The Effects of a Structured Curriculum on Preschool Effectiveness: A Field Experiment

Online Appendix

Mari Rege
Ingunn Størksen
Ingeborg F. Solli
Ariel Kalil
Megan M. McClelland
Dieuwer ten Braak
Ragnhild Lenes
Svanaug Lunde
Svanhild Breive
Martin Carlsen
Ingvald Erfjord
Per Sigurd Hundeland

A1. Intervention

Preschool curriculum

Scientifically based: The preschool curriculum consists of age-appropriate skill-building activities in mathematics, language, executive functioning, and social skills. The pedagogical principles for the curriculum build on research related to teacher-child relationships (

Pianta 1999) and to playful learning (Weisberg, Hirsh-Pasek, and Golinkoff 2013). For young children, learning takes place during play, both during free play and guided play. During guided play, teachers intentionally prepare and introduce books, games, activities or toys, and engage children in exploring specific content themes. Additionally, the activities were designed with inquiry as a pedagogical principle that emphasizes children's own explorations (Wells 1999).

The activities were inspired by existing curricula with promise for positive effects. A multi-disciplinary team of researchers (including researchers with many years of experience working as preschool teachers in Norway) studied existing curricula in detail. We selected curricula based on two criteria: a) Potential gain for improvement in executive functioning, language and mathematics development, and b) Promise for implementation in the Norwegian culture and preschool context, i.e. a playful learning approach (Weisberg, Hirsh-Pasek, and Golinkoff 2013). This narrowed down the list of curricula and pedagogical approaches we used as inspiration to the following: I Can Problem Solve (Shure 1992), Interactive Book Reading (Mol, Bus, and de Jong 2009), Building Blocks (Clements and Sarama 2011), California Preschool Curriculum Framework (California Department of Education 2016), Tools of the Mind (Bodrova and Leong 2007), and Red Light, Purple Light (Schmitt et al. 2015).

Based on the selected curricula, we drafted outlines for more than 130 learning activities, stimulating the targeted skill domains. The main objective was to provide material and inspiration for teachers to create a more intentional and structured practice. For each activity, we organized the text according to the following headings: Intention, Preparation, Implementation, and Materials needed. Our goal was that the teachers would use the activities thoughtfully and intentionally, and not just as fun games. Still, all activities were designed to be engaging and meaningful for the children, stimulating them to be active and collaborative, in accordance with theory of playful learning.

Embedded in Norwegian culture and context: As part of the teacher training, the teachers were asked to try out the drafted activities with the current five-year-olds in their own preschool centers and to give the researchers feedback and suggestions. Notably, teachers were instructed that this piloting should not be done with the treatment children. Throughout the year, each teacher had to provide us with oral and written feedback on multiple activities. At the end of the school year, we had critical and constructive feedback from many teachers on each of the activities. They also suggested alternative activities that we integrated into the curriculum. This resulted in a substantial revision embedding the playful learning activities to Norwegian culture and context.

Published as a Book: The curriculum was published as a book called "Lekbasert læring" (playful learning) in Norwegian (Størksen et al. 2018). The book is written in Norwegian to fit with the Norwegian preschool context. In addition to 130 learning activities, the book contains an introduction to the theoretical and empirical evidence on which the curriculum builds. In particular, it emphasizes curricula that intentionally and systematically target skills in language, mathematics, executive functioning, and social skills, as key ingredients in

quality preschool. Moreover, it gives an introduction to the importance of positive and stimulating relationships between teachers and children, and the playful learning approach.

The book includes the following:

- Brief introduction to evidence base (importance of child-teacher relationship, the playful learning approach, and important skills to stimulate in early childhood)
- Practical implementation guidelines, including templates to plan each day, week, and month.
- 130 playful learning activities stimulating skills in mathematics, language, executive functioning, and social skills.
- Fall and spring curricula referring to relevant pages for all 130 activities in the book
- Templates for activities

Importantly, the curriculum was not a detailed program intended to dictate teacher practice on an everyday basis. We provided a suggestion for activities suited for each month (fall and spring plan). However, teachers were encouraged to develop their own unique approach to the curriculum and to augment it with their own ideas. The activities were flexible in terms of challenge and complexity, allowing teachers to match their practice to children's skill levels. Teachers were encouraged to enhance children's learning opportunities by continuously giving them new challenges. The treated preschool centers signed a contract to spend at least eight hours a week engaging the five-year-olds in the curriculum and split the time between activities stimulating skills in mathematics, language, executive functioning, and social skills.

