European Economic Review 154 (2023) 104413

Contents lists available at ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/eer

Reducing the gender gap in early learning: Evidence from a field experiment in Norwegian preschools☆

Andreas Fidjeland [a,b,*], Mari Rege [b], Ingeborg F. Solli [b], Ingunn Størksen [c]

[a] Nordic Institute for Studies of Innovation, Research, and Education, Norway
[b] University of Stavanger Business School, Norway
[c] Norwegian Centre for Learning Environment and Behavioural Research in Education, University of Stavanger, Norway

ARTICLE INFO

JEL classification:
I20
H42

Keywords:
Gender gap
Field experiment
School readiness
Child development
Early childhood education
Preschool

ABSTRACT

Substantial gender differences in key academic skills appear even before children start formal schooling. Although increasing investments in early-childhood programs is motivated by efforts to promote equality of opportunity in education, program attendance seems to have less effect on boys. In this field experiment, we investigate whether a more structured curriculum can help preschools reduce the gender gap in early learning. While girls have higher skills at baseline, we find that the intervention primarily benefits boys, thereby reducing the gender skill gap, with effects persisting into formal schooling.

1. Introduction

An extensive literature documents persistent gender gaps in academic outcomes, across a variety of educational contexts (e.g., Bedard and Cho, 2010; Autor and Wasserman, 2013; DiPrete and Buchmann, 2013; OECD, 2015). Girls consistently outperform boys. Moreover, boys disproportionately dropout of school, have behavioral problems and special needs. These negative outcomes in early schooling spill over into adulthood — with links to college enrollment, unemployment, and even crime (Goldin et al., 2006; Vincent-Lancrin, 2008; Fortin et al., 2015).[1] Although the origins of these gaps are not fully understood, there is evidence of substantial gender differences appearing even before children start formal schooling (Magnuson and Duncan, 2016; Brandlistuen et al., 2020). In many countries, a perception that boys are often less school ready than girls of similar age has caused an increasing prevalence of delayed school entry for boys. However, there is little evidence to suggest that such "academic redshirting" has long-term positive effects on child development and educational attainment (Deming and Dynarski, 2008).

☆ We are thankful for helpful comments and suggestions from Hans H. Sievertsen, Kristiina Huttunen, Henning Hermes, John Eric Humphries, Maximiliaan Thijssen, May Linn Auestad, and from seminar participants at the University of Stavanger, as well as conference participants at the AEA meetings, the AEFP conference, EALE, the Lund Ph.D. Workshop in the Economics of Education, and the annual meeting of the Norwegian Association for Economists. Funding from the Norwegian Research Council, Norway, grant numbers 237973 and 270703, as well as Sørlandet Knowledge Foundations, Norway and the Agder County administrations, Norway is gratefully acknowledged. The project was preregistered in the AEA registry, code AEARCTR-0002241.

* Correspondence to: Nordic Institute for Studies of Innovation, Research, and Education, P. O. Box 2815 Tøyen, 0608 Oslo, Norway.
E-mail addresses: andreas.fidjeland@nifu.no (A. Fidjeland), mari.rege@uis.no (M. Rege), ingeborg.f.solli@uis.no (I.F. Solli), ingunn.storksen@uis.no (I. Størksen).

[1] Even though, over the past decades, women have surpassed men by substantial (and increasing) margins in terms of educational attainment, it is still the case that women, on average, earn less than men and are disproportionately less likely to hold powerful positions in society (DiPrete and Buchmann, 2013). However, while women might still face barriers to capitalizing on their education in the labor market, the earnings gap has halved over the past 40 years despite increasing wage inequality overall (Blau and Kahn, 2007).

https://doi.org/10.1016/j.euroecorev.2023.104413
Received 26 April 2022; Received in revised form 17 February 2023; Accepted 22 February 2023
Available online 8 March 2023
0014-2921/

In contrast, early-childhood education and care (ECEC) programs have attracted increasing interest from policymakers as research has demonstrated how participation in such programs can improve school readiness, with potential long-term gains in academic and labor market outcomes (Berlinski et al., 2008; Heckman et al., 2010; Melhuish, 2011; Felfe et al., 2015; Cornelissen et al., 2018; Felfe and Lalive, 2018). Offering universal ECEC programs could therefore be an efficient tool for policymakers to ensure that all children are adequately prepared for the transition to formal schooling. However, the variety of contexts, designs, and features in existing programs makes the research literature far from unified with respect to the conditions necessary for child development and fostering school readiness (White et al., 2015; Phillips et al., 2017). Even less is known about the distribution of potential effects, and about whether the conditions for realizing these benefits are similar for all children (Duncan and Magnuson, 2013; Phillips et al., 2017). The evidence on gender-specific returns to ECEC enrollment is mixed (Magnuson et al., 2016), but several studies report results indicating that girls might benefit more in terms of skill development than boys (e.g., Goodman and Sianesi, 2005; Anderson, 2008; Havnes and Mogstad, 2015; Felfe et al., 2015; Cornelissen et al., 2018; Fessler and Schneebaum, 2019). However, so far the literature provides little evidence on why this might be the case, or how preschools can help mitigate these gaps in early learning.

