AARHUS UNIVERSITY

Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Jensen, B., Jensen, P., & Rasmussen, A. W. (2017). Does Professional Development of Preschool Teachers Improve Children's Socio-Emotional Outcomes? *Labour Economics*, 45, 26-39. DOI: [10.1016/j.labeco.2016.11.004](https://doi.org/10.1016/j.labeco.2016.11.004)

Publication metadata

Title:	Does professional development of preschool teachers improve children's socio-emotional outcomes?
Author(s):	Bente Jensen, Peter Jensen, Astrid Würtz Rasmussen
Journal:	Labour Economics
DOI/Link:	https://doi.org/10.1016/j.labeco.2016.11.004
Document version:	Accepted manuscript (post-print)

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Does Professional Development of Preschool Teachers Improve Children's Socio-Emotional Outcomes?*

Bente Jensen[†]

Peter Jensen[‡]

Astrid Würtz Rasmussen[§]

Revised version, September 2016

Abstract: From 2011 to 2013 a randomized controlled trial has been run in Danish preschools to obtain evidence on improvements of early childhood education by providing training to the preschool teachers. The purpose of the intervention is to improve children's socio-emotional outcomes (measured by SDQ), especially for socially disadvantaged children. The teachers in intervention preschools received extra training in implementation of new pedagogical activities and principles, whereas teachers in control preschools did not receive any training. The results show improvements in several subscales of the SDQ scale. However, heterogeneity analysis is not able to demonstrate larger improvements for socially disadvantaged children than for other children.

JEL classification: I21, J13, J24

Keywords: Preschools, childcare, child development, intervention, RCT, professional development

* Acknowledgements: We thank the Danish Ministry of Children, Gender Equality, Integration and Social Affairs and the Centre for Strategic Research in Education (CSER) for financial support. Comments from Sven Bremberg, Frank Niklas, and Marianne Simonsen as well as participants at the EALE 2013 conference, the ESPE 2013 conference, the 8th Nordic Summer Institute in Labor Economics, the Applied Micro Seminar at DIW Berlin, and the Brownbag Lunch Workshop at Aarhus University are highly appreciated. We are further grateful for the comments from the editor and two anonymous referees. The usual disclaimer applies.

[†] Aarhus University, Tuborgvej 164, DK-2400 København NV, Denmark. Email: bj@edu.au.dk

[‡] Aarhus University, Fuglesangs Alle 4, DK-8210 Aarhus V, Denmark. Email: pje@econ.au.dk (corresponding author)

[§] Aarhus University, Fuglesangs Alle 4, DK-8210 Aarhus V, Denmark. Email: awr@econ.au.dk

1. Introduction

In recent years, it has become increasingly clear that the foundation for children's future success in life is laid very early in life (see e.g. Cunha et al., 2006). All children, and socially disadvantaged children in particular, benefit from early childhood education and care (Blau and Currie, 2006; Burger, 2010; Duncan and Sojourner, 2013). Positive effects of early childhood education and care are observed for cognitive skills as well as for socio-emotional skills of the children (Pianta et al., 2009; Heckman et al., 2010) and socio-emotional skills actually seem to be as important as cognitive abilities (Heckman and Kautz, 2012). Several studies have focused on long-term effects, documenting the importance of early childhood education for later life outcomes.¹ Much of this evidence relates to the US (see e.g. Duncan and Magnuson, 2013), but there is also evidence from Europe showing that early childhood education and care has positive effects (Anders et al., 2016; Melhuish et al., 2015).

The advantages gained from effective early interventions are best sustained when they are followed by continued high-quality learning experiences (Heckman, 2008). Hence, it is very important to provide high-quality childcare and to obtain more knowledge of how to design early childhood education programs to improve school readiness, increase educational attainment, and lay the foundation for future labor market success. Datta Gupta and Simonsen (2010) and Duncan and Magnuson (2013) point out that center-based care is generally more beneficial than home-based care and family daycare. Thus, a logical next step is to investigate possibilities for improving the quality of existing programs of formal center-based early childhood education and care in order to further enhance child outcomes.

In this study, we focus on professional development (PD) of preschool teachers as a potential route to improved childhood education in formal center-based care. PD can take many different forms and can occur as either pre-service or in-service programs, and the aim is to increase the knowledge

¹ A survey of many of these studies is given by Almond and Currie (2011).

and skills of preschool teachers. It can take the shape of workshops or collaborative group work between teachers, or be framed around technological innovations improving accessibility to and content of professional development interventions. Thus, PD covers a widespread range of activities such as formal coursework, consultation, coaching, and reflective supervision, and politicians increasingly express interest in and earmark resources for PD interventions in early childhood education settings, often focusing at disadvantaged children (Farran and Hofer, 2013).

Professional development is a promising way of improving the teaching quality of early childhood education and care and thereby improving both child outcomes (Schachter, 2015) and the quality of teacher-child interactions in preschools (Mashburn et al., 2008). Pianta et al. (2009) give an extensive survey of the (primarily US-based) evidence on the effects of preschool teacher qualifications showing that teacher qualifications and PD of preschool teachers are important for child development. Powell and Diamond (2010), Schachter (2015), and Zaslow et al. (2010) provide recent reviews of professional development interventions in which PD is associated with better child outcomes. For Denmark, Bauchmüller et al. (2014) find that a high percentage of BA-educated staff is positively associated with child development. Chetty et al. (2011) also find that teacher qualifications are positively associated with better cognitive outcomes. However, the literature on the effects on children of PD interventions is scarce if focus is not on US interventions. Results from a meta-analysis focusing on European PD interventions point in the direction of positive short-run effects on child outcomes (Jensen and Rasmussen, 2016), i.e. confirming the positive results found in US interventions.

In this study, we provide evidence from a randomized controlled trial on the effects of a PD intervention in Danish preschools, VIDA.² The intervention is targeted at all 3-5-year-old children in universal daycare but with special focus on disadvantaged children. The objective is for teachers

² VIDA is a Danish acronym for *Knowledge-based efforts for socially disadvantaged children in daycare*.

to provide the individual child with learning opportunities in ways that match the child's skill level. During a total of 17 days distributed across two years, preschool teachers are trained to improve their practices, refresh and update their pedagogical knowledge and expertise, and thereby improve child outcomes. The intervention trains preschool teachers to work on an evidence-based platform. The training sessions focus on theoretical aspects, and they further include educational training targeting the interplay between theoretical knowledge and practice. As an important part of the implementation into practice, the teachers learn to reflect on how to transfer the theoretical knowledge into practice. This final part is crucial to ensure that the teachers are able to implement the pedagogical methods and tools into renewed practices. A focus point is to improve every child's skills by being responsive to the child's needs and potentials, working with the individual child and peer groups (inclusion), and thus enhance the well-being and learning of *all* children, but hopefully with a relatively larger effect for socially disadvantaged children.

Our contribution is to investigate the effect of the VIDA intervention on children's socio-emotional outcomes and thus provide new empirical evidence on the effects of targeting children through professional development of preschool teachers. Furthermore, we focus on the effect of PD on children's socio-emotional outcomes, which is an under-investigated area, especially compared to studies on children's language and literacy skills (Schachter, 2015). The evidence is from a relatively large randomized controlled trial taking place in Denmark. The specific research questions are: Does pedagogical training based on learning and knowledge sharing (the VIDA intervention) improve socio-emotional outcomes of children in preschools? If yes, does the effect vary among different groups of children (e.g. due to gender) or based on differences of preschools (e.g. turnover rate of staff)?

Our results show that an intervention based on professional development of preschool teachers leads to improvements in the socio-emotional skills of the children, especially as regards emotional

symptoms and conduct problems. However, we *cannot* conclude that the intervention is successful in producing larger improvements for socially disadvantaged children than for other children.

The remainder of the paper is structured as follows. In Section 2, we briefly describe the institutional settings of preschools in Denmark, before we describe the VIDA intervention in detail in Section 3. Section 4 includes information on the experimental design and data along with tests on attrition and balancing of the sample. We present the empirical framework in Section 5 and the main results are presented in Section 6. Results on heterogeneity are included in Section 7, before the sensitivity analysis in Section 8. Section 9 discusses the various results and finally concludes with some policy recommendations.

2. Institutional Settings of Preschools in Denmark

In this section, we briefly present the organization of Danish preschools in the relevant period, 2011-2013. Danish parents use organized childcare on a large scale. For example, in 1998, almost 70% of 1-year-old children attended a formal childcare arrangement, and more than 90% of 3-5-year-olds attended childcare. Attendance rates have continued to rise and today 97% of 3-5-year-olds attend childcare (Statistics Denmark, 2015).

The majority of Danish childcare facilities are organized and operated by the municipalities. They organize family daycare where young children (typically 0-2-year-olds) are cared for by a child-minder, usually in her or his private home. Moreover, they provide center-based early childcare at nursery centers for the 0-2-year-olds, center-based preschools for children aged 3-5, and age-integrated institutions (for the 0-5-year-olds) with separate nursery and preschool compartments. In this study, we focus on 3-5-year-old children in center-based preschools and age-integrated institutions. These children are most often cared for in groups across ages, i.e. they are generally not subdivided into groups of 3-year-olds, 4-year-olds, and 5-year-olds.

Children are allocated to the institutions by the local municipality. On average, half of the institutions are run directly by the municipality; the others are independent but heavily subsidized by the municipality. The latter can be considered semi-private non-profit organizations, and they have more autonomy and more parental influence than a municipality-run institution, but they basically function as a municipal institution as they are under municipal monitoring and supervision and also subject to the same legislation on educational quality and provisional conditions for hygiene and safety.