In the following, we will give detailed examples of specific learning activities in the curriculum.

Mathematics activities: The curriculum covers activities stimulating number and quantitative thinking, in addition to measurement, geometry, and statistics. For example, within numeracy, children are engaged in the game *Marve Larve* (*Marve the Caterpillar*). In this activity, children make their own Caterpillar with beads and pearls on a string. They also make a "house" out of a matchbox labeled with the same numeral as well as the same number of dots as the number of beads. Furthermore, children compare the lengths of all the children's caterpillars (smaller, longer, smallest, etc.). Then, the teacher mixes all the children's caterpillars and houses, and the children are asked to match them into the right houses. The teacher plays a game and lets the Caterpillar partly pop out of the house, and children and adults count the number of beads outside the house and try to figure out how many beads are still inside the house. This is an example of how a fun game in the book corresponds with quite advanced mathematics such as equations (e.g. $3 + x = 6$) introduced later in school. Another example is the *Geometric photo safari.* In this game, the group discusses various geometrical shapes and identify such shapes in the environment outside the preschool center. The children take photos of geometrical shapes and then discuss the results during circle time at the end of the session.

Language activities: Important for the language activities, is the theory section of the book that describes how language can be divided into three separate but overlapping components: content, form, and use. In this way, teachers are conscious of different developmental areas within language. Furthermore, this section reviews the principles of interactive book reading (Mol, Bus, and de Jong 2009), including teacher preparation, pre-reading with children to create engagement for the book, reading sessions, focus words, retelling and dialogue, and finally, book-related activities such as drawing and drama. In these drama activities, children are given props related to the story that has been read so that they can dramatize the content and continue to practice new words and concepts. Language activities for children are related

to either interactive book reading or other kinds of language games. For example, in a language game called *I am a letter!,* children are given one letter each on a piece of paper, and then the teacher challenges them to form words with these letters by moving around so that the letters come in the correct position. In another language game, each child is given a short word, and then the group is challenged to move around so that they form a sentence.

Activities related to executive functioning: The theoretical part of the book explains how self-regulation relies on underlying executive functioning processes (including attentional or cognitive flexibility, working memory, and inhibitory control) (Blair and Raver 2015). In accordance with this, the activities involve children's ability to use attentional flexibility, working memory, and inhibitory control. For example, in the activity *The bear is sleeping, the snail is sleeping* teachers rewrite a traditional Norwegian nursery rhyme so that not only does the bear fall asleep and wake up, other animals fall asleep and wake up again too. The children pay attention and listen carefully to the song and to what kind of animals appear as the teacher sings. During the song, they dramatize the sleeping and awakening of these different animals with different speeds and movements. In this game, children use attention, working memory, and inhibition to be able to follow the song and mime the animals and their movements. In another activity called *The ready, steady, go game* children are to run on "ready, steady, go!" and to inhibit their impulse to run on other instructions such as "ready, steady, gorilla!".

Activities related to social competence: The book explains how social competence relies on skills such as self-control, assertiveness, responsibility, co-operation, and empathy (emotional competence), and games are included that stimulate these skills. For example, in the activity called *Mailing a hug*, children think of a relative or a friend that might enjoy encouragement and write or draw a message that can encourage the receiver. Through this activity, the children imagine the experiences and emotions of another person. Likewise, children identify and express their own emotions through drawing and drama activities. In an activity called *The gingerbread man* children express their emotions through colors within the outlines of a gingerbread man and talk about these emotions and how they feel inside.

Teacher Training

The teacher training consisted of a credit-based university class prior to the year of curriculum implementation and coaching during the year of implementation. The class provided the preschool teachers with key insights from the theoretical and empirical research literature on which the curriculum builds. Importantly, the class was practice-oriented. The class consisted of four two-day lecture gatherings over a period of eight months. Between class gatherings, teachers practiced playful learning activities with the current five-year-olds in their preschool center (not the children in our study), reported on feasibility, and reflected on how their experiences aligned with the theoretical and empirical literature covered in class. They also suggested new games and activities for the curriculum.

Since our baseline assessments with teachers told us that they spent much less time on playful learning within mathematics compared to other skill domains, we chose to give more attention to mathematics in the teacher training. More precisely almost 40 percent of teaching hours were spent on mathematics.