One potential explanation for the gender gap in school readiness is that girls and boys spend their time in childcare differently. Most existing programs in the U.S. (e.g., *Head Start*) as well as many European countries, including Norway, are based on *whole-child* curricula which puts significant emphasis on free play and child-initiated activities, while only providing general guidance to teachers on how to foster productive learning environments (Jenkins et al., 2018). As a result, children often spend large portions of the day not engaged in learning activities, or engaged in free play without adult interaction and involvement (Karlsen and Lekhal, 2019). In these unstructured preschool settings, prior research indicate that girls are more likely to engage in activities that promote school readiness and skill development, for example by seeking out interactions with the teachers (Tonyan and Howes, 2003; Ruble et al., 2006; Early et al., 2010; Størksen et al., 2015; Stangeland et al., 2018). In contrast, boys engage more in spontaneous and physical behavior, shifting attention between activities rapidly and interacting with adults to a lesser extent. This suggests that boys may be less exposed to many of the stimulating learning activities that girls seem inclined to engage in of their own accord.

These observational insights suggest that a more structured curriculum where the teachers take a more active role might be particularly beneficial for boys, thereby decreasing gender gaps in early learning. We investigate this hypothesis using data from a randomized controlled trial (RCT) carried out in the context of the universal preschool system in Norway (see Rege et al., 2021, for aggregate treatment results and Rege et al., 2017 for the AEA RCT Registry entry). The intervention introduced a more intentional practice through a structured, comprehensive curriculum for groups of five-year-olds in their final year of preschool, with the goal of improving their school readiness. This practice contrasts with the prevalent Norwegian ECEC pedagogical philosophy, which largely centers on child-initiated free play.

To that end we recruited 71 ECEC centers, and provided teachers in the treatment group with a curriculum encompassing age-appropriate intentional skill-building activities to implement for all their five-year-olds. This curriculum was coupled with a thorough professional-development program as well as further support throughout the intervention. The activities were embedded within a playful learning approach and targeted key school-readiness skills in the areas of mathematics, language, and executive functioning.

We rely on data collected through detailed one-to-one assessments by certified testers blind to treatment status to investigate the effects of the intervention. We find a substantial gender gap in school readiness at baseline, and this gap is not mitigated by business-as-usual pedagogical practice. Moreover, consistent with our hypothesis, we find that the average improvement in school readiness brought about by the intervention appear to be entirely driven by a treatment effect of about 20 percent of a standard deviation (0.2σ) on the boys. In contrast, we find little evidence that the curriculum had any effect on the girls. Lack of precision prevents us from making conclusive statements about specific skill domains, but we present point estimates that in terms of magnitude and direction consistently points toward the curriculum having a positive effect on boys and little effect on girls. This is true both for the posttreatment assessment and for a one-year follow-up at the end of first grade of primary school, where we find that the positive effects on boys persist across the transition to formal schooling. We also find suggestive evidence that boys at the bottom of the skill distribution at baseline are the ones who improve the most. In a heterogeneity analysis we estimate decreasing treatment effects as we move up the rank distribution of baseline scores, suggesting that the efficacy of the intervention decreases with initial skill level. For girls, we find no such relationship.

Our study generates results with implications for the design of ECEC curricula and pedagogical practice. In particular, our results are relevant for the many countries experiencing a push toward universal provision of preschool programs, often with the policy objective of facilitating equal opportunities for early learning. We contribute to meeting the need for robust evidence on how the design and content of such programs might have heterogeneous impacts across children. The curriculum in existing universal programs tend to be fairly nonspecific and unstructured, emphasizing the value of free play, autonomy and spontaneous engagement between teacher and child (Engel et al., 2015; White et al., 2015). This holistic approach to child development gives substantial discretion to teachers regarding the pedagogical content, but implementing these types of curricula in a manner that translates to child development requires substantial skill, leading to heterogeneity in the quality of teaching across preschools (Jenkins et al., 2018; Rege et al., 2018). Our results suggest that implementing a more structured curriculum can help teachers support the development of both boys and girls, potentially reducing gender gaps in early learning. In addition, the persistence of the effects as the children transition to formal schooling suggests that narrowing learning gaps in preschool can help reduce gender gaps in later academic achievement as well.

Our study also informs the literature on the nature and origins of academic gender gaps in several ways. First, we contribute with evidence that speaks to the heterogeneous impacts of childcare participation, using extensive, high-quality assessment data. Second, we provide causal evidence on a plausible channel through which these heterogeneous impacts might develop, as we provide evidence suggesting that structured and intentional practice is an important channel for promoting the development of boys in particular. Third, our study design allows us to follow the development of early learning gaps in the crucial transition from early-childhood care to formal education. By studying the effects of the curriculum in both the preschool and formal school setting, our results could also inform the literature of how gender gaps in educational outcomes later in school might be connected to the type of program the children were exposed to in preschool. In general, our results underscore the need for a better understanding of what constitutes process quality in early-childhood education (Phillips et al., 2017) – as opposed to structural quality, which has been the focus of much economic research – in order to fully understand how ECEC participation might have heterogeneous impacts on child development.