Parental payment is the same for institutions within a municipality, but fees vary across municipalities. The municipality decides how much of the cost parents should bear, subject to an upper cap set by national law. Thus, parents usually pay around 20-30% of the actual cost of a preschool slot,³ but the price is reduced if siblings attend preschool at the same time, and there is no fee for very low-income families. A free slot is provided to families with an income level below DKK 164,101⁴ (for the year 2015). For single parents (sole providers) the income level below which a free slot is provided is further adjusted upwards to DKK 221,508.

3. The VIDA Intervention

The VIDA intervention is rooted theoretically in the bioecological model of human development, which focuses on the social contexts in which children live and views the child as an inseparable part of a wider system (Bronfenbrenner and Morris, 2006). According to this approach, the child's context (family, daycare, school, or neighborhood) is critical for understanding the child's learning and development. Recent research on life cycle skill formation focuses on dynamic complementarity with early investments in children generally making later investments more

³ With some variation across municipalities.

⁴ This is equivalent to USD 24,275 (July 26, 2016: USD 100~DKK 676)

beneficial and also emphasizes that non-cognitive skills, such as the child's socio-emotional skills, can enhance the cognitive skills (Heckman, 2006; Cunha et al., 2006).

The VIDA intervention focuses on improving the quality of the preschool environment by training the preschool teachers instead of targeting the children directly. Since the effects of the intervention on the children are indirect, it is crucial for the success of the intervention that the training and implementation in the preschools builds on responsive and supportive interactions between teacher and child.⁵ We measure the success of the intervention by changes in the socio-emotional development for the children.

The principal and one teacher from each intervention preschool are offered participation in a training program consisting of a series of educational sessions and workshops whose overall purpose is to qualify the preschool teachers to work on an evidence-based platform and provide the children with learning and development opportunities. As PD is a learning process for the teachers involved (Zaslow et al., 2010), explicit attention is devoted to ensuring that teachers understand and use the targeted practices of the VIDA approach to child development. The training program runs over a two-year period with a total of 17 full days of training⁶: 7 full-day educational sessions the first year (scheduled every second week over a four-month period), 7 full-day educational sessions the second year (again scheduled every second week over a four-month period) and 3 full-day practice-based sessions in the second year. In addition, the principals are each year offered a two-day course and a workshop on facilitating the organizational learning processes in the preschools. This setup ensures support and training for the participants to implement their new knowledge into

⁵ The design of the intervention is based on research on professional development using a perspective of innovation in the public sector (Sundbo, 2003; Jensen et al., 2007) and research on practice-based innovation and organizational learning (Ellström, 2010; Brown and Duguid, 1991). The intervention relies heavily on the principle of organizational learning (Gheradi, 2011), i.e. a process where some employees acquire new knowledge and use this knowledge to change work-related practices (e.g. introduce new activities and practices) that are then systematically adopted and practiced by all teachers in the preschool. The main tool for transmitting knowledge is critical-reflection groups.

⁶ The training program is coordinated by a steering committee with members from the participating municipalities and the education team. This committee is responsible for organizing the practical aspects of implementing the intervention.

practice in the daily work at the preschool, which is known to strengthen the effect of the intervention (Cameron, 2012; Pianta et al., 2009; Hamre et al., 2012).

The full intervention consists of three elements: knowledge, reflection, and activities. Knowledge is provided by the training program, then knowledge forms the basis for reflection about practices, and reflection subsequently leads to the development of new activities. These three elements alternate during the two-year period of the intervention.

The training program provides new and refreshes old knowledge based on a theoretical background in which children are viewed as active learners, in which resources (as opposed to deficits) are emphasized, and where changes are made to the whole system instead of focusing on just one child. Preschool teachers learn to analyze individual child progress and learning in the relevant social context in the preschool and they are trained in implementing a pedagogical principle of inclusive learning environments. The training program covers such topics as social heritage, the role of pedagogical principles, inclusion, exclusion, and reflection tools.⁷

Once teachers have completed the training, they are expected to transform the theoretical knowledge and their renewed analytical skills into activities. These activities are developed and implemented locally by preschool teachers in order to meet the needs of the individual child and the specific social issues in the particular group of children.⁸ An activity lasts minimum one month, and the entire preschool works together in implementing and evaluating the activity. Examples of activities include: improving children's management of conflict situations, improving children's friendships with peers by creating, entering, and maintaining social relationships, and improving

⁷ The teachers are provided with an IT-based reflection tool developed for the VIDA program.

⁸ Engaging preschool teachers in developing activities and practices at a local level ensures that challenges within the specific preschool are targeted, as e.g. improving efforts for socially disadvantaged children. Preschools in areas with many socially disadvantaged children do not get more resources from the intervention than other preschools. However, it is possible that such preschools generally are subsidized more from the municipality than other preschools.

children's communicative skills by telling narratives in the peer groups.⁹ During the activity, the teachers are supposed to evaluate the impact on the children (and adjust the activity, if necessary). Thus, this increases the preschool teachers' awareness of the impact of their approach on children's development and skills. An important aspect that is trained is the teachers' ability to communicate with the children in different situations (e.g. in the classroom or on the playground) in a way that guides and mentors the children towards the goal of the particular activity. An important part of this approach is that the teachers are trained in setting goals and conducting *ongoing* evaluation of the individual child as well as of the pedagogical process. The activity thereby transforms educational principles into new technical and pedagogical standards, which all teachers in the preschool follow.

Thus, the initial step of the teachers' learning process takes place as a top-down process that presents participants with predefined topics and content in workshops and courses. Through reflection, everyday experiences are related to the research-based knowledge, and the teachers are using this to change their practice. The process transforms what was initially top-down, course-based theoretical knowledge into bottom-up, practice-oriented teacher learning and innovation.

⁹ All activities take place in inclusive learning environments and the expected main drivers of children's learning are 1) adult-initiated learning activities with a range of selected themes based on the national curriculum, e.g. social and emotional (personal) development, language-stimulating activities, physical development, mathematics and science, creativity and arts, 2) child-initiated learning activities exploring the same range of selected themes, 3) supporting activities that train children's language and motor skills, 4) stimulation of children's learning and wellbeing through active involvement, 5) stimulation of children's curiosity and concentration through educational games, and 6) recognition of child progress by encouraging the child to explore new aspects of his/her personality in the peer group and to embark on new activities independently.

4. Experimental Design and Data

4.1 Experimental design

The VIDA intervention was implemented using a cluster-randomized controlled trial.¹⁰ Random assignment occurred at the preschool level. The participating preschools (and control preschools) were selected from three different municipalities in Denmark: Brøndby, Gentofte, and Randers.^{11 12} The randomization was done using a stratification procedure to ensure that the preschools were balanced with respect to their share of socially disadvantaged children. The share of socially disadvantaged children was calculated from information about the parents' level of education, social welfare dependency, family status and income (but the stratification procedure did not necessarily balance on each of those characteristics). This information was obtained from Statistics Denmark (for further details, see Jensen et al., 2011). A total of 58 preschools were randomly assigned to the intervention group and the control group, resulting in each group containing 29 preschools.

4.2 Data and validity of outcome measure

The data for this study comes from a variety of sources. Background information on preschools was obtained from questionnaires to the administrations of the municipalities, while background information on children and their parents was obtained from administrative registers at Statistics Denmark.

¹⁰ Two different versions of the intervention were tested in the RCT, but our study focuses only on the basis intervention. The other version of the intervention also focuses on parents and therefore it is less clear where potential effects on the children originate from. The results from this other version are reported in a report in Danish and are qualitatively similar to those presented here.

¹¹ Brøndby and Gentofte are municipalities located in the outskirts of Copenhagen, the capital city of Denmark, while Randers is north of Aarhus, the second largest city in Denmark.

¹² Originally, the intervention was also implemented in a fourth municipality. However, the data on child outcomes for the control group in this municipality showed unexpected high SDQ-scores. This amounts to relatively few children in the final sample of children present for the full period, and thus could not be detected in the original balancing tests. The available data did not allow us to investigate the reasons behind this further and the current study therefore only includes data from three municipalities.

Each child has a unique personal identifier (due to the Danish person registry) and we use this identifier to obtain data on child characteristics from administrative registers in Statistics Denmark. Furthermore, we link the child to his/her parents and in a similar way obtain data on the home and family background. For each child, we have information on age, ethnicity, and gender. For the parents, we have information on educational level, income, and marital status.

Our analyses control for observable characteristics of preschools that might be correlated with child outcomes. This information is provided by the administrations of the municipalities for each calendar year of the study. The information includes size of each preschool (i.e. the number of children ages 3-5 served), the number of staff, the composition of the preschools' staff with respect to gender and education, and the turnover rate of the staff.

We might expect that the effect of the intervention is influenced by the turnover of preschool staff. Turnover is generally high in Denmark due to flexible hiring and firing rules, and it is therefore not surprising that turnover is also high in Danish preschools (Bauchmüller et al., 2014). However, turnover is highest among the untrained staff, and since the untrained staff are not core participants in the professional development provided by the intervention, this should not be a concern. The turnover rate is included in the analyses to examine this question further.