Teachers were spread across a large region in the southern part of Norway called Agder. In order to make the class feasible for all teachers, we arranged all lecture gatherings twice, once in Eastern Agder and once in Western Agder. Fulfillment of class participation, practice and assignments gave 15 credit points in the Norwegian university system, in which full-time students are supposed to complete 60 credit points a year. All the preschool teachers passed

the class. Absence from sessions or classes was low, with less than 10 full-day absences due to health issues across all teachers.

In addition to the credit base university class, teachers were coached during the intervention year in two gatherings and phone meetings. In the two gatherings (September 2016 and March 2017), teachers reviewed insights from research on systematic curricular focus and were challenged to reflect in groups on how their curriculum implementation matched with the intentions in the project. They were also asked to reflect on challenges and successes. Thereafter, the teachers shared ideas and experiences with the entire class, and the instructors participated in the discussion by listening to the preschool teachers' concerns and guiding them on how to address these concerns.

Additionally, the teachers had scheduled phone meetings with their coach two times each semester and could schedule additional phone meetings to discuss any immediate questions or concerns. One of the scheduled phone meetings per semester was dedicated to mathematics and the other to language, executive functioning, social skills, and pedagogical approaches. Again, mathematics was emphasized, since we knew from our baseline assessments with teachers that mathematics was less emphasized in preschool centers. The scheduled phone meetings were conducted as semi-structured interviews. For example, in the first meeting, we asked: 1. a very broad opening question allowing them to come up with whatever they felt relevant; 2. whether their center administration gave practical support and facilitated for the group; 3. their experiences with guiding the assistants; 4. whether they believed they succeeded in building a trusting relationship with children; 5. whether activities met criteria for playful learning, and finally; 6. experiences from daily activities. Throughout the conversations, we listened to their concerns and guided them on how to address these concerns. All treated teachers participated in the scheduled phone meetings.

Typically, each preschool teacher had one or two assistants when implementing the curriculum, depending on the size of the child group. The trained preschool teacher had the main responsibility to train the assistants. However, assistants also received a one-day training in the intervention material and research project in general (1/2 day), and more specifically on interactive book reading (1/2 day). This is because in groups with more than six five-year-olds, which were most groups, we recommended that the children were divided into two groups which alternated between the language and mathematics activities, with the assistant in charge of the language activities.

A2: Assessment, Measures, and Control Variables

Assessment:

We assessed the children in August 2016 (baseline, T1), June 2017 (post-intervention, T2), and March 2018 (follow-up, T3). The T1, T2, and T3 assessments used the same test battery, which took approximately 40 minutes for each child. All assessments were one-to-one with a trained and certified tester, blind to treatment status. All testers had to hand in a police certificate stating that they had no record of offenses that would make them unsuitable for working with children. The testers used computer tablet instruments with a validated test battery developed for transition between preschool and school.

Tester training consisted of one full day of theory and practice related to our computer tablet test battery. Testers were then instructed to visit pilot preschool centers (T1 and T2) or schools (T3) to practice the test on the computer tablets with children in the relevant age group. A week later, the testers came back to discuss their experiences and to take the certification. The certification involved conducting all tests in the battery while one of the

researchers made systematic notes according to a certification form. For minor mistakes, testers got reminders and feedback and subsequently received their certification.

The T1 and T2 assessments were conducted at three central locations, including a large Science museum. For the other two locations, we hired personnel from the science museum to come and arrange an activity day for the children. All children in the participating preschool centers in the Agder project were invited to these activity days with assessments, and all preschool centers participated. Children were allowed to use the facilities and activities provided by the science museum for the full day. At a scheduled time the preschool centers met for assessment. Children's names were replaced with personal codes on stickers attached to their clothes, and the children were then guided to individual test stations for assessment. This way each preschool center was exposed to many different testers, which made it possible to detangle center effects from tester effects in analyses. In T3 (spring of 1st grade), testers traveled to the schools to conduct the assessment. We collaborated with the school administration who facilitated by guiding the participating children out of the classroom for the assessment. All children received a small gift for their participation (e.g. a ruler, a gym bag, or a pencil case with illustrations from the computer tablet assessment printed on them).

Measures:

We assessed skills in mathematics, language, working memory, and inhibitory control. The latter two are important components of executive functioning. Unfortunately, we weren't able to reliably measure social competence in this study due to a lack of tests validated in a Norwegian context.