2. The scope and origins of gender gaps in early learning

The extant literature on gender gaps in early learning suggests that girls start formal schooling with a significant advantage in school-readiness skills (e.g., DiPrete and Jennings, 2012; Magnuson and Duncan, 2016; Brandlistuen et al., 2020). Girls consistently score better on a range of measures of early learning, such as literacy and language skills (Lundberg et al., 2012; Stangeland et al., 2018), self-regulation and executive functioning (Matthews et al., 2009; Lenes et al., 2020), and social and behavioral skills (DiPrete and Jennings, 2012). While the evidence on early gender gaps in mathematics is more mixed,[2] previous studies in the Norwegian context have uncovered substantial skill differences favoring girls in this domain as well (Brandlistuen et al., 2020).

Although gender gaps at school start are not necessarily *caused* by the preschool programs children may have attended, several studies report results suggesting that such programs might be particularly beneficial for the development of girls, although gender differences in program effectiveness are rarely an explicit focus in those studies (Magnuson et al., 2016). Anderson (2008) provides a prominent example by revisiting the Perry Preschool, Abecedarian, and Early Training projects to find that these model programs benefited girls over boys by a difference of about 0.4σ. While differential effects of such magnitudes have rarely been replicated elsewhere, a meta-analysis of 23 early preschool programs revealed that girls benefited significantly more than boys in terms of cognitive, achievement, behavioral, and mental-health measures, although the differences were small (Magnuson et al., 2016). In contrast, boys benefited much more on other school outcomes such as detention and need for special education. There is also some evidence that programs affect boys more in the long term (Domond et al., 2020; Gray-Lobe et al., 2023).

There are at least three plausible mechanisms through which preschool programs might affect boys and girls differently. First, if girls enter preschool age with better-developed pre-academic skills – and skills beget skills (Cunha and Heckman, 2007) – we would expect different developmental trajectories for boys and girls during preschool. However, not only do boys and girls typically enter preschool at roughly equal levels of development (Magnuson et al., 2016), there is seemingly no gender difference in the aptitude for developing early language and mathematics skills (Spelke, 2005).

Second, preschool teachers might be inclined to foster learning environments better suited for girls' development. For example, the lack of male role models in ECEC might be particularly detrimental for boys (Sumsion, 2005), and there is indeed some evidence that increasing the share of male teachers would be beneficial for their development (Gørtz et al., 2018; Drange and Rønning, 2020). Other studies have indicated that teachers have different expectations toward the behavior of boys and girls, where the latter to a larger extent is expected to behave in self-regulated manners, such as sitting still and waiting for their turn (Lenes et al., 2020).

Third, girls and boys might spend their time in preschool differently, particularly in unstructured settings. Broadly, girls have been found to be more likely to spend their time in cognitively stimulating activities during free-play time, while boys are more likely to engage in spontaneous and physical behavior (Tonyan and Howes, 2003; Ruble et al., 2006; Early et al., 2010). Further, girls are more likely to interact with adults, hence developing higher-quality relationships with teachers. In turn, the degree of teacher–child interaction has been found to be a consistent indicator of classroom quality as well as a predictor of child development in preschool settings (Mashburn et al., 2008). Indeed, Howes et al. (2008) find that effective preschool classrooms are characterized by intentionality in the way teachers engage and interact with the children, and that learning opportunities are more plentiful in classrooms where teachers manage time more actively. Moreover, Early et al. (2010) note that the pattern of gender-stereotypical time-use was not observed when the teachers selected activities for the children.

In this context, a strand of research highlights the potential benefits of more structured curricula for child development (Diamond et al., 2007; Clements and Sarama, 2011; Weiland and Yoshikawa, 2013; Schmitt et al., 2015; Dillon et al., 2017). These studies argue that preschool staff can target school-readiness competences through an intentional and systematic approach to learning situations. This could be particularly important for boys, who might need more support and scaffolding from teachers to engage in stimulating activities (Størksen et al., 2015). In this paper, we therefore ask whether providing staff with such structured learning activities to carry out with *all* the children could be a way for preschools to improve boys' school readiness, so that they start formal schooling on a more equal footing with girls.