Data on child outcomes are in this study restricted to socio-emotional skills which are measured by letting preschool teachers assess each child using the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997). The SDQ measures the strengths as well as the weaknesses of the child's socio-emotional development and it contains 25 questions. SDQ is a well-established instrument within child research and it has been used widely in other similar analyses, also as a pre- and post-measurement (Goodman and Goodman, 2009), and in intervention studies (e.g. Ford et al., 2009, Stewart-Brown et al., 2004, Mathai et al., 2003, and Nordhov et al., 2012). We treat the five

subscales of the SDQ as separate outcomes, i.e. we analyze the following five outcomes: 1) emotional symptoms, 2) conduct problems, 3) hyperactivity, 4) peer relationship problems, and 5) prosocial behavior. In addition, we analyze the total SDQ score which is obtained as the sum of the first four subscales, and the SDQ impact score which is obtained from the impact supplement of SDQ. This latter score reflects whether the preschool teacher thinks that the child has any emotional or behavioral difficulties and, if so, to what extent these difficulties result in distress and social impairment (for further details, see Goodman, 1999). Data on child outcomes was collected both immediately prior to the start of the intervention period in March 2011 (as a pre-test or baseline measure) and after the completion of the intervention period in March 2013 (as a post-test).¹³

The SDQ has high reliability and validity. A recent review of 48 studies of the psychometric properties of the SDQ concludes “that the psychometric properties of the SDQ are strong, particularly for the teacher version.” (Stone et al., 2010, p. 254). They also show that the teacher-reported SDQ has higher reliability over time compared to the parent version. More recently, Niclasen et al. (2012) has performed an extensive study of the validity of the Danish SDQ using a very large sample with more than 70,000 respondents. They similarly conclude “that the psychometric properties of the Danish version of the SDQ are strong, particularly for the teacher version.” (Niclasen et al., 2012, p. 6). The choice of the teacher-reported SDQ as our outcome measure obviously has both advantages and disadvantages. The main reasons behind the choice are the well-established validity of the SDQ for measuring children’s socio-emotional development and the feasibility of using it in a large-scale study as the current. Firstly, obtaining measures of children’s socio-emotional development requires familiarity with and knowledge about the children. Using the parent-reported SDQ could be an alternative choice, but children might exhibit different

¹³ Child outcomes were also measured in March 2012 as a midway measure. We do not use this information in our main estimations, but do instead exploit it to perform robustness checks on our results and for checking the validity of the teacher-reported SDQ measure.

behavior in different settings (home vs preschool). Secondly, limited resources prohibit using externally administered tests of the children (ideally blinded or masked to the intervention). Hence, we are using teacher-reported SDQ because it is a cost-effective and feasible solution to obtaining the outcome measure in this large-scale study. Thirdly, preschool teachers are able to refer to a broader experience with children and can draw on their educational background when assessing each child. Finally, using parents as respondents may lead to low response rates and non-representative samples, with underrepresentation of at-risk children (Demarest et al., 2013).

The main concern with using the teacher-reported SDQ is the potential bias that may be induced by letting the preschool teachers provide the outcome measure themselves. The teachers are obviously part of the experiment and therefore they are not blinded to the intervention. However, the teachers in the present study did not know that the SDQ measure would be used as the main outcome measure in later evaluations of the intervention. Shadish et al. (2002) also discuss various threats to validity, including those related to the outcome measure. They outline a number of different possible effects on the outcome measures, making it clear that it is not possible *a priori* to sign the direction of a potential bias on the impact measure.

If teachers in the intervention group are influenced (consciously or unconsciously) by their knowledge of the intervention and their expectations of children's situation, they might give the children a better assessment in the post-test (similar to a "Hawthorne effect"). This would bias our impact results towards a more positive effect of the intervention. Evidence of such behavior has been documented in earlier studies, e.g. by Ainsworth et al. (2015).

If teachers in the control group are trying to show (again, consciously or unconsciously) that they can do as well as the teachers in the intervention group, they might give children in the control group a better assessment (similar to a "John Henry effect"; Shadish et al., 2002; Duflo et al., 2007).

If such effects were present, then our impact results would be biased towards a less positive effect of the intervention. In that case, the estimated effect would be a lower limit on the possible impact measure and we might be less concerned about the bias.¹⁴ In this current study, the main concern is a potential positive bias of treated children's outcomes.

To check for possible "Hawthorne effects", we exploit the midway SDQ outcome information.¹⁵ An earlier study of a similar Danish intervention (Jensen et al., 2013) investigated possible "Hawthorne effects" as well as "John Henry effects", but did not find any evidence of such effects. We replicate their analysis and exploit that we have information on post-test measures available for children who did not participate in the full period, but only entered the preschools at a later time (at the collection of the midway information). This group of children has not been exposed to the intervention in the same magnitude as the children who have been in the preschools for the entire intervention period. Hence, we focus on whether there are significant differences in SDQ scores between those children that were observed by the preschool teachers for the full period and those children that entered later. If the effect from the full intervention were a change in children's behavior, significant differences would be expected between the 'full period' children and the 'late entrants'. If the changes were due to biased reporting from the preschool teachers after the completion of the intervention, no differences would be expected between the 'full period' children and the 'late entrants'. In the control group, no differences would be expected between children who were in the sample for the full period and those who entered later.

Table 1 shows results from estimations of models where the dependent variable is the normalized SDQ score and the main explanatory variable is an indicator variable for 'late entrants'. We control

¹⁴ Saretsky (1972) provides evidence of "John Henry effects" in an educational experiment where teachers in the control group performed much better than usual. Stewart-Brown et al. (2004) find evidence of similar effects among parents in the control group in a parenting program.

¹⁵ No information is available on which of the preschool teachers filled out the questionnaire for a given child, so it is not possible to account for preschool teacher fixed effects in the main analysis.

for important child characteristics (gender, age, and ethnicity). The top panel shows the results for the intervention group and the bottom panel shows the corresponding results for the control group. For the intervention group, we see from Table 1 that there are large differences between ‘full period’ children and ‘late entrants’. All these differences are in the right direction and most of them are significant (for hyperactivity we do not expect a significant difference since this SDQ subscale do not show any impact from the intervention in the later analysis; see Section 6). For the control group, we see from Table 1 that there are no significant differences (with the exception of the SDQ impact score at a 10% significance level) between ‘full period’ children and ‘late entrants’. This is as expected, as none of the children actually received the intervention. These results confirm that the intervention creates real differences in SDQ scores at the child level and that potential effects of the intervention are not “Hawthorne effects” due to biased assessments by the preschool teachers who have participated in the intervention. These results along with the general high validity of the SDQ measure makes us trust that our results using the teacher-reported SDQ actually measure the true effect of the VIDA intervention on child outcomes.

Table 1. Investigation of possible “Hawthorne effects” in the SDQ outcome measures

	Emotional symptoms	Conduct problems	Hyper- activity	Peer relation- ships	Prosocial behavior	Total SDQ score	SDQ impact score
<i>Intervention group (n=819)</i>							
Late entrants	0.148	0.127*	0.156	0.198*	-0.199*	0.202**	0.067
	(0.105)	(0.069)	(0.107)	(0.108)	(0.123)	(0.099)	(0.085)
<i>Control group (n=676)</i>							
Late entrants	0.025	-0.016	0.133	0.086	-0.055	0.083	0.224*
	(0.158)	(0.123)	(0.162)	(0.165)	(0.149)	(0.169)	(0.122)

Notes: * $p < 0.10$; ** $p < 0.05$. Standard errors in parentheses. Standard errors are adjusted for clustering at the preschool level. The dependent variables are normalized SDQ scores. The estimations also control for gender, age, age squared, and ethnicity. The results are robust towards including the full set of explanatory variables.

4.3 Analytic sample and attrition

The final analytic sample used in this study consists of all the children that are enrolled in intervention and control preschools for the full period of the intervention from March 2011 to March 2013. This ensures that the children in the intervention preschools have had the opportunity to experience the full effect of the intervention, which is the effect that is relevant from a policy perspective. For all these children, we have measurements of their socio-emotional skills both before the intervention starts (baseline) and after it is completed. This final sample consists of 686 children, with 396 in intervention preschools and 290 in control preschools.

The initial sample of children relevant for the analysis contains all children who were enrolled at the start of the intervention period and who were born after 2006.¹⁶ This initial sample consists of 875 children, with 472 in intervention preschools and 403 in control preschools. Thus, the overall attrition rate was 22%, but with an uneven distribution across the intervention and control groups with significantly different attrition rates of 16% and 28%, respectively. The main reason for attrition in this sample is residential moves, as families with young children are often in need of different (and larger) housing arrangements. When the families change their residential location, they typically move their children to a preschool near their new home. We have compared attritors with non-attritors (as suggested by Duflo et al., 2007) by balancing tests between the two groups, both for the total sample and separately for control and intervention groups. The results of these tests are reported in Appendix Table A.1 and show that most outcome measures and characteristics do not differ between attritors and non-attritors. This is in particular true for the control group where we find no differences despite having the highest attrition rate, whereas there are a few differences for the intervention group (which generally transfer into differences for the total sample as well). Regarding the outcome measures, we find differences for prosocial behavior and the SDQ impact score. A further look at the data shows that children with weaker outcomes (on these two outcome measures) are more likely to attrit from the intervention group, but as shown below in Table 2 all the pre-test outcome measures are still balanced across the control and intervention groups in the final sample. Regarding the child and family characteristics, we find differences for mothers' education and living with a single parent. Again, it is children with weaker characteristics that are more likely to attrit from the intervention group, which may introduce a downward bias in the estimated effects of the intervention if the intervention is more beneficial for socially disadvantaged children (as intended). On the other hand, it might introduce an upward bias if the

¹⁶ In addition, the preschools had older children enrolled at the start of the intervention period, but they were never potential participants in the final analysis, since they would be admitted to grade 0 of primary school before the end of the period.

intervention is more beneficial for children with stronger characteristics. However, these concerns turn out to be of less relevance, as we show in Section 7 that no such heterogeneous effects can be found.

In addition to the main analytic sample, we also have data on the children who either left or entered the preschools during the intervention period. These children are characterized by missing either a pre-test or a post-test measure and for those in intervention preschools by not having been exposed to the full intervention. Hence, we cannot include these children in our main analysis, but these data allow us to perform a midway analysis with an enlarged sample (see further in Section 8).

4.4 Summary statistics and balancing tests

In this section, we present summary statistics for our analytic sample in Tables 2, 3 and 4. These tables also report results of balancing test for the allocation to control and intervention groups. Corresponding tables for the initial sample, i.e. before attrition, are found in the Appendix Tables A.2, A.3 and A.4.