Mathematics skills were assessed with the *Ani Banani Math Test* (ABMT; ten Braak and Størksen 2021). The ABMT is a playful mathematics test on a tablet application, which includes items covering three areas of mathematics – numeracy, geometry, and problem-solving. Children help a monkey with different tasks, such as counting bananas and setting the table with enough plates for birthday party guests. All correct answers were given one point. Due to technical problems with the tablet application at T3, data for 5 out of the 18 items of the ABMT was not recorded correctly and therefore omitted in the analyses for all assessment periods T1 to T3. All correct answers were given one point, and the total score was calculated as the sum across the 13 items. This short version correlates strongly (r = .58) with the Preschool Early Numeracy Skills test (PENS) in kindergarten and significantly predicts mathematic achievement in 1st (r = .529) and 5th grade (r = .553). Internal consistency was considered adequate (Cronbach's alpha = .60).

Two tests were conducted to measure language; one pertaining to vocabulary and the other to phonological awareness. Vocabulary was assessed with the *Norwegian Vocabulary Test* (NVT; Størksen et al. 2013). The NVT is a typical expressive vocabulary task including 20 words. Illustrations appeared on a tablet screen and the child was subsequently asked to name them. Cronbach's alpha was high; α = .81. Children's phonological awareness was assessed with a 12-item *blending task* that is part of the official literacy screening battery from The Norwegian Directorate for Education and Training. For each task, a target word was presented in its individual phonemes by the experimenter and children had to indicate the corresponding alternative from four presented images on a tablet screen. All correct answers were given one point. For both tests, the total score was calculated as the sum across all items.

Three tests were conducted to measure executive functioning. The *Head-Toes-Knees-Shoulders task* (HTKS; McClelland et al. 2014) integrates attention, inhibitory control, body control, and working memory demands into a short task of behavioral self-regulation

appropriate for children aged 4 to 8 years. It has strong reliability and validity and is significantly related to other measures of self-regulation and children's academic outcomes. These results have been replicated in many recent studies across the world. The task includes three blocks with 10 items each. Responses were scored with two points when correct, one point when the child made an incorrect movement but ended up with the correct response, and zero points for incorrect responses. Cronbach's alpha was sufficient; $\alpha = .76$. In the *Hearts and Flowers task* (Davidson et al. 2006), children had to press a key on the same side of the stimulus when they saw a heart and on the opposite side when the stimulus was a flower. The task has 57 items and number of correct responses were counted. The measure is designed to assess inhibitory control and cognitive flexibility skills and has been widely used with young children. Cronbach's alpha was $\alpha = .89$. For both tests, the total score was calculated as the sum across all items. The third test of executive functioning was the *Forward/Backward Digit Span* subtest from the Wechsler Intelligence Scales for children-III (Wechsler, 1991), which measures working memory. Digits were read aloud, one digit per second, and the children were asked to repeat the sequence of digits. First, they had to repeat digit sequences in the same order as they heard them, and then in reversed order. The number of digits in each sequence increased as the test continued. The test was automatically discontinued after two subsequent errors. The total score reflects the number of repeated sequences forward plus the number of repeated sequences backwards.

From these six tests conducted at each assessment, we created three outcome measures:

- Math: Percent correct answers at the ABMT test.
- Executive Functioning: Mean score of the standardized "HTKS", "Hearts and Flowers" and "Digit Span" tests. If missing on one of the tests, the other tests constitute Executive Functioning.
- Language: Mean score of the standardized "Phonological Awareness" and NVT tests. If missing on one of the tests, the other test constitutes Language.

All three outcome measures are standardized within each period to mean 0 and standard deviation 1.

Control variables:

From registry data we constructed the following control variables that entered into our analyses:

- Gender: Indicator for female
- Birth month: Continuous variable running from 1 (December born) to 12 (January born)
- Mother's and father's education: Continuous variable for number of years education, running from 10 (compulsory schooling) to 18 (Master's degree).
- Mother's and father's earnings: Income from work (employment and self-employment). In categories of 50.000 NOK, but re-coded as a continuous variable for the analyses.
- Immigrant status: Indicator for whether one or both parents are immigrants from a non-western country.

Data on parental characteristics was not available for 3 percent of the children, likely because these children/families were recent immigrants to Norway at the time, and still not recorded in the Norwegian administrative registers. Missing values were replaced by 0, and indicators for missing were included in the analyses. Finally, we also constructed indicators for late consent

(parental consent received after the preschool center's treatment status was known); for missing all T1 test scores; and for randomization block.