3. Institutional background

Norway invests heavily in ECEC by subsidizing all preschool centers that adhere to governmental regulations. Parental payments are capped at approximately USD 300 per month per child, which reflects about 15 percent of the total cost of childcare

[2] Illustratively, while Magnuson and Duncan (2016) report a small advantage in mathematics for boys using data from the 2010 wave of the Early Childhood Longitudinal Study–Kindergarten Cohorts, they find a gap of similar magnitude favoring girls in the 1998 wave. In a similar vein, Robinson and Lubienski (2011) find little evidence of gender differences in math in kindergarten, yet teachers consistently rate girls as better in math. Girls also earn better grades in math throughout school, even when controlling for performance on achievement tests (Kenney-Benson et al., 2006).

enrollment (Norwegian Directorate for Education and Training, 2019). As a result, Norway is among the OECD countries with the highest public spending on ECEC. Children are typically enrolled in ECEC between ages 1 and 2, and all children are guaranteed enrollment in a center in their municipality. These rights and subsidies have led to near-universal take-up. As of 2020, 92.2 percent of all children aged 1–5 years were enrolled in formal childcare, and 97.5 percent of all five-year-olds.[3] Once enrolled, most children remain in ECEC until they start compulsory schooling in August of the calendar year in which they turn six.

The Norwegian center-based ECEC is founded on a social pedagogical tradition emphasizing free play and child-initiated activities in preschool child groups (Engel et al., 2015). While play is seen as an activity that may facilitate learning, a cornerstone of the Norwegian philosophy is that play has intrinsic value and is a goal in and of itself. There is no set curriculum to guide the provision of ECEC. Rather, centers are given substantial discretion in how to structure daily activities for the children, following the goals of a framework plan that loosely outlines the purpose, values, and learning areas of Norwegian ECEC (OECD, 2015).[4] Against this backdrop, current pedagogical practice emphasizes unstructured and spontaneous play to such an extent that it is often given priority over adult-driven activities even when those are preplanned and scheduled (Synodi, 2010). Indeed, Karlsen and Lekhal (2019) find in their case studies that 60 percent of the ECEC day consisted of free-play activities and that center staff spent almost half of that time away from play situations, indicating that children spend a significant portion of their time without interacting with adults.

This emphasis on free and autonomous play contrasts with the school-readiness approach found in parts of the US and UK, in which employing skill-targeted curricula to prepare children for formal schooling is a more explicit pedagogical objective. The structured yet playful curriculum intervention investigated here should be seen as a step toward a more intentional school-readiness perspective in Norway. While the intervention does not abandon the tenets of the social-pedagogical tradition, it does incorporate some of the intentional and structured practices of the school-readiness philosophy.

4. Experimental design and measures

4.1. Experimental design

We recruited participants from two Norwegian counties. At first we invited all 30 municipalities in those counties to sign up for the project, of which 15 did. Then we invited all publicly regulated childcare centers operating in those 15 municipalities to participate. Out of 190 centers, 72 signed up. One center in the control group later withdrew from the project, leaving us with a sample of 71 participating centers.

We deemed it necessary to rely on voluntary participation in order to maximize the likelihood of successful project implementation. Since project participants were not only carrying out the intervention in the preschools, but also important contributors to the development of the curriculum itself, recruiting a highly motivated and competent pool of participants was a crucial component for ensuring fidelity, and has been recognized as a key success criterion in the implementation theory literature (Bertram et al., 2015; Størksen et al., 2021). This self-selection mechanism likely contributed to the high level of compliance and fidelity in the implementation of the intervention, However, as we would expect that the preschool teachers who participated in the project are more invested in and motivated for the project, on average, than the general population of preschool teachers in Norway our focus on implementation quality is likely to come at the expense of the generalizability of our results.

For the randomization procedure, we split centers in 15 blocks, matched for size and geographic location. The resulting blocks consisted of 4 to 6 centers, with the total number of children ranging from 29 to 92. Parental consent was collected prior to randomization, but we accepted late consenters because of the lengthy time between initial collection and the start of the intervention. Of 701 parental consents collected (92 percent consent rate), 18.8 percent were submitted after our initial deadline. The late consenters were skewed toward the treatment group. This could be because teachers, charged with collecting the consent sheets, might have been more invested in the project and worked harder to get parents to sign up once they became aware that they would be in the treatment group. We therefore include an indicator for late consent in all our estimations. However, excluding all late consenters from the sample yields broadly similar results (see Table D.5 in the appendix).

As the curriculum was developed for the project and hence not covered by existing preschool-teacher training, the project period started with teachers in the treatment group receiving training in the form of a 15 ECTS (corresponding to half a semester full-time) university course on the pedagogics and practices of *playful learning*.[5] This training also allowed us to obtain extensive feedback from the practitioners on the curriculum and revise accordingly. The teachers subsequently implemented the curriculum in the fall of 2016. We conducted our *baseline* assessment immediately prior to implementation (T1). The treatment group then proceeded with the intervention for nine months, before we conducted the *postintervention* assessment in late spring 2017 (T2). We reconnected with the children one year later for a *follow-up* assessment once at the end of first grade (T3).