Table 2 shows means and standard deviations of pre-tests and post-tests for each normalized SDQ measure¹⁷ for the main analytic sample, separately for the intervention and control groups. There is some variation across the two groups and across outcome measures, but in general the means for the first four subscales show a clear decrease between pre-tests and post-tests, implying better outcomes for the children over time. This also leads the mean of the total SDQ score (equal to the sum of the first four subscales) to show a strong decrease over time. The fifth subscale, prosocial behavior, is measured on a positive scale and its mean shows a clear increase over time. Thus, on average children experience an improvement over time (i.e. during the intervention period), in the sense that

¹⁷ To facilitate comparison and assessment of effect sizes, all SDQ scores have been normalized by their standard deviation in the pre-test control group.

they experience fewer problems and have a better prosocial behavior.¹⁸ The balancing tests show that the pre-test outcome measures are balanced across intervention and control groups. However, the figures in the table indicate that the children in the intervention group experience larger improvements than the children in the control group. This is examined more closely in the econometric analysis below.

¹⁸ This is most likely related to the fact that children have more self-control and encounter fewer behavioral problems as they grow older, and it is as such not related to the intervention. This autonomous improvement over time or by age of the children has also been noticed in other studies using the SDQ scale (see e.g. Datta Gupta and Simonsen, 2010).

Table 2. Summary statistics for the normalized SDQ measures and balancing tests for pre-test measures

		Pre-test		Post-test	
SDQ subscale	Obs.	Mean	SD	Mean	SD
Emotional symptoms					
Control	290	0.80	1.00	0.81	0.96
Intervention	396	0.88	1.06	0.64	0.95
Balancing test (p-val.)		0.48			
Conduct problems					
Control	290	0.61	1.00	0.63	1.05
Intervention	396	0.68	1.05	0.40	0.74
Balancing test (p-val.)		0.50			
Hyperactivity					
Control	290	0.95	1.00	0.78	1.05
Intervention	396	1.00	1.06	0.67	0.95
Balancing test (p-val.)		0.59			
Peer relationship prob.					
Control	290	0.70	1.00	0.50	0.90
Intervention	396	0.71	1.03	0.37	0.69
Balancing test (p-val.)		0.89			
Prosocial behavior					
Control	290	2.59	1.00	3.01	0.84
Intervention	396	2.71	0.95	3.22	0.70
Balancing test (p-val.)		0.32			
Total SDQ score					
Control	290	1.08	1.00	0.95	1.08
Intervention	396	1.16	1.08	0.74	0.87
Balancing test (p-val.)		0.54			
SDQ impact score					
Control	284	0.37	1.00	0.55	1.17
Intervention	386	0.37	1.07	0.29	0.86
Balancing test (p-val.)		0.99			

Notes: All SDQ scores have been normalized by their standard deviation in the pre-test control group. Originally, all five subscales of SDQ are measured on a 0 to 10 scale, the total SDQ score is the sum of the first four subscales (hence, its original scale is from 0 to 40), and the SDQ impact score is measured on a 0 to

6 scale. Original scores are available from the authors or can be found in our previous working paper version. The balancing tests are for equality of means in the control and intervention group. The p-values are from regressions of outcome measures on an intervention indicator variable with standard errors adjusted for clustering at the preschool level (as in Avvisati et al., 2014).

Tables 3 and 4 show means and standard deviations for the background variables for the children in the main analytic sample, separately for the control and intervention groups. The child and family characteristics are presented for the year the child turns 3 and we mainly present variables that are included in the empirical analysis. 88% of the children are of Danish origin meaning that the mother, the father, or both parents are of Danish origin. Between 73% and 82% of mothers and fathers have at least 12 years of education, i.e. they have completed high school or vocational training on top of compulsory schooling. The administrative registers offer detailed information about living arrangements. 59-68% of the children are living in a household with married parents (at child age 3), and 12-22% are living with a single parent. The rest are living in cohabiting households either with both or only one of the biological parents. Hence, 78-88% of the children live with two parents, who are either married or cohabiting.

Looking at the preschool characteristics in Table 4, it is seen that the average size of the preschools is around 75 children, the staff-child ratio is around 0.20, and more than half of the staff is BA-educated. The average stability of the staff is close to 90%, meaning that the average annual turnover rate is around 10%. In the empirical analysis, we include the stability of staff at the start of the intervention period to avoid issues of endogeneity. This initial stability of staff is highly correlated with subsequent stability of staff during the intervention period, with a year-to-year correlation of around 0.70 for the intervention group.

Table 3. Summary statistics (by group) and balancing tests for child and family characteristics

Variable	Control			Intervention			Balancing tests
	Obs.	Mean	SD	Obs.	Mean	SD	p-values
Boy (0/1)	290	0.56	0.50	396	0.49	0.50	.12
Danish origin (0/1)	288	0.88	0.34	392	0.88	0.33	.86
12 years of education, mother (0/1)	281	0.79	0.41	388	0.82	0.39	.66
12 years of education, father (0/1)	278	0.73	0.45	382	0.82	0.38	.05*
Income less than 10% of income distribution (0/1)	266	0.09	0.29	368	0.10	0.30	.82
Parents married (0/1)	288	0.59	0.49	392	0.68	0.47	.09*
Living with single parent (0/1)	288	0.22	0.41	392	0.12	0.32	.01**

Notes: Background information for children is collected at the end of the year the child turns 3. The balancing tests are for equality of means in the control and intervention group. The p-values are from regressions of row variables on an intervention indicator variable with standard errors adjusted for clustering at the preschool level (as in Avvisati et al., 2014). * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

The balancing tests in Tables 3 and 4 show that most background characteristics are balanced across intervention and control groups. They only test that rejects equality of the means at a 5% significance level is for the indicator variable for living with a single parent, while the test for father's education is on the borderline. The children in the control group have slightly weaker characteristics, in the sense that they are more often living with a single parent and their fathers are less educated. These baseline differences are not in themselves a concern, but they underline the need to include these characteristics as explanatory variables in the empirical analysis. In addition, we cannot rule out the possibility of unobserved differences, but our empirical specification will be able to control for (some of) these differences.

Table 4. Summary statistics (by group) and balancing tests for preschool characteristics

Variable	Control			Intervention			Balancing tests
	Obs.	Mean	SD	Obs.	Mean	SD	p-values
Size of preschool (no. of children)	282	78.52	28.53	396	71.77	23.98	.44
Staff-child ratio	277	0.19	0.04	396	0.21	0.04	.33
Share of BA-educated staff	277	0.56	0.08	396	0.60	0.11	.18
Stability of staff (1- annual turnover)	277	0.90	0.14	396	0.86	0.14	.33
Change of principal of preschool							
(0/1)	277	0.19	0.39	396	0.04	0.20	.17
Stable preschool (turnover rate below							
median) (0/1)	277	0.55	0.50	396	0.46	0.50	.57

Notes: Background information for preschools is collected at the start of the intervention period. The balancing tests are for equality of means in the control and intervention group. The p-values are from regressions of row variables on an intervention indicator variable with standard errors adjusted for clustering at the preschool level (as in Avvisati et al., 2014). * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

It is also seen from the tables that not all variables are observed for all 686 children. In a few cases, we do not have information about e.g. ethnicity, parents' education or characteristics of the staff of the preschool. Such missing values occur in around 3-4% of all observations. Observations with missing values are left out in the relevant estimations, thereby reducing the sample size in specific estimations.

5. Empirical Methods

To examine the effects of the intervention on child outcomes, we employ a value-added specification of an educational production function (see Todd and Wolpin, 2007, for further details

and discussion of this specification). The basic econometric model is specified in the following way:¹⁹

$$(1) \quad y_i = \mathbf{X}_i\boldsymbol{\beta} + \delta I_i + \gamma y_i^{pre} + u_i,$$

where y_i is the post-test outcome measure for child i and y_i^{pre} is the pre-test (baseline) measure. Including this measure allows us to control for unobserved characteristics that might be correlated with prior achievement. $\mathbf{X}$ is a vector of explanatory variables that include child and family characteristics (gender and ethnicity of the child, parental education level, and parents' marital status). It also includes a set of preschool characteristics (the size of the preschool, the staff-to-child ratio, preschool teachers' educational composition, turnover rate of teachers, and change of principal). These variables are included to control for individual- and preschool differences that might influence the outcome measures. I_i is an indicator variable for whether the child is enrolled in an intervention preschool, which makes δ our main parameter of interest.

The effect of the intervention might have been estimated by the simple difference of the means of the post-test outcome measure for the intervention and control groups (due to the random assignment),²⁰ but we have chosen to include as many explanatory variables as possible in the specification to account for any baseline differences in the final analytic sample and to increase the statistical power (thereby following the advice of Lipsey et al., 2012). Hence, the estimated coefficient of interest in Equation 1 is δ which measures the direct effect of the intervention net of any other characteristics at the child, family and preschool levels. In addition to the estimated intervention effects, we also calculate effect sizes by dividing the estimated coefficients by the full

¹⁹ A similar specification has been employed by Claessens et al. (2014). Raver et al. (2009) also model children's behavior problems in preschool by a specification that is essentially similar although their exposition of the model is different.

²⁰ The results of this simple estimation are shown in Table 5 in Section 6.

sample standard deviation for the dependent variable. Thereby we allow comparisons across effects and to other studies.

We estimate the model separately for each outcome measure, i.e. the five subscales of the child's SDQ index. Since the randomization occurs at the preschool level and we also include other preschool characteristics in our specification, we adjust the standard errors of the estimates to take account of this clustering. The clustered standard errors also account for any other non-independence of children within preschools.