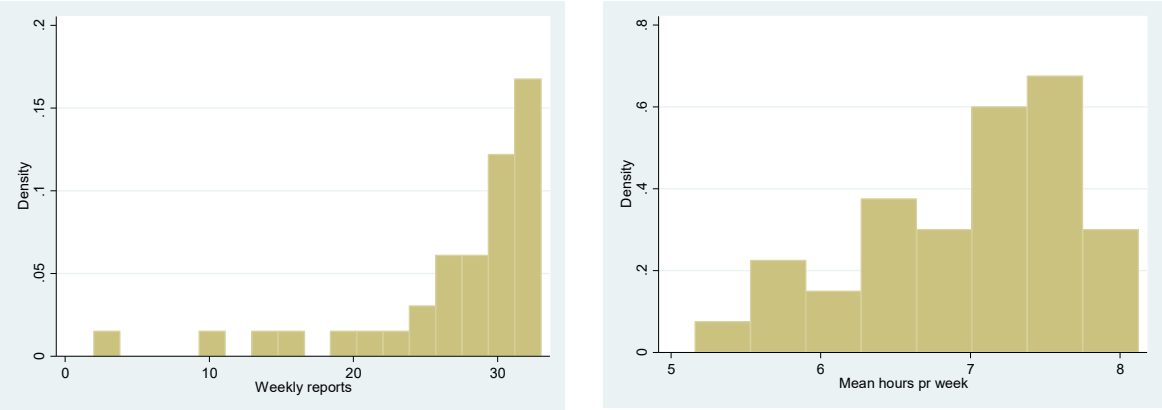
A3. Compliance

We assessed treatment compliance in a brief weekly electronic questionnaire where teachers reported on fidelity of implementation, including how many hours spent implementing the learning activities. The teachers were told to spend eight hours per week on the activities in mathematics, language, executive functioning, and social skills. We requested response to the questionnaire for a total of 34 weeks, which excluded five vacation weeks during fall, Christmas, winter and Easter. The first questionnaire was in early September 2016, and the last in June 2017. Figure A1 (left) reports number of responses submitted by each preschool center. As can be seen, the majority of centers submitted reports in most of the 34 weeks.

In the weekly reports, teachers were asked how many hours were spent on the learning activities during the previous week. Among the 974 weekly reports submitted, 67 percent reported spending eight or more hours on the learning activities the previous week, and only in 16 percent of the reports, the teacher reported they spent less than six hours on learning activities. In the open comment field that was included in the questionnaire, teachers gave reasons for not complying with the eight hours they had committed to. Reasons typically included teacher absence due to health issues or other practical issues that prevented them from following their plans.

Finally, Figure A1 (right) shows the distribution of average number of hours per week across the year spent on learnings activities for each preschool center. We find that 60 percent of all centers spent at least 7 hours per week on the learning activities.

Figure A1: Distributions of number of weekly reports (left) and average weekly hours spent on learnings activities (right) for each preschool center. N = 36 centers.



Additionally, we assessed the teachers' perceived relevance, importance, and benefit of the intervention in anonymous evaluations of the credit base university class (spring 2016, response rate 85 percent) and the curriculum (spring 2017, response rate 80 percent). All teachers agreed that they found the material and activities in the class relevant for their work as practitioners. Their overall rating of the course was on average 4.9 on a five-point scale. When assessing the curriculum, all teachers agreed to the statements "The children have

enjoyed working with the learning activities" and "The children have learned a lot from working with the learning activities." All but one teacher agreed that they would continue to use the curriculum with the five-year-olds in the next preschool year.

A4. Cost-Effectiveness

Total Costs of Curriculum Intervention in 2016 NOK	Per teacher	All teachers (n=41)
	2016 NOK	2016 NOK
Teacher time spent on education	89,000	3,649,000
Teacher time spent on implementation	222,000	9,102,000
Material	12,000	492,000
Cost of education and coaching*	38,000	1,558,000
Total	**361,000**	**14,801,000**

* Number taken from: Reiling, R. B., Snåre, M., Finnanger, E., Rikter-Svendsen, T., Bjørnstad, S., & Aamodt, P. O. (2014), Hva koster en student? En kostnadskartlegging av universiteter og høgskoler.

Cost per student in 2016 NOK**	**37,951**

** Number of children receiving the intervention: 390 (includes children with no parental consent for assessment)

Cost per student in 2016 USD***	**4,298**

*** 1 USD = 8.83 NOK, January 3rd 2016

Cost effectiveness ratio (increase in SD per 2016 USD 1000)	
Full population (treatment impact=.13 SD)	**0.03**
Targeted towards low quality centers (treatment impact= .22 SD)	**0.05**
Accounting for training benefit for three future cohorts	**0.22**

A5. Robustness Tables

In Table A1 we report our main results when children with late consent or missing all T1 test scores are excluded, and in Table A2 we report our main results for each of the six tests.