Throughout the intervention period, the control group followed a business-as-usual condition. However, teachers in the control centers were informed that they would receive the training, curriculum documents, and any accompanying materials when the participating children left for primary school. This was made clear early in the process, in an effort to mitigate any discouragement effects in the control group that might bias the results.[6]

[3] Aggregate data on participation is available at https://www.ssb.no/en/utdanning/barnehager/statistikk/barnehager.

[4] An English-language version is available at https://www.udir.no/globalassets/filer/barnehage/rammeplan/framework-plan-for-kindergartens2-2017.pdf.

[5] Størksen et al. (2021) detail the theories underpinning the practical implementation of the project.

[6] Our project did not involve parents, and it was up to the centers to keep them informed. As we did not survey the parents, we cannot speak to how aware they were of the project or whether they may have responded in a compensatory manner. However, because the activities in the new curriculum did not differ

4.2. Intervention content

Our intervention is a bundle of several components. The main feature is the preschool curriculum, consisting of structured and intentional skill-building activities centered around *playful learning*. Concretely, we provided a booklet with 130 activities, most of which would be familiar to Norwegian preschools.[7] However, they differed in both design and content, as well as in the level of intentionality with which they were to be implemented. Examples of activities include puzzles and games to cultivate mathematical thinking, dialogical reading to stimulate language, and stories and images where the children had to identify emotions.[8]

We encouraged teachers to develop their own approach to the curriculum. The activities were flexible, allowing teachers to adapt their difficulty and complexity to the needs of their child group. The only requirement was for teachers to commit to spending at least 8 h a week doing activities with their five-year-olds. Given that nearly all Norwegian children of this age spend 30–40 h a week in childcare, our intervention took up only a modest proportion of the preschool schedule. Even so, it represented a substantial increase in time devoted specifically to stimulating school-readiness skills, which, according teacher reports, ranged from 0 to 3 h a week prior to the intervention.

To improve implementation quality, we assisted the teachers throughout the intervention. Three times per semester, a team member would call the teachers to answer questions or discuss challenges they faced. The teachers were also required to answer a weekly questionnaire in which they reported on implementation fidelity. Because child groups are typically mixed-age, the five-year-olds had to be separated from the rest of their group for the activities. The teachers also needed time to prepare the week's activities. To cover the centers' costs associated with hiring additional and temporary staff we provided them with lump-sum transfers equivalent to the cost of a part-time (50 percent) position for nine months. Similarly, we covered the costs of a 50-percent position for four months as teachers participated in the training course.

4.3. Measures

We center our assessment around the construct of *school readiness* (Bennett and Tayler, 2006). Conceptually, this encapsulates skills that support a successful transition to formal schooling, and constitute the foundation on which later learning is achieved. The assessment consisted of age-appropriate, validated tests for measuring early skills. All tests were conducted in one-to-one sessions using a tablet computer, both for the child to interact directly with and for the tester to record answers and scores on.

We used the following measures to assess the children's skills:

Numeracy — To measure early mathematical skills, we used the Ani Banani Math Test (ABMT), which assesses the understanding of numeracy, geometry, and problem solving using a playful tablet application. The children help a monkey with different tasks, such as counting bananas and setting the table with enough plates for their birthday-party guests. ten Braak and Størksen (2021) assess the psychometric properties of the ABMT and find strong predictability of later mathematical achievement. They do note signs of gender bias in three items, but this was not consistent across samples, and nor was gender predictive of the latent construct. Still, for robustness we run our analyses both with and without those three items, finding that this does not affect results (see Table D.6 for details).

Language — To assess vocabulary, we use a short version of the Norwegian Vocabulary Test (NVT; (Størksen et al., 2013)). The children were presented with images on the tablet and asked to name the object depicted. To assess phonological awareness, we used a 12-item blending task that is part of the official literacy screening battery of the Norwegian Directorate for Education and Training. A word is presented in phonemes, and the child has to select the one out of four options on the tablet that corresponds to that word.

Executive functioning — To assess executive functioning, we used three tests: Wechsler's Digit Span Test (Wechsler, 2003) for measuring working memory; the Head–Toes–Knees–Shoulders task (McClelland et al., 2014) to assess behavioral self-regulation; and the Hearts and Flowers task (Davidson et al., 2006), which is widely used to measure cognitive flexibility in young children.

From these tests we created three outcome measures (numeracy, language, and executive functioning) by standardizing scores within each wave to a mean of 0 and a standard deviation of 1. We then averaged across tests within each domain and re-standardized the resulting index. We also averaged across indices and re-standardized to construct a sum-score measure.

We conducted the T1 and T2 assessments at local science museums. Participating centers were invited to spend a day at the museum, and we assigned a time slot for assessment. At that point, the children were lined up at the assessment station by the center staff. Children standing in line were continuously assigned to the next available tester, meaning that child-to-tester matching was naturally randomized. Testing took place over several days, and centers from both the treatment and control groups were invited each day. The testers were blind to treatment status, and they had been trained and certified prior to data collection. At T3, the children had moved on to primary school and hence spread across multiple sites. For this wave, testers traveled to schools, where staff let children leave the classroom to be assessed.

radically from existing practices and represented only a small part of the regular schedule, we believe that such adjustments from parents are unlikely. Also, opportunities to move children between centers, for example to enter the treatment group, are very limited in the short-term, as centers do not typically have available slots in the middle of the academic year.