We first use the model specified above to estimate the main effects of the intervention. We further investigate whether these effects are heterogeneous across different groups. To avoid estimating the effects on very small subsamples (and thereby losing explanatory power), we have chosen to include interaction terms between the specific subgroups and the intervention indicator.

To check the robustness of our estimates, we also employ two alternative specifications. The first of these is the more restrictive version of the value-added model where the parameter on the pre-test outcome measure is set to one ($\gamma = 1$) and the resulting dependent variable becomes the gain in the outcome measure from pre-test to post-test. The gain is then expressed solely as a function of the explanatory variables. The second alternative specification is based on treating the data as individual-level panel data (with only two observations, from the pre-test in 2011 and the post-test in 2013, respectively) and specifying a standard panel data model with unobserved individual-specific fixed effects. The intervention effect can then be estimated as the coefficient to a time-varying intervention indicator (see Imbens and Wooldridge, 2007, for further details). In both of these alternative specifications, the standard errors of the estimates are adjusted for clustering at the preschool level.

6. Main Effects

We estimate the main effects of the intervention using the value-added specification outlined in the previous section. Our main results are based on the analytic sample consisting of children who participate in the full period from 2011 to 2013. For these children, we have measurements of their socio-emotional skills both before the intervention starts and after it has completed. We use the latter as the dependent variable and the former as an additional explanatory variable, as prescribed for the value-added model. The effect of the intervention is given by the estimated coefficient of the intervention preschool indicator, δ , which basically is the mean difference between children in the intervention and control preschools, net of any other characteristic at the child, family and preschool levels. The effect is measured after the intervention has been completed in 2013.

The estimation results are reported in Table 5 along with simple mean differences in post-test outcomes between children in the control and intervention groups. Since normalized SDQ measures are used, all results on effects are given as effect sizes (i.e. measured in standard deviations of the outcome). The middle panel shows the estimates without any additional explanatory variables (i.e. only the pre-test (baseline) value of the outcome measure has been included along with the intervention indicator) and the bottom panel shows the estimates with a full set of explanatory variables included to control for child, family and preschool characteristics.²¹ As expected, the magnitude and significance of the simple mean differences reported in the top panel are remarkably similar to the estimates of the value-added model without additional explanatory variables. Moreover, the estimates from the two value-added models are also similar in magnitude and significance, which is reassuring since it indicates that the initial randomization has not been undermined. However, the preferred set of estimates is from the model with all explanatory

²¹ We have also estimated a value-added model with family income (in the lowest decile) as an additional explanatory variable. Since data on family income is missing for many observations, this specification can only be estimated on a substantially smaller sample and this further sample selection could lead to bias. Reassuringly, the estimates from this specification are only marginally different from the reported estimates.

variables included. We report only the coefficient estimates for the pre-test value of the outcome measure and the intervention indicator, but coefficient estimates for the other explanatory variables are available from the authors upon request.

The overall picture is that the intervention has had positive effects on several SDQ subscales (shown by negatively estimated coefficients for all the outcomes, except the positive coefficient estimate for prosocial behavior, which is in the expected direction). The results show that the intervention significantly improves two of the five SDQ subscales: emotional symptoms and conduct problems. The three remaining subscales are also improved: for peer relationship problems and prosocial behavior at a 10% significance level, but for hyperactivity the effect is not significant. In addition, it should be noted that both the total SDQ score (representing the total of the first four subscales) and the SDQ impact score are also significantly improved by the intervention.

The effect sizes of all the significant improvements are around 0.25. Effect sizes in the range 0.2-0.4 are considered moderate by Hattie (2008) and compare favorably to effect sizes on similar outcomes found in experimental evaluations of Head Start and Early Head Start which are in the range 0.11-0.19 (see Almond and Currie, 2011). They are, however, below the effect sizes found by Raver et al. (2009) who find improvements in teacher-reported behavior problems corresponding to effect sizes in the range 0.53-0.89, but these are only for low-income children.

Table 5. Main effects (effect sizes), from estimations of value-added models, on the sample of children for the full period, 2011-2013

	Emotional symptoms	Conduct problems	Hyper- activity	Peer relation- ships	Prosocial behavior	Total SDQ score	SDQ impact score
<i>Simple differences of post-test means (t-tests)</i>							
Intervention	-0.168** (0.074)	-0.232*** (0.068)	-0.116 (0.077)	-0.130** (0.061)	0.206*** (0.059)	-0.216*** (0.075)	-0.256*** (0.078)
<i>Value-added specification with only pre-test outcome measure as explanatory variable</i>							
Intervention	-0.191** (0.091)	-0.261*** (0.073)	-0.140 (0.102)	-0.134 (0.081)	0.174** (0.084)	-0.248** (0.102)	-0.247** (0.105)
Pre-test outcome	0.298*** (0.046)	0.387*** (0.043)	0.430*** (0.038)	0.224*** (0.049)	0.284*** (0.040)	0.418*** (0.036)	0.321*** (0.047)
<i>Value-added specification with full set of explanatory variables</i>							
Intervention	-0.234** (0.110)	-0.257*** (0.080)	-0.125 (0.101)	-0.154* (0.080)	0.169* (0.089)	-0.266** (0.111)	-0.248** (0.098)
Pre-test outcome	0.292*** (0.046)	0.362*** (0.042)	0.368*** (0.040)	0.226*** (0.052)	0.242*** (0.044)	0.384*** (0.033)	0.298*** (0.050)

Notes: * p<0.10; ** p<0.05; *** p<0.01. Standard errors in parentheses. Standard errors are adjusted for clustering at the preschool level. The full set of explanatory variables consists of: child gender, child ethnicity, mother's education (at least 12 years), father's education (at least 12 years), living with single parent, size of preschool, staff-child ratio, share of staff with education, staff stability, change of principal, and municipality indicators. No. of observations is 686 for the top panels and 639 for the bottom panel.

7. Heterogeneous Effects

Our main results show that several subscales of SDQ are improved by the intervention. However, these main effects reflect the average effects for the intervention group as a whole, and there may be important differences in the effects across various subgroups. In particular, we are interested in investigating whether the intervention (as intended) is more beneficial for disadvantaged children than for other children. Therefore, we now turn to estimations of specifications where we allow the effects to be heterogeneous across subgroups.

To avoid estimating the effects on very small subsamples (and thereby losing explanatory power), we have chosen to use the value-added specification of the previous section augmented with an interaction term between the specific subgroup and the intervention indicator. In Table 6, we report the results from a number of different specifications investigating such heterogeneous effects, representing differences across children's gender, parents' education, children's ethnicity, and stability of the staff of the preschool.²²

In Table 6, we consider 5 potential sources of heterogeneous intervention effects and provide a total of 35 estimates of interaction coefficients, each of which represents a potential differential effect of the intervention. Thus, it is a relevant concern that some of these coefficients may be significantly different from zero by mere chance (under the joint null hypothesis of no differential effects).

Therefore, we account for this multiple testing by computing adjusted p -values using Bonferroni-Holm corrections (Duflo et al., 2007; Kling and Liebman, 2004). In the table, we report the results of both the separate significance tests (known as per-comparison significance) and the adjusted significance tests (known as family-wise significance; here we consider all 35 coefficients as one family) for the interaction coefficients.

²² There are many other possibilities for defining subgroups, but we have restricted the number of different specifications here, since we want to avoid arbitrary group definitions and the danger of too much data mining. In addition, using many different specifications would amplify the problem of multiple testing (see further details later).

Table 6. Heterogeneous effects (effect sizes), from estimations of value-added models, on the sample of children for the full period, 2011-2013

	Emotional symptoms	Conduct problems	Hyper- activity	Peer relation- ships	Prosocial behavior	Total SDQ score	SDQ impact score
<i>Panel A: Children's gender (n=335 boys vs. 304 girls)</i>							
Intervention	-0.246 (0.148)	-0.146 (0.096)	-0.091 (0.111)	-0.105 (0.137)	0.194* (0.104)	-0.207 (0.146)	-0.280* (0.115)
Boy * intervention	0.023 (0.159)	-0.204(**) (0.098)	-0.063 (0.143)	-0.091 (0.154)	-0.047 (0.110)	-0.109 (0.139)	0.059 (0.143)
<i>Panel B: Mother's education, at least 12 years (vs. less than 12 years) (n=525 vs. 114)</i>							
Intervention	-0.138 (0.223)	-0.388* (0.204)	-0.134 (0.216)	-0.048 (0.187)	0.012 (0.210)	-0.231 (0.238)	0.022 (0.163)
Medu * intervention	-0.116 (0.211)	0.158 (0.193)	0.011 (0.193)	-0.128 (0.167)	0.188 (0.197)	-0.042 (0.197)	-0.326(*) (0.176)
<i>Panel C: Father's education, at least 12 years (vs. less than 12 years) (n=506 vs. 133)</i>							
Intervention	-0.096 (0.219)	-0.198 (0.167)	-0.028 (0.197)	0.127 (0.160)	0.061 (0.146)	-0.064 (0.209)	-0.218 (0.189)
Fedu * intervention	-0.170 (0.224)	-0.074 (0.164)	-0.125 (0.175)	-0.348(*) (0.181)	0.133 (0.134)	-0.250 (0.192)	-0.045 (0.214)

Panel D: Ethnicity, Danish child (vs. child of non-Danish origin) (n=569 vs. 70)

Intervention	0.079	-0.161	0.078	-0.065	0.090	-0.024	0.061
	(0.220)	(0.162)	(0.210)	(0.193)	(0.238)	(0.217)	(0.196)
Danish *	-0.351	-0.108	-0.228	-0.100	0.088	-0.271	-0.346
intervention	(0.235)	(0.188)	(0.232)	(0.227)	(0.212)	(0.246)	(0.226)

Panel E: Preschool's stability of staff, high stability (vs. low stability) (n=318 vs. 321)

Intervention	-0.159	-0.220*	-0.292*	-0.097	0.232*	-0.293*	-0.215
	(0.120)	(0.115)	(0.148)	(0.117)	(0.130)	(0.154)	(0.131)
Stability *	-0.136	-0.068	0.301	-0.103	-0.114	0.048	-0.059
intervention	(0.171)	(0.135)	(0.182)	(0.144)	(0.179)	(0.182)	(0.197)

Notes: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. Standard errors in parentheses. Standard errors are adjusted for clustering at the preschool level. Asterisks for the interaction coefficients are given in () or [], where the former is used for per-comparison significance and the latter for multiple testing (family-wise) significance. Notice that no asterisks in [] are given in the table, since no interaction coefficients are significant using multiple testing significance. All specifications include: pre-test outcome measure, child gender, child ethnicity, mother's education (at least 12 years), father's education (at least 12 years), living with single parent, size of preschool, staff-child ratio, share of staff with education, staff stability, change of principal, and municipality indicators. Total no. of observations is 639, except for the SDQ impact score where it is 608.