Table A1. Treatment effect on test scores at post-intervention (T2) and in the one-year follow-up (T3). Sample excluding children with (A) late consent, and (B) not participating in pre-intervention assessment.

	Post-intervention (T2)				Follow-up (T3)			
	Sum score	Math	EF	Language	Sum score	Math	EF	Language
Panel A: Sample excluding children with late consent								
Treat	0.070	0.094	0.071	-0.002	0.122	0.216*	0.047	0.036
	(0.069)	(0.091)	(0.059)	(0.073)	(0.079)	(0.078)	(0.070)	(0.081)
Wild P	0.322	0.323	0.244	0.976	0.143	0.0147	0.511	0.680
N	532	530	532	528	534	534	533	532
Adj. R2	0.601	0.411	0.510	0.489	0.518	0.364	0.379	0.469
Panel B: Sample excluding children not participating in pre-intervention T1 assessment								
Treat	0.111	0.152	0.114+	-0.001	0.136+	0.209**	0.072	0.052
	(0.064)	(0.096)	(0.056)	(0.073)	(0.073)	(0.062)	(0.062)	(0.093)
Wild P	0.101	0.168	0.0614	0.988	0.0841	0.00660	0.234	0.603
N	625	623	625	621	629	629	628	627
Adj. R2	0.625	0.450	0.504	0.540	0.542	0.377	0.399	0.500

Note: ** p<0.01, * p<0.05, + p<0.1. Each column presents regression coefficient of treated (standard error) using ordinary least squares. For both assessment periods: We control for baseline test scores, gender, birth month, parental characteristics (mother and father's education level, earnings, an indicator for non-western country of birth, and indicators for randomization block. In all models in Panel A we add a control for not having participated in the T1 assessment, and in Panel B for late consent. All regressions are clustered on randomization block. We have utilized the boot-test package in Stata to do a few-cluster-correction of the p-value, reported in the table as Wild-P.

Table A2: Treatment effect on all six test scores at post-intervention (T2) and in the one-year follow-up (T3).

