if child reads fairy tales at least once a day
Painting	=1 if child paints at least once a day
Visiting library (times 100)	=1 if child visits a library at least once a day

Notes: Information based on NEPS-Starting Cohort 2.

Table A2: Means and standardized bias by treatment status

Variable	Treat. group mean	Control group mean		Standardized bias (%)	
		unmatched	matched	unmatched	matched
<b>Kindergarten characteristics</b>					
Public <sup>+</sup>	0.34	0.28	0.31	13.6	9.5
Share low SES	17.64	14.30	17.51	17.5	0.7
Share low SES squ	725.51	555.24	685.49	13.1	3.0
Share high SES	12.15	13.74	11.89	-9.2	1.6
Share high SES squ	439.22	482.94	415.07	-4.2	2.3
Share migration	26.44	18.42	26.09	33.3	1.5
Share migration squ	1346.70	871.93	1310.40	24.8	1.9
Monthly fee (in €)	76.61	78.01	74.81	-1.9	2.1
Focus: foreign language	0.07	0.14	0.06	-21.9	2.2
Focus: math	0.20	0.12	0.23	23.2	-6.3
<b>Kindergarten teacher</b>					
Female	0.93	0.89	0.94	15.9	-3.1
High-school degree	0.36	0.29	0.35	14.2	0.8
Migrational background	0.12	0.10	0.13	7.2	-1.1
Number further educ	1.96	1.99	1.91	-2.1	2.2
<b>Family background</b>					
Female <sup>+</sup>	0.50	0.47	0.52	6.2	-4.0
East-German <sup>+</sup>	0.09	0.19	0.10	-28.2	-1.1
Child: mother tongue Slavic <sup>+</sup>	0.04	0.03	0.04	3.2	-4.6
Child: mother tongue Turkic <sup>+</sup>	0.04	0.03	0.04	5.3	1.7
Child: other mother tongue <sup>+</sup>	0.05	0.03	0.04	6.9	1.4
Parent: mother tongue Slavic <sup>+</sup>	0.08	0.07	0.09	6.2	-0.8
Parent: mother tongue Turkic <sup>+</sup>	0.05	0.03	0.04	4.0	2.2
Parent: other mother tongue <sup>+</sup>	0.06	0.05	0.05	4.0	2.2
Partner: mother tongue Slavic <sup>+</sup>	0.07	0.05	0.08	8.7	-4.1
Partner: mother tongue Turkic <sup>+</sup>	0.05	0.04	0.04	6.2	2.8
Partner: other mother tongue <sup>+</sup>	0.07	0.06	0.07	2.9	1.0
Both parents foreign language <sup>+</sup>	0.14	0.11	0.14	8.2	0.4
Household income (in € 1000)	2.21	2.42	2.17	-8.1	1.4
Household income squ	7.97	16.73	8.03	-8.1	1.4
Siblings	1.02	0.97	0.97	5.4	4.6
Parent: years educ	10.72	11.08	10.48	-6.1	4.0
Parent: years educ squ	151.38	158.87	146.96	-7.7	4.3
Partner: years educ	8.79	9.34	8.76	-8.0	0.5
Partner: years education squ	127.52	138.64	126.35	-10.0	1.0
Parent's × partner's educ	122.21	131.54	119.51	-8.7	2.4
Parent employed <sup>+</sup>	0.49	0.52	0.50	-6.5	-2.0
Partner employed <sup>+</sup>	0.69	0.68	0.69	1.1	-1.1
<b>Child's activities at home</b>					
Reading (picture) books	0.66	0.66	0.66	-1.6	0.1
Writing letters	0.33	0.31	0.34	1.9	-1.8
Games including numbers	0.36	0.34	0.35	4.5	1.1
Reading fairy tales	0.09	0.10	0.10	-5.3	-4.3
Painting	0.22	0.20	0.21	5.7	2.7
Visiting library (times 100)	0.00	0.17	0.00	-0.7	-0.2
Median absolute bias				5.95	1.80

Notes: Own calculations based on NEPS-Starting Cohort 2. Observations: 1,793. Epanechnikov kernel, common support, bandwidth 0.06. Full set of age in month dummy variables included but not shown. Dummy variables in the table are marked with "+", the given mean is the share of 1.

Table A3: Estimation results without conditioning on child's activities at home

	Regression		Matching		Matching+Regression	
	Simple diff. (1)	DiD (2)	Simple diff. (3)	DiD (4)	Simple diff. (5)	DiD (6)
<b>Estimation results</b>						
Coefficient	0.111*	0.135**	0.126	0.179***	0.114**	0.167***
S.E.	(0.058)	(0.058)	(0.091)	(0.067)	(0.058)	(0.055)
<b>Conditioning variables</b>						
Kindergarten	✓	✓	✓	✓	✓	✓
Teacher	✓	✓	✓	✓	✓	✓
Family	✓	✓	✓	✓	✓	✓
Activities	--	--	--	--	--	--
<b>Observations</b>						
Total	1,793	1,793	1,793	1,793	1,793	1,793
Treated	668	668	668	668	668	668