[7] The booklet was also published after the end of the project, and is now widely available to practitioners and researchers (Størksen et al., 2018).

[8] See appendix material in Rege et al. (2021) for a more comprehensive description of the curricular content.

Table 1
Descriptive statistics and balance test.

	Boys			Girls		
	Control	Treat	Difference	Control	Treat	Difference
Child characteristics						
Birth month	6.380	6.249	−0.265	6.139	6.091	0.028
	(3.153)	(3.260)	(0.386)	(3.307)	(3.091)	(0.305)
Immigrant	0.114	0.161	0.041	0.128	0.224	0.091
	(0.319)	(0.369)	(0.045)	(0.336)	(0.418)	(0.059)
Mother education	14.333	14.128	−0.115	14.433	14.123	−0.224
	(2.495)	(2.520)	(0.260)	(2.602)	(2.635)	(0.291)
Father Education	13.896	13.656	0.260	13.676	13.786	0.007
	(2.426)	(2.422)	(0.291)	(2.640)	(2.532)	(0.310)
Mother earnings	344,680	329,301	−17,695	345,774	333,160	−12,373
	(225,887)	(200,712)	(31,008)	(216,475)	(200,750)	(30,897)
Father earnings	571,014	565,968	1,805	525,675	559,337	38,296
	(268,260)	(262,916)	(27,804)	(256,313)	(284,157)	(29,256)
Baseline scores						
T1 sum score	−0.123	−0.068	0.016	0.172	0.039	−0.116
	(1.050)	(0.970)	(0.133)	(0.964)	(1.021)	(0.081)
T1 math	−0.117	−0.090	0.010	0.140	0.080	−0.055
	(0.982)	(1.019)	(0.113)	(0.915)	(1.024)	(0.130)
T1 EF	−0.115	−0.044	0.047	0.064	0.089	0.045
	(1.055)	(0.982)	(0.141)	(0.958)	(1.019)	(0.090)
T1 language	−0.065	−0.030	−0.017	0.212	−0.075	−0.270
	(0.986)	(0.941)	(0.110)	(1.069)	(1.018)	(0.126)
Missing T1 scores	0.070	0.052	−0.020	0.046	0.029	−0.024
	(0.257)	(0.222)	(0.032)	(0.211)	(0.167)	(0.026)
N	142	193	335	151	175	326

Note: The columns provide means (standard deviations) for child characteristics and T1 test scores separately by gender and treatment status for the T3 analytic sample. The columns labeled *Difference* represent the estimated coefficient (standard error) from regressing each covariate against treatment status, while controlling for randomization block. Regressions are also clustered on the block level.

5. Data and empirical strategy

5.1. Sample

Of the 701 children for whom we collected parental consent, 658 participated in the baseline assessment while 650 participated in the postintervention assessment. For the follow-up we were able to locate and assess 661 children. Although we did not explicitly balance the sample on gender, shares of boys and girls were equal in each wave, with the difference in absolute numbers ranging from 2 at T1 to 9 at T3.

We construct our analytical samples from the children observed in the T2 and T3 waves, respectively, and run our analysis on these samples separately. Although this means that the T2 and T3 samples are slightly different, there is substantial overlap as 620 children participated in both waves. For the 33 children (4.8 percent of the gross sample) who were not assessed at baseline, we impute their pre-test scores using predicted values based on child and parent characteristics. Specifically, we regress each pre-test measure on gender, birth month, parental education and earnings, immigrant status, and indicators for preschool center, and use the resulting estimates to impute the expected value of the pre-test conditional on the child's observable characteristics. We add an indicator for missing baseline scores in all our estimations. As a robustness check we also replicate our analysis after excluding observations with imputed pre-scores, with results presented in Table D.7 in the appendix. If anything, our results are even stronger when those with impute pre-test scores are excluded from the analysis.

As is evident from our contact rate across waves, attrition was generally low. More importantly, as we show in Table D.2, attrition rates were balanced across gender and treatment status.

5.2. Summary statistics

We combine our assessment data with registry data from Statistics Norway relating to child and parent characteristics. The variables used in our analyses are listed in Table 1, where we report means and standard deviations for the T3 analytical sample separately across gender and treatment status.[9] Birth month is a running variable taking a value of 1 (December) to 12 (January), so that a higher value indicates an older child. Immigrant status is denoted by an indicator taking the value 1 if the child's mother or father is a non-Western immigrant. Mother's and father's education is measured in years of schooling, and their annual earnings

[9] The appendix provides similar information for the T2 sample in Table D.1.

are measured in Norwegian kroner on a running scale rounded to the nearest 50,000. We also report summary statistics on the baseline scores of the children. The values presented correspond to the average of the subgroup relative to a sample mean of 0, in standard deviation units. In the final row, we report the proportion of children without baseline scores.