None of the adjusted p -values shows any significantly different effects across subgroups defined on the basis of child, family and preschool characteristics. Even the per-comparison significance tests only reveal very few significant differences. Hence, the analysis of heterogeneous effects is not able to demonstrate larger improvements for socially disadvantaged children than for other children. There are, however, a number of interesting patterns to be noticed in some of the augmented specifications.

When we look at the differential effects due to parents' education and ethnicity there is some indication that the children with highly educated parents and of Danish origin are benefitting the most from the intervention. We observe a consistent pattern in the coefficients of the interaction terms indicating that the effects of the intervention are stronger for children of highly educated parents (fathers in particular) and for Danish children. Furthermore, the effects of the intervention seem to be very low for children of non-Danish origin. However, it should be noted that the group of children of non-Danish origin is quite diverse and consists of children from very different backgrounds, e.g. non-Western immigrants but also children from other Scandinavian countries. Despite noticing these patterns, it should be stressed that none of the observed effects are significant, neither in per-comparison testing nor in multiple testing, and we do not pursue the analysis of them further here, but it may be interesting to analyze them further in future research.

To analyze the potential importance of the stability of staff, we have divided the preschools into two subgroups: those with a high stability of the staff (an annual turnover rate below the median) and those with a low stability (an annual turnover rate above the median).²³ The results do not show any significant heterogeneity in the effects across these two subgroups of preschools. We looked more closely into the question of the stability of staff, since this is an issue of great interest in relation to the current intervention. If we instead of splitting the group of preschools at the median, select a threshold of 30% turnover²⁴ then the results (not reported here) indicate that the intervention is more effective in high-stability preschools where it leads to improvements for children attending these preschools. For children in preschools with a high turnover rate among the staff, the intervention fails to create improvements. This result is not surprising as the intervention is based

²³ The median of the turnover rate is 12.5%.

²⁴ The selection of the threshold of 30% turnover is based upon an investigation of different threshold values and resulting sizes of the subgroups. Since most preschools experience some turnover of their staff during a year (and one person may constitute 5-10% of the staff), very low threshold values have no explanatory power. We also estimated a version of the model with a continuous measure of turnover in the interaction term and that specification also indicates that the effects were stronger for more stable preschools.

on training some teachers in each preschool, and the newly trained teachers thereafter share their knowledge and ideas with the rest of the teachers. Together they adapt and implement the new knowledge in the preschool. A highly changing pool of teachers from year to year could make it difficult to absorb the new changes in the preschool and to make them a common foundation for the pedagogical work. However, since these subgroups were not defined *ex ante* and therefore can be considered a result of data mining, these results on stability should rather be considered as suggestions for a further evaluation focusing on these effects (as recommended by Duflo et al., 2007).

To summarize, the general picture emerging from the analysis of heterogeneous effects is that the intervention has not been able to produce larger improvements for socially disadvantaged children than for other children. There are even some indications, although not significant, that the intervention is actually less beneficial for socially disadvantaged children than for other children, contrary to the stated intentions.

8. Sensitivity Analyses

8.1 Alternative Specifications

This section reports results from various sensitivity analyses where we have employed a number of different specifications to investigate the robustness of our main results. Our main results come from a value-added model that includes a large set of explanatory variables to control for child, family, and preschool characteristics. To check the robustness of our estimates, we employ two alternative specifications: a restricted value-added model and a panel data model with individual-specific fixed effects (see also the description in Section 5).²⁵

²⁵ Detailed results from these alternative specifications are not reported here, but are available on request.

The restricted value-added specification models the changes in the outcome measures from pre-test to post-test as a function of the explanatory variables, i.e. it corresponds to fixing the parameter on the pre-test outcome measure to a value of one. Since the results reported in Section 6 show that the estimates of these parameters lie around 0.3, we would prefer the main effects from the unrestricted value-added model as they are better supported by the data. However, the main effects estimated from the restricted value-added model are qualitatively similar to those obtained from the unrestricted model, although they in most cases are stronger in the sense that their magnitude are larger and they are more significant.

As another alternative specification, we employ a panel data model with individual-specific fixed effects. This allows us to estimate the intervention effects controlling for all unobserved time-constant characteristics, but the presence of the fixed effects does not allow us to include time-constant explanatory variables.²⁶ In general, the main effects estimated by the panel data approach are very similar to those obtained from the restricted value-added model, which means that they are stronger than (larger magnitude and higher significance) but qualitatively similar to those obtained from the unrestricted value-added model. We can thus conclude that the main effects are fairly robust against changes in the specification of the model, and that the effects reported in Section 6 are probably rather conservative estimates of the intervention effects. When using the alternative specifications to investigate heterogeneous effects, the overall picture is confirmed (in particular when multiple testing is accounted for): the intervention has not been able to produce larger improvements for socially disadvantaged children than for other children.

Including the control variables also provides us with information about which background characteristics determine the socio-emotional skills of the children (as measured by SDQ). As

²⁶ The panel data model also includes time indicators which capture the autonomous improvement over time in SDQ scores.

several previous studies, we find that girls generally have better outcomes on all five SDQ dimensions and that parents' education is a strong predictor for child outcomes across all five SDQ dimensions, with better outcomes measured for children whose parents have more education. Ethnicity affects some of the outcomes, but not in a consistent way.

8.2 Midway Analysis

Another way to investigate the robustness of the effects is to perform a midway analysis of the effects of the intervention. Instead of analyzing the effects after the full period of the intervention, it is possible to analyze the effects half way through the intervention, i.e. after the first year of the intervention, since all outcomes have also been measured in March 2012. Such a midway analysis can be performed on two different samples, either on the same sample as above, i.e. the children enrolled for the full period of 2011-2013, or on an enlarged sample consisting of all children enrolled in 2011 and 2012. About half of the latter will have left the preschools in 2013, because they have become old enough to be admitted to grade 0 of primary school. Basing the analysis on this latter sample will thus give us more precise estimates due to the larger number of observations. We can also perform a separate analysis on the children that have joined the preschools in 2012 and still are enrolled in 2013.

Results on the first subperiod (2011-2012) do not show any significant effects of the intervention, despite more precise estimates. This holds for the five SDQ subscales, the total score, and the impact score. Results on the second subperiod (2012-2013) also do not show any significant effects of the intervention, again despite more precise estimates. One possible explanation of the results for the second subperiod is that the children included in this analysis are quite young children, since they have just entered the preschools between the start in 2011 and the midway measurement in 2012. If the intervention has stronger effects on older children than on younger children, this could

explain these results. However, the age aspect will not be investigated further here as it falls outside the scope of our analysis.

The conclusion of this sensitivity analysis is that the effects cannot be found in any of the two subperiods, but that the full period is required to obtain significant effects of the intervention. We are not surprised about this result as it probably takes some time for the new pedagogical routines to be implemented and to have a direct effect on the children.

9. Discussion and Conclusion

International studies have shown that early interventions can have positive effects in the sense that (especially) socially disadvantaged children are more likely to prosper, in both the short and the long run, if they receive social and intellectual stimulation from an early age. However, the studies do not tell us about benefits for children in a system of universal daycare, where all children participate regardless of their socio-economic background, as is the case in Denmark with an attendance rate of 97% among children aged 3-5.

We evaluate a professional development intervention (the VIDA intervention) to investigate whether education, training and knowledge sharing of preschool teachers (and principals) affect children's socio-emotional outcomes. The intervention was evaluated using a cluster-RCT design with random assignment at the preschool level. The intervention was implemented in 29 preschools (and compared with 29 control preschools) in three Danish municipalities.

The intervention distinguishes itself from other interventions found in existing international studies in two ways. Firstly, it focuses on professional development of preschool teachers to implement an intervention aimed at enhancing children's socio-emotional development via an inclusive pedagogy, i.e. it is founded theoretically in the bioecological and contextual perspective. Secondly, the

intervention is designed for implementation in universal public daycare where almost all children attend. The preschool teachers are trained to implement a specific learning and developmental program and thereby to change the entire system (interactions, inclusion, and teacher qualification). The professional development focused on teaching the participants to develop activities and renew practices that could improve the individual child's development through responsive supportive interaction and through inclusive learning environments. The preschool teachers trained this approach through ongoing educational and training sessions where they learned to implement the program.

Our focus is on socio-emotional competences, i.e. child behavior and well-being, as measured by the five subscales of the SDQ scale (Goodman, 1997). Effects are analyzed by comparing baseline measures of SDQ to measures collected at the end of the intervention. The results show that the professional development of preschool teachers provided by the intervention has positive effects on development and behavior among children aged 3-5. The intervention significantly improves children's emotional development and causes a reduction in emotional and behavioral problems with an effect size around 0.25. However, the results also show that the intervention is not successful in producing larger improvements for socially disadvantaged children than for other children. Finally, the results indicate that a high turnover rate among the staff of the preschools may act as a barrier to a successful outcome of the intervention, but that aspect requires further evaluation.