	Post-intervention (T2)						Follow-up (T3)					
	Math	HTKS	Digit Span	Hearts & Flowers	Phon. Awareness	Vocabulary	Math	HTKS	Digit Span	Hearts & Flowers	Phon. awareness	Vocabulary
Panel A: Full sample												
Treat	0.155	0.028	0.216*	0.078	0.066	-0.042	0.228**	0.034	0.028	0.075	0.052	0.021
	(0.093)	(0.051)	(0.079)	(0.072)	(0.102)	(0.035)	(0.063)	(0.087)	(0.062)	(0.077)	(0.125)	(0.043)
Wild P	0.145	0.600	0.0180	0.292	0.503	0.239	0.00430	0.718	0.649	0.345	0.716	0.633
N	650	645	641	635	645	648	661	659	653	660	658	659
Adj. R2	0.440	0.288	0.403	0.391	0.202	0.639	0.364	0.208	0.299	0.250	0.174	0.607
Panel B: Preschool center quality												
Treat	0.204	0.000	0.165	0.165	0.045	0.049	0.342**	-0.063	0.117	0.111	0.112	0.153
	(0.131)	(0.111)	(0.099)	(0.119)	(0.128)	(0.063)	(0.085)	(0.128)	(0.084)	(0.110)	(0.102)	(0.088)
Treat*High	-0.063	0.043	0.048	-0.097	0.030	-0.143	-0.223*	0.116	-0.177	-0.099	-0.170	-0.217+
	(0.147)	(0.147)	(0.154)	(0.123)	(0.147)	(0.096)	(0.081)	(0.190)	(0.140)	(0.132)	(0.172)	(0.120)
Wild P (treat)	0.190	1	0.154	0.208	0.724	0.450	0.00760	0.668	0.201	0.375	0.296	0.117
Wild P (t*h)	0.690	0.766	0.758	0.466	0.839	0.159	0.0217	0.607	0.243	0.456	0.372	0.112
N	636	631	627	621	631	634	648	646	640	647	645	646
Adj. R2	0.439	0.284	0.396	0.380	0.201	0.626	0.348	0.185	0.304	0.249	0.156	0.584
Panel C: Baseline skills												
Treat	0.167	0.170	0.157	0.025	-0.013	-0.123+	0.233*	-0.015	0.089	-0.062	0.017	-0.025
	(0.112)	(0.097)	(0.096)	(0.110)	(0.090)	(0.063)	(0.105)	(0.141)	(0.067)	(0.123)	(0.135)	(0.068)
Treat*High	-0.020	-0.286	0.080	0.157	0.170+	0.198+	-0.023	0.044	-0.122	0.234+	0.035	0.140
	(0.125)	(0.171)	(0.096)	(0.116)	(0.079)	(0.095)	(0.146)	(0.133)	(0.142)	(0.119)	(0.147)	(0.107)
Wild P (treat)	0.198	0.117	0.139	0.817	0.886	0.0845	0.0480	0.922	0.200	0.619	0.898	0.728
Wild P (t*h)	0.873	0.123	0.417	0.193	0.0542	0.0506	0.878	0.734	0.410	0.0648	0.815	0.216
N	650	645	641	635	645	648	661	659	653	660	658	659
Adj. R2	0.446	0.286	0.410	0.385	0.204	0.634	0.350	0.185	0.311	0.258	0.156	0.589
Panel D: Parent education												
Treat	0.152	-0.062	0.179+	0.048	0.168	-0.059	0.234*	0.003	-0.039	-0.045	0.001	0.057
	(0.108)	(0.077)	(0.087)	(0.125)	(0.138)	(0.078)	(0.099)	(0.141)	(0.113)	(0.140)	(0.192)	(0.079)
Treat*High	-0.000	0.160	0.041	0.083	-0.169	0.058	-0.017	-0.006	0.076	0.217	0.106	-0.012
	(0.121)	(0.105)	(0.103)	(0.146)	(0.140)	(0.100)	(0.128)	(0.159)	(0.139)	(0.205)	(0.183)	(0.093)
Wild P (treat)	0.186	0.426	0.0460	0.716	0.245	0.468	0.0416	0.988	0.792	0.754	0.996	0.502
Wild P (t*h)	0.997	0.147	0.691	0.578	0.255	0.602	0.900	0.974	0.620	0.305	0.622	0.901
N	639	634	630	624	634	637	649	647	642	648	646	647
Adj. R2	0.439	0.276	0.395	0.379	0.206	0.617	0.345	0.171	0.282	0.249	0.152	0.565

Note: ** p<0.01, * p<0.05, + p<0.1. Each column in each panel presents regression coefficients with standard errors in parenthesis, using OLS. All models control for gender, birth month and parental characteristics (education, earnings and indicator for non-western country of birth), baseline test scores, indicators for late consent and not having participated in the T1 assessment, and randomization block, all regressions clustered on randomization block. Preschool center quality is measured as the preschool center fixed effect (center average covariate adjusted test score). High/low center quality is split at median value. Parental education is measured as the average of mother's and father's number of years of education. High/low parental education and high/low baseline skills are split at median value. We have utilized the boottest package in Stata to do a few-cluster-correction of the p-value, reported in the table as Wild-P(treat) and Wild-P(t*h).

Table A3. Attrition

	T1	T2	T3
Treat	-0.017	-0.003	0.020
	(0.023)	(0.013)	(0.019)
Wild P	0.458	0.821	0.337
N	691	691	691
Adj. R2	0.007	0.004	0.025

Note: Each column presents regression coefficient of treated (standard error) using ordinary least squares. Outcome variables are constructed as indicators for participating in assessments at T1, T2, and T3. In all models we control for and cluster on randomization block. We have utilized the boot-test package in Stata to do a few-cluster-correction of the p-value, reported in the table as Wild-P.

Table A4. Association between center quality and center characteristics

	Model 1	Model 2	Model 3	Model 4
Hours in all age groups	-0.009	-0.104		
	(0.077)	(0.091)		
Hours in 5-year group			0.106	0.112+
			(0.064)	(0.065)
Center size		0.000		0.000
		(0.000)		(0.000)
Teacher/child ratio		7.972+		5.695
		(4.686)		(4.003)
Tenure		0.015		0.011
		(0.022)		(0.021)
N	52	52	52	52
Adj R2	-0.020	-0.004	0.032	0.031

Note: + $p<0.1$. Outcome variable is the standardized quality indicator. Each column presents regression coefficient (standard error) of center characteristics using ordinary least squares.