Table 1 also presents results (in the columns labeled *Difference*) from tests to determine whether child and parent characteristics and baseline scores are balanced across treatment status within each gender. The test consists of regressing the covariate on treatment status while controlling for randomization block. For both genders, background characteristics are sufficiently balanced: no differences significant at conventional levels are uncovered. The magnitudes are also too small to be economically meaningful. We find that treated boys, on average, score somewhat higher at baseline than those in the control group, but their scores are not significantly different. However, we do find a gap in the girls' language score which is of a meaningful magnitude. Such imbalances, even though they might occur by random chance, highlight the importance of controlling for baseline performance, which we do in all our preferred specifications.

5.3. Empirical strategy

We leverage the randomization to treatment to identify the gender-specific effects of our intervention. To quantify these effects, we use ordinary least squares to estimate models of the form

$$y_{i,c} = \alpha + \gamma_1(Boy_i \times T_c) + \gamma_2(Girl_i \times T_c) + \delta Girl_i + \boldsymbol{\beta X_i} + \epsilon_{i,c} \quad (1)$$

where $y_{i,c}$ is the score for the outcome of child i enrolled in center c. Treatment status is denoted by the indicator T_c taking the value 1 if the child's center was randomized to treatment. We interact the treatment indicator with gender so that γ_1 and γ_2 capture the average treatment effect of being treated for boys and girls separately, enabling us to test whether these effects are statistically different from 0. We also report results from tests to determine whether the effects are statistically different from each other. Our preferred specification includes controls for baseline test scores, block fixed effects, and a vector of child and parent background characteristics.[10] We also add indicators for turning in the consent sheet on time and for being assessed at baseline. $\epsilon_{i,c}$ is the error term. We estimate the models separately for each outcome measure, and for T2 and T3 scores.

For our skill-heterogeneity analysis, we extend (1) to include indicators for specific segments of the test-score distribution at baseline. Hence we estimate the model

$$\begin{aligned} y_{i,c} = &\alpha + \phi_1(Boy_i \times T_c \times BS_i^{Boy}) + \phi_2(Girl_i \times T_c \times BS_i^{Girl}) \\ &+ \gamma_1(Boy_i \times T_c) + \gamma_2(Girl_i \times T_c) + \theta_1 BS_i^{Boy} \\ &+ \theta_2 BS_i^{Girl} + \delta Girl_i + \boldsymbol{\beta X_i} + \epsilon_{i,c} \end{aligned} \quad (2)$$

on the T3 sample, where BS_i^{Boy} is an indicator taking the value 1 if child i is a boy with a baseline score in the relevant segment, and BS_i^{Girl} is the female equivalent. We focus on those scoring in the bottom 10, bottom 25, bottom 50, top 25 and top 10 percent. The coefficient ϕ captures the marginal treatment effect of being a child in the particular segment relative to other treated children of the same gender.

We compute standard errors robust to serial correlation by clustering at the randomization block level. A potential concern with this approach is that 15 clusters are too few to provide reliable inference (Cameron and Miller, 2015). Therefore, we also use two alternative approaches that are more robust to small-sample issues. First, we account for the small number of clusters by performing a Wild-T bootstrap procedure. Second, we perform a permutation test (randomization inference) where we randomly reassign treatment status within blocks to estimate a distribution of placebo treatment effects with which we can compare our true effect estimate (Abadie et al., 2020; Athey and Imbens, 2017). In our results, we report p-values obtained from estimations both with and without these corrections.

6. Results

6.1. Descriptive evidence

We begin our presentation of the results with a discussion of the descriptive evidence on early gender gaps in our sample. In Fig. 1 we present differences in school-readiness skills measured by test-score performance at baseline. The bars indicate the gap in average scores between girls and boys in standard deviation units (σ). Assuming we can interpret the *sum score* as a measure of the child's overall school readiness, we find that girls are, on average, about 0.15σ more ready for school than boys. Moreover, we find that girls score better in all domains, with the largest discrepancy found for mathematics (0.2σ).

While average scores may provide useful information, they may also mask important variation over the skill distribution. In Fig. 2 we present density plots for the distribution of baseline scores. We find that boys are more likely to score in the bottom half of the distribution than girls. For all measures, there are about 60 percent boys among those scoring 1.5σ or more below the mean. However, there are equal numbers of boys and girls scoring 1.5σ or more *above* the mean. This suggests that the discrepancies are

[10] We also report results from regressions without baseline scores, and without child and parental covariates in the appendix. See Table D.3 and D.4 for details.