Comparing the findings of this study with existing international studies, it is noticeable that we do not find stronger effects on socially disadvantaged children than on their more privileged peers. Previous studies (e.g. of the effects of the Perry Pre-School Project) have often found very strong effects of interventions on socially disadvantaged children, but these interventions have typically been targeted towards this group and the control group has consisted of children who have not

attended any daycare. In our case, at least two aspects are probably explaining why we obtain a different result. Firstly, the intervention has been implemented in a daycare system where socially disadvantaged children are already attending preschools, i.e. they are also in the control group and thereby participating in activities rooted in the Act on Educational Curricula. Secondly, the intervention is based on professional development of the preschool teachers and as such it is much less targeted towards socially disadvantaged children. Basing an intervention on professional development of the preschool teachers may be effective in many cases, but it contains risky components since a high turnover among the preschool teachers may erode the basis of the intervention. This is consistent with other recent research on Danish register data indicating that a high structural quality of preschools (measured e.g. by staff-child ratios and educational level of the staff) can have positive long-term effects (see Bauchmüller et al., 2014).

A possible limitation of this study is that some of the participating municipalities have had other interventions at the same time as the VIDA intervention. One municipality has implemented a so-called ‘vision project’ for all preschools in the municipality. This should affect intervention and control preschools in the same way, but potentially leads to a (downward) bias in the effects of the current intervention. Another municipality initiated an ‘inclusion project’ in 2013 by offering post-education training of preschool teachers. We do not believe that this project interferes with the results from the current intervention as the project started almost at the same time as our final data collection. To control for potential municipality differences, all reported estimations included municipality fixed effects. Another limitation of the current study is the focus on children’s socio-emotional outcomes as measured by SDQ scores. SDQ is based on teacher ratings which potentially are subjective, although we have demonstrated that our results most likely are not influenced by such bias. However, in terms of generalizability of the results, it would also be nice to have objective data on child outcomes from cognitive tests.

In summary, the current study has contributed to fill some of the knowledge gap about how professional development of preschool teachers may affect children's socio-emotional outcomes. We have provided evidence showing that it is possible to improve children's outcomes through an intervention focused on preschool teachers, but that it is not effective as a special way to target socially disadvantaged children in the current context. The evidence also points to some challenges in implementing organizational changes and learning approaches to professional development, as it requires some stability of the staff. Future research needs to more carefully consider the importance of combining the professional development with additional measures for increasing the stability of the staff.

10. References

- Ainsworth, H., Hewitt, C., Torgerson, C., Higgins, S., Wiggins, A. & Torgerson, D. (2015). Sources of bias in outcome assessment in randomised controlled trials: a case study. *Educational Research and Evaluation* 21(1), 3-14.
- Almond, Douglas, and Janet Currie (2011). Human capital development before age five. In *Handbook of Labor Economics, Volume 4b, Ch. 15*, ed. O. Ashenfelter and D. Card, 1315-1486, Amsterdam: Elsevier.
- Anders, Yvonne, Hannah Ulferts, Paul Leseman and Edward Melhuish (2016). *Effects of ECEC on academic outcomes in literacy and mathematics: Meta-analysis of European longitudinal studies*. CARE project; Curriculum Quality Analysis and Impact Review of European Early Childhood Education and Care (ECEC). Available at <http://ecec-care.org/resources/publications/>
- Avvisati, Francesco, Marc Gurgand, Nina Guyon and Eric Maurin (2014). Getting Parents Involved: A Field Experiment in Deprived Schools. *Review of Economic Studies* 81, 57-83.
- Bauchmüller, Robert, Mette Gørtz, and Astrid Würtz Rasmussen (2014). Long-Run Benefits from Universal High-Quality Preschooling. *Early Childhood Research Quarterly* 29, 457-470.
- Blau, David, and Janet Currie (2006). Pre-school, day care and after-school care: Who's minding the kids? In *Handbook of the Economics of Education, Volume 2, Ch. 20*, ed. E. Hanushek and F. Welch, 1163-1278, Amsterdam: Elsevier.
- Bronfenbrenner, Urie and Pamela A. Morris (2006). The bioecological model of human development. In *Handbook of child psychology: Theoretical models of human development, Volume. 1*, ed. W. Damon and R. M. Lerner, 793-828, New York: Wiley.

Brown, John S., and Paul Duguid (1991). Organizational Learning and Communities-of-Practice: Towards a unified view of working, learning and innovation. *Organization Science* 2, 40-57.

Burger, Kaspar (2010). How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds. *Early Childhood Research Quarterly* 25, 140-165.

Cameron, Claire E. (2012). A transactional Model of Effective Teaching and Learning in the Early Childhood Classroom. In *Handbook of Early Childhood Education*. ed. R. C. Pianta, 278-296, New York: Guildford Press.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane W. Schanzenbach, and Danny Yagan (2011). How does your kindergarten classroom affect your earnings? Evidence from project STAR. *The Quarterly Journal of Economics* 126, 1593-1660.

Claessens, Amy, Mimi Engel, and F. Chris Curran (2014). Academic content, student learning, and the persistence of preschool effects. *American Educational Research Journal* 51, 403-434.

Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov (2006). Interpreting the Evidence on Life Cycle Skill Formation. In *Handbook of the Economics of Education, Volume 1, Ch. 12*, ed. E. Hanushek and F. Welsh, 697-812, Amsterdam: Elsevier.

Datta Gupta, Nabanita, and Marianne Simonsen (2010). Non-cognitive child outcomes and universal high quality child care. *Journal of Public Economics* 94, 30-43.

Demarest, S., Van der Heyden, J., Charafeddine, R., Tafforeau, J., Van Oyen, H., & Van Hal, G. (2013). Socio-economic differences in participation of households in a Belgian national health survey. *The European Journal of Public Health* 23(6), 981-985.

- Duflo, Esther, Rachel Glennerster and Michael Kremer (2007). Using randomization in development economics research: A toolkit. In *Handbook of Development Economics, Volume 4, Ch. 61*, ed. T. P. Schultz and J. A. Strauss, 3895-3962, Amsterdam: Elsevier.
- Duncan, G. J., & Magnuson, K. (2013). Investing in Preschool Programs. *Journal of Economic Perspectives*, 27(2), 109-132.
- Duncan, Greg J., and Aaron J. Sojourner (2013). Can intensive early childhood intervention programs eliminate income-based cognitive and achievement gaps? *The Journal of Human Resources* 48, 945-968.
- Ellström, Per-Erik (2010). Practice-based innovation: a learning perspective. *Journal of Workplace Learning* 22, 27-40.
- Farran, D.C. and Hofer, K.G. (2013). *Evaluating the quality of early childhood education programs*, in Saracho, O. N., & Spodek, B. (Eds). *Handbook of Research on the Education of Young Children*. 3rd Edition: Routledge.
- Ford T, Hutchings J, Bywater T, Goodman A, Goodman R (2009). Strengths and Difficulties Questionnaire Added Value Scores: evaluating effectiveness in child mental health interventions. *British Journal of Psychiatry* 194, 552-558.
- Gheradi, Silvia (2011). Organizational Learning: The Sociology of Practice. In *The Blackwell Handbook of Organizational Learning and Knowledge Management*, ed. M. Easterby-Smith and M. A. Lyles, 43-65, Oxford: Blackwell Publishing.
- Goodman, Robert (1997). The strengths and difficulties questionnaire: a research note. *Journal of Child Psychology and Psychiatry* 38, 581–586.

Goodman, Robert (1999). The extended version of the strengths and difficulties questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry* 40, 791–799.

Goodman, Anna, and Goodman, Robert (2009). Strengths and difficulties questionnaire as a dimensional measure of child mental health. *Journal of the American Academy of Child and Adolescent Psychiatry* 48, 400-403.

Hamre, Bridget, Jason T. Downer, Faiza M. Jamil, and Robert C. Pianta (2012). Enhancing Teachers' Intentional Use of Effective Interactions with Children. In *Handbook of Early Childhood Education*. ed. R. C. Pianta, 507- 532, New York: Guildford Press.

Hattie, John A. C. (2008). *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Heckman, James J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science* 312, 1900-1902.

Heckman, James J. (2008). Schools, Skills and Synapses. *Economic Inquiry* 46, 289-324.

Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics* 94, 114-128.

Heckman, James J., and Tim Kautz (2012). Hard evidence on soft skills. *Labour Economics* 19(4), 451-464.

Imbens, Guido, and Jeffrey Wooldridge (2007). What's new in econometrics? Lecture notes (lecture 10), NBER Summer Institute (available at www.nber.org).

Jensen, Bente, Anders Holm, Camilla Wang, Dorte Kousholt, Ib Ravn, Michael S. Larsen, Ove S. Rasmussen, Peter Berliner, Thomas Y. Andersen, and Ulrik Brandi (2011). Vidensbaseret indsats over for udsatte børn i dagtilbud – modelprogram, Statusrapport 1. VIDA-forskningsserien 2011:1, Aarhus Universitet.

Jensen, Bente, Anders Holm, and Sven Bremberg (2013). Effectiveness of a Danish early year preschool program: A randomized trial. *International Journal of Educational Research* 62, 115-128.

Jensen, P. and Rasmussen A.W. (2016). *Professional Development and its Impact on Children in Early Childhood Education and Care: A Meta-Analysis based on European Studies*. CARE project; Curriculum and Quality Analysis and Impact Review of European Early Childhood Education and Care (ECEC). Available at <http://ecec-care.org/resources/publications/>

Jensen, Morten B., Björn Johnson, Edward Lorenz, and Bengt Å. Lundvall (2007). Forms of knowledge and modes of innovation. *Research Policy* 36, 680-693.