Notes: Own calculations based on NEPS-Starting Cohort 2. Every cell states the estimated effect of the participation in a nationwide-used language training program on the grammar skills (odd columns), respectively, the grammar-math skill difference (even columns), obtained by a separate estimation. Outcome variables are standard normalized with mean 0 and standard deviation (SD) 1. Thus coefficients can be interpreted as percentage change of a SD of the outcome variable. Columns (1) and (2): ordinary least square regression with conditioning variables; columns (3) and (4): propensity score (PS) matching with Epanechnikov kernel on the common support with bandwidth 0.06; columns (5) and (6): regression-adjusted matching using the PS as regression weights. Given sets of conditioning variables were used for Probit estimation of the PS and for regression-adjustment. Missing values in the conditioning variables were replaced with 0. The replacement is indicated by a dummy for each variable with missing values. Coefficients of the conditioning variables are not reported but available on request. Kindergarten-level clustered standard errors (S.E.) in parentheses. Significance: \*  $p \leq 0.1$ , \*\*  $p \leq 0.05$ , \*\*\*  $p \leq 0.01$ .

Table A4: Nearest neighbor matching and radius matching

	Nearest neighbor				Radius			
	Matching		Matching+Regression		Matching		Matching+Regression	
	Simple diff. (1)	DiD (2)	Simple diff. (3)	DiD (4)	Simple diff. (5)	DiD (6)	Simple diff. (7)	DiD (8)
<b>Estimation results</b>								
Coefficient	0.153*	0.168**	0.129**	0.154***	0.136	0.142*	0.118**	0.135**
S.E.	(0.090)	(0.072)	(0.059)	(0.058)	(0.087)	(0.072)	(0.058)	(0.057)
<b>Conditioning variables</b>								
Kindergarten	✓	✓	✓	✓	✓	✓	✓	✓
Teacher	✓	✓	✓	✓	✓	✓	✓	✓
Family	✓	✓	✓	✓	✓	✓	✓	✓
Activities	✓	✓	✓	✓	✓	✓	✓	✓
<b>Observations</b>								
Total	1,553	1,553	1,553	1,553	1,714	1,714	1,714	1,714
Treated	668	668	668	668	668	668	668	668

Notes: Own calculations based on NEPS-Starting Cohort 2. Every cell states the estimated effect of the participation in a nationwide-used language training program on grammar skills (odd columns), respectively, the grammar-math skill difference (even columns), obtained by a separate estimation. Outcome variables are standard normalized with mean 0 and standard deviation (SD) 1. Thus coefficients can be interpreted as percentage change of a SD of the outcome variable. Columns (1) and (4): propensity score (PS) 5-to-1 nearest neighbor matching on the common support with caliper 0.25 SD of the PS; columns (5) and (8): radius matching with caliper 0.5% of the PS. Given sets of conditioning variables were used for Probit estimation of the PS and for regression-adjustment. Missing values in the conditioning variables were replaced with 0. The replacement is indicated by a dummy for each variable with missing values. Coefficients of the conditioning variables are not reported but available on request. Kindergarten-level clustered standard errors (S.E.) in parentheses. Significance: \*  $p \leq 0.1$ , \*\*  $p \leq 0.05$ , \*\*\*  $p \leq 0.01$ .



Table A5: Robustness checks for kernel estimation

Robustness check	Matching		Matching+Regression	
	Simple diff.	DiD	Simple diff.	DiD
<b>Bandwidth of Epanechnikov kernel</b>				
(A) Bandwidth = 0.03	0.140 (0.093)	0.181** (0.071)	0.123** (0.058)	0.168*** (0.057)
(B) Bandwidth = 0.09	0.115 (0.089)	0.178** (0.067)	0.109* (0.057)	0.156*** (0.056)
(C) Bandwidth = 0.12	0.108 (0.088)	0.161** (0.053)	0.107* (0.056)	0.151*** (0.056)
<b>Change of conditioning variables</b>				
(D) Only significant controls	0.128 (0.090)	0.176** (0.076)	0.128** (0.056)	0.157** (0.064)
(E) No mother tongue missings	0.093 (0.083)	0.163** (0.069)	0.091 (0.058)	0.150** (0.059)
<b>Exact matching on...</b>				
(F) ...German mother tongue	0.015 (0.076)	0.137** (0.069)	0.084 (0.058)	0.127** (0.062)
(G) ...mother tongue, gender, and birth year	0.024 (0.077)	0.122* (0.069)	0.079 (0.059)	0.115* (0.061)

*Notes:* Own calculations based on NEPS–Starting Cohort 2. Information based on the final sample used for the baseline estimations. Matching algorithm: Epanechnikov kernel matching on the common support. Panels (A) to (C): bandwidth as given in the table, all other robustness checks use a bandwidth of 0.06; panel (D): uses only variables significant at the 10% level to estimate the propensity score (PS), panel (E): observation excluded if mother tongue information is missing, 1.593 observations used; panel (F): exact matching on German mother tongue indicator; panel (G): exact matching on German mother tongue indicator, gender, and year of birth. Coefficients of the conditioning variables are not reported but available on request. Kindergarten-level clustered standard errors (S.E.) in parentheses. Significance: \* $p \leq 0.1$ , \*\* $p \leq 0.05$ , \*\*\* $p \leq 0.01$ .