Table A5. Summary statistics for blocks

Block	Treat mean	Z2 mean (SD)	Z3 mean (SD)	#children	#centers
1	0.515	0.012	0.247	33	6
		(0.994)	(0.839)		
2	0.446	-0.155	-0.017	56	4
		(0.934)	(0.941)		
3	0.559	-0.363	-0.200	34	4
		(1.174)	(1.366)		
4	0.677	0.049	0.195	31	4
		(0.839)	(0.739)		
5	0.391	-0.476	-0.329	46	6
		(1.176)	(1.107)		
6	0.471	0.159	-0.008	34	4
		(1.009)	(1.005)		
7	0.692	-0.145	-0.418	39	5
		(0.989)	(1.012)		
8	0.554	0.049	-0.060	92	6
		(0.918)	(1.020)		
9	0.618	0.176	0.317	55	4
		(0.972)	(0.880)		
10	0.643	-0.024	-0.043	42	4
		(0.934)	(1.161)		
11	0.484	0.108	0.117	62	6
		(1.032)	(0.886)		
12	0.609	-0.035	-0.050	69	6
		(0.966)	(0.955)		
13	0.514	0.217	0.031	35	4
		(0.785)	(0.823)		
14	0.448	0.248	0.147	29	4
		(1.161)	(1.014)		
15	0.676	0.245	0.145	34	4
		(1.027)	(1.091)		

Appendix References

Blair, Clancy, and C. Cybele Raver. 2015. "School Readiness and Self-Regulation: A Developmental Psychobiological Approach." *Annual Review of Psychology* 66 (1):711-731.

Bodrova, Elena, and Deborah J Leong. 2007. "Tools of the mind." *Columbus, OH: Pearson*.

California Department of Education. 2016. California Preschool Curriculum Frameworks.

Clements, Douglas H, and Julie Sarama. 2011. "Early childhood mathematics intervention." *Science* 333 (6045):968-970.

Davidson, Matthew C., Dima Amso, Loren Cruess Anderson, and Adele Diamond. 2006. "Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching." *Neuropsychologia* 44 (11):2037-2078.

McClelland, Megan M., Claire E. Cameron, Robert Duncan, Ryan P. Bowles, Alan C. Acock, Alicia Miao, and Megan E. Pratt. 2014. "Predictors of early growth in academic achievement: The Head-Toes-Knees-Shoulders task." *Frontiers in Psychology* 5.

Mol, Suzanne E., Adriana G. Bus, and Maria T. de Jong. 2009. "Interactive Book Reading in Early Education: A Tool to Stimulate Print Knowledge as Well as Oral Language." *Review of Educational Research* 79 (2):979-1007.

Pianta, Robert C. 1999. *Enhancing relationships between children and teachers*. Washington, DC, US: American Psychological Association.

Schmitt, Sara A, Megan M McClelland, Shauna L Tominey, and Alan Acock. 2015. "Strengthening school readiness for Head Start children: Evaluation of a self-regulation intervention." *Early Childhood Research Quarterly* 30:20-31.

Shure, Myrna B. 1992. *I can problem solve (kindergarten and primary grades): An interpersonal cognitive problem-solving program for children*: Research Press.

Størksen, Ingunn, Ingunn T. Ellingsen, Maren S. Tvedt, and Ella MC Idsøe. 2013. "Norsk vokabulartest (NVT) for barn i overgangen mellom barnehage og skole: Psykometrisk vurdering av en nettbrettbasert test." *Spesialpedagogikks forskningsdel* 04/13:40 - 54.

Størksen, Ingunn, Dieuwer ten Braak., Svanhild Breive, Ragnhild Lenes, Svanaug Lunde, Martin Carlsen, Ingvald Erfjord, Per Sigurd Hundeland, and Mari Rege. 2018. *Lekbasert læring - et forskningsbasert førskoleopplegg fra Agderprosjektet* Oslo: GAN Aschehoug.

ten Braak, Dieuwer, and Ingunn Størksen. 2021. "Psychometric properties of the Ani Banani Math Test." *European Journal of Developmental Psychology*: 1-18.

Wechsler, David. 1991. *WISC-III: Wechsler intelligence scale for children: Manual*: Psychological Corporation.

Weisberg, Deena Skolnick, Kathy Hirsh-Pasek, and Roberta Michnick Golinkoff. 2013. "Guided Play: Where Curricular Goals Meet a Playful Pedagogy." *Mind, Brain, and Education* 7 (2):104-112.

Wells, Gordon. 1999. *Dialogic inquiry: Towards a socio-cultural practice and theory of education*: Cambridge University Press.