Kling, Jeffrey, and Jeffrey Liebman (2004). *Experimental Analysis of Neighborhood effects on Youth*. Working paper No. RWP04-034, KSG, Harvard University.

Lipsey, Mark W., Kelly Puzio, Cathy Yun, Michael A. Hebert, Kasia Steinka-Fry, Mikel W. Cole, Megan Roberts, Karen S. Anthony, and Matthew D. Busick (2012). Translating the statistical representation of the effects of education interventions into more readily interpretable forms. NCSER 2013-3000, U.S. Department of Education.

Mashburn, Andrew J., Robert C. Pianta, Bridget K. Hamre, Jason T. Downer. Oscar A. Barbarin, Donna Bryant, Margaret Burchinal, Diane M. Early, and Carollee Howes (2008). Measures of

classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development* 79, 732-749.

Mathai, J., Anderson, P. and Bourne, A. (2003). Use of the Strengths and Difficulties Questionnaire as an outcome measure in a child and adolescent mental health service. *Australasian Psychiatry* 11, 334-337.

Melhuish, E., Ereky-Stevens, K., Petrogiannis, K., Ariescu, A., Penderi, E., Rentzou, K., Tawell, A., Slot, P., Broekhuizen, M., & Leseman, P. (2015). *A review of research on the effects of Early Childhood Education and Care (ECEC) upon child development*. CARE project; Curriculum Quality Analysis and Impact Review of European Early Childhood Education and Care (ECEC). Available at <http://ecec-care.org/resources/publications/>

Niclasen J, Teasdale TW, Andersen A-MN, Skovgaard AM, Elberling H, Obel C (2012). Psychometric Properties of the Danish Strength and Difficulties Questionnaire: The SDQ Assessed for More than 70,000 Raters in Four Different Cohorts. *PLoS ONE* 7(2), 1-8.

Nordhov, Solveig Marianne; Rønning, John Andreas; Ulvund, Stein Erik; Dahl, Lauritz Bredrup; Kaaresen, Per Ivar (2012). Early Intervention Improves Behavioral Outcomes for Preterm Infants: Randomized Controlled Trial. *Pediatrics* 129 (1), E9 - E16.

Pianta, Robert C., W. Steven Barnett, Margaret Burchinal, and Kathy R. Thornburg (2009). The effects of preschool education: what we know, how public policy is or is not aligned with the evidence base, and what we need to know. *Psychological Science in the Public Interest* 10, 49-88.

Powell, D. R., and Diamond, K. E. (2010). Improving the Outcomes of Coaching-Based Professional Development Interventions, in Neuman, S. B., and Dickinson, D. K. (Eds). *Handbook of Early Literacy Research. Volume 3*: Guilford Publications.

- Raver, C. Cybele, Stephanie M. Jones, Christine Li-Grining, Fuhua Zhai, Molly W. Metzger, and Bonnie Solomon (2009). Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial. *Journal of Consulting and Clinical Psychology* 77, 302-316.
- Saretsky, G. (1972). The OEO PC experiment and the John Henry effect. *Phi Delta Kappan* 53, 579-581.
- Schachter, R. E. (2015). An Analytic Study of the Professional Development Research in Early Childhood Education. *Early Education and Development* 26, 1057-1085.
- Shadish, W. R., Cook, T. D., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalised causal inference*. Boston, MA: Houghton Mifflin.
- Statistics Denmark (2015). www.statistikbanken.dk. Tables PAS11 and BEF5 (accessed February 3, 2015).
- Stewart-Brown S, Patterson J, Mockford C, Barlow J, Klimes I, Pyper C. (2004). Impact of a general practice based group parenting programme: Quantitative and qualitative results from a controlled trial at 12 months. *Archives of Disease in Childhood* 89, 519–525.
- Stone, L .L., Otten, R., Engels, R. C. M. E., Vermulst, A. A. & Janssens, J. M. A. M. (2010). Psychometric Properties of the Parent and Teacher Versions of the Strengths and Difficulties Questionnaire for 4-to 12-Year-Olds: A Review. *Clinical Child and Family Psychology Review* 13(3), 254-274.
- Sundbo, Jon (2003). Innovation and Strategic Reflexivity: An Evolutionary Approach Applied to Services. In *The International Handbook of Innovation*, ed. L. V. Shavinina, 97-114. New York: Pergamon Press.

Todd, Petra E., and Kenneth I. Wolpin (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human Capital* 1, 91-136.

Zaslow, M., Tout, K., Halle, T., Whittaker, J. V., & Lavelle, B. (2010). *Toward the identification of features of effective professional development for early childhood educators. Literature review.*

U.S. Department of Education, Washington, D.C.

A. Appendix

Table A.1. Comparison of attritors and non-attritors for the total sample and separately for control and intervention groups

	Total sample p-values	Control group p-values	Intervention group p-values
Outcome measures			
Emotional symptoms	.77	.29	.62
Conduct problems	.05*	.10	.27
Hyperactivity	.26	.45	.31
Peer relationship prob.	.28	.93	.08*
Prosocial behavior	.09*	.84	.01***
Total SDQ score	.14	.26	.28
SDQ impact score	.02**	.39	.02**
Child and family characteristics			
Boy (0/1)	.69	.45	.93
Danish origin (0/1)	.17	.39	.26
12 years of education, mother (0/1)	.22	.75	.01**
12 years of education, father (0/1)	.60	.26	.22
Income less than 10% of income distribution (0/1)	.94	.54	.14
Parents married (0/1)	.03**	.43	.05*
Living with single parent (0/1)	.11	.56	.03**
Preschool characteristics			
Size of preschool (no. of children)	.53	.55	.51
Staff-child ratio	.21	.33	.74
Share of educated staff	.89	.86	.56
Stability of staff (1- annual turnover)	.44	.18	.45
Change of principal of preschool (0/1)	.15	.13	.32
Stable preschool (turnover rate below median) (0/1)	.64	.37	.76

Notes: The tests are for equality of means in the group of attritors and the group of non-attritors. The p-values are from regressions of row variables on an attrition indicator variable with standard errors adjusted for clustering at the preschool level (as in Avvisati et al., 2014).). * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table A.2. Summary statistics and balancing tests for the normalized SDQ measures in the initial sample (pre-test measures)

SDQ subscale	Obs.	Balancing		
		Pre-test	tests	
		Mean	SD	p-values
Emotional symptoms				
Control	403	0.81	1.00	
Intervention	472	0.83	1.04	0.81
Conduct problems				
Control	403	0.63	1.00	
Intervention	472	0.65	1.00	0.83
Hyperactivity				
Control	403	0.94	1.00	
Intervention	472	0.99	1.03	0.63
Peer relationship prob.				
Control	403	0.72	1.00	
Intervention	472	0.79	1.13	0.52
Prosocial behavior				
Control	403	2.58	1.00	
Intervention	472	2.61	1.01	0.81
Total SDQ score				
Control	403	1.11	1.00	
Intervention	472	1.16	1.09	0.63
SDQ impact score				
Control	391	0.39	1.00	
Intervention	461	0.44	1.19	0.60

Notes: All SDQ scores have been normalized by their standard deviation in the pre-test control group. Originally, all five subscales of SDQ are measured on a 0 to 10 scale, the total SDQ score is the sum of the first four subscales (hence, its original scale is from 0 to 40), and the SDQ impact score is measured on a 0 to 6 scale. Original scores are available from the authors. The balancing tests are for equality of means in the control and intervention group. The p-values are from regressions of outcome measures on an intervention indicator variable with standard errors adjusted for clustering at the preschool level (as in Avvisati et al., 2014).

Table A.3. Summary statistics (by group) and balancing tests for child and family characteristics in the initial sample

Variable	Control			Intervention			Balancing tests
	Obs.	Mean	SD	Obs.	Mean	SD	p-values
Boy (0/1)	403	0.54	0.50	472	0.50	0.50	.16
Danish origin (0/1)	399	0.86	0.35	464	0.87	0.34	.92
12 years of education, mother (0/1)	385	0.80	0.40	446	0.80	0.40	.95
12 years of education, father (0/1)	384	0.74	0.44	442	0.81	0.39	.11
Income less than 10% of income distribution (0/1)	365	0.10	0.30	432	0.09	0.29	.77
Parents married (0/1)	399	0.58	0.49	464	0.66	0.47	.10*
Living with single parent (0/1)	399	0.21	0.41	464	0.14	0.35	.10*

Notes: Background information for children is collected at the end of the year the child turns 3. The balancing tests are for equality of means in the control and intervention group. The p-values are from regressions of row variables on an intervention indicator variable with standard errors adjusted for clustering at the preschool level (as in Avvisati et al., 2014). * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table A.4. Summary statistics (by group) and balancing tests for preschool characteristics in the initial sample

Variable	Control			Intervention			Balancing tests
	Obs.	Mean	SD	Obs.	Mean	SD	p-values
Size of preschool (no. of children)	382	77.10	29.90	471	71.23	24.52	.50
Staff-child ratio	375	0.19	0.05	471	0.21	0.04	.22
Share of educated staff	375	0.57	0.11	471	0.61	0.11	.21
Stability of staff (1- annual turnover)	375	0.88	0.15	471	0.86	0.14	.63
Change of principal of preschool							
(0/1)	375	0.24	0.43	471	0.04	0.19	.06*
Stable preschool (turnover rate below							
median) (0/1)	375	0.52	0.50	471	0.46	0.50	.72

Notes: Background information for preschools is collected at the start of the intervention period. The balancing tests are for equality of means in the control and intervention group. The p-values are from regressions of row variables on an intervention indicator variable with standard errors adjusted for clustering at the preschool level (as in Avvisati et al., 2014). * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.