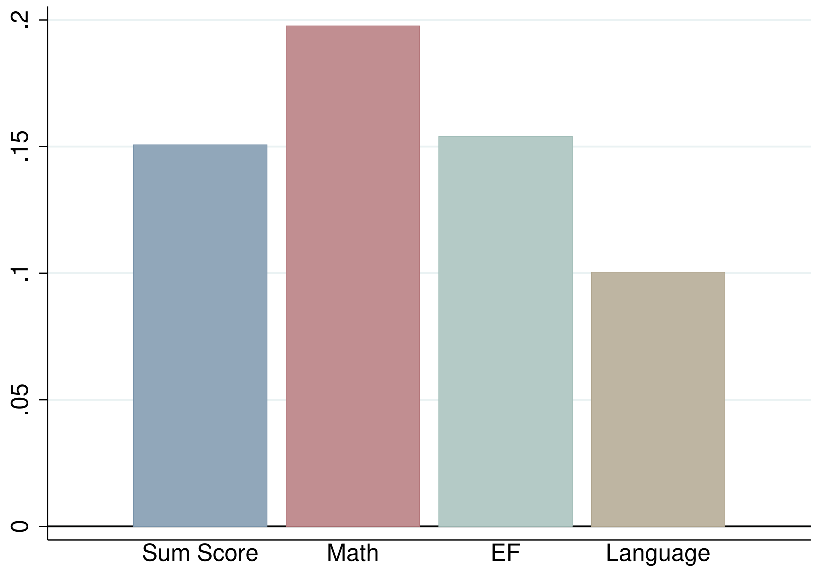

Fig. 1. Gender skill gap prior to treatment.
Note: The bars represent the difference in average scores between boys and girls on our four school-readiness measures at baseline. Higher values on the *y*-axis indicate a larger gap in favor of girls. Scores are standardized so that values represent difference in skills in standard deviation units.

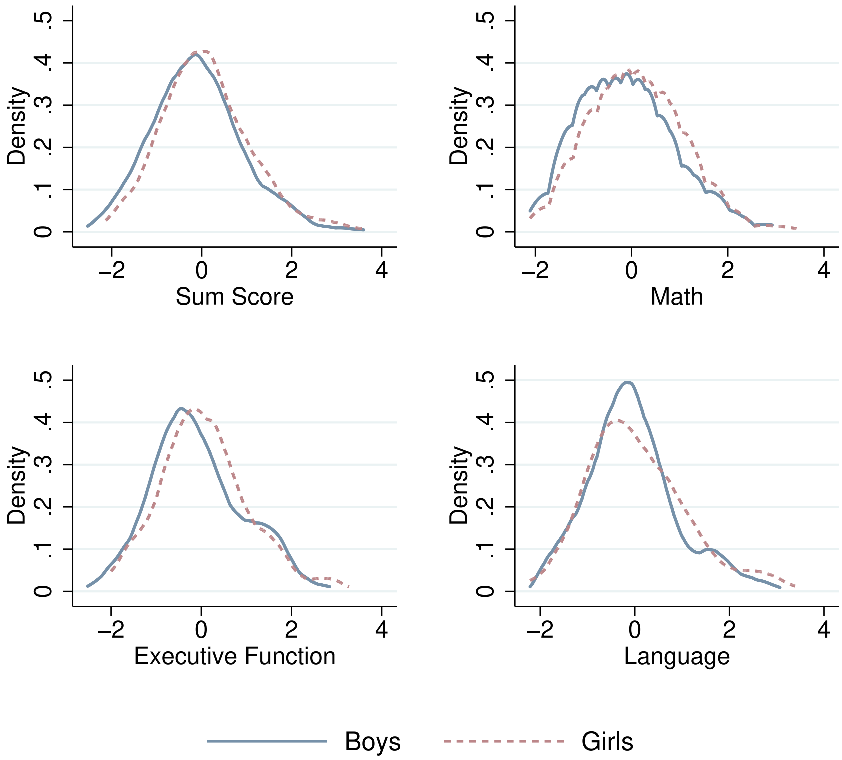

Fig. 2. Distribution of baseline scores by gender.
Note: The figure presents density plots of the baseline test scores on our four school-readiness measures. The scores are standardized so that 0 corresponds to the sample mean on that measure. We use epanechnikov kernels in all plots.

driven in large part by low-achieving boys at the bottom of the distribution. To determine whether boys in the lower parts of the skill distribution are particularly responsive to our intervention, we perform a skill-heterogeneity analysis in Section 6.3.

Table 2
Gender specific treatment effects.

	Post-Intervention (T2)				*Follow-Up (T3)*			
	Sum score	Math	EF	Language	Sum score	Math	EF	Language
Treatment effect	0.191	0.197	0.201	0.059	0.235	0.330	0.137	0.108
Boys	(0.079)	(0.120)	(0.071)	(0.083)	(0.097)	(0.089)	(0.110)	(0.099)
Treatment effect	0.046	0.110	0.041	−0.041	0.025	0.117	−0.028	−0.027
Girls	(0.069)	(0.090)	(0.083)	(0.086)	(0.089)	(0.114)	(0.055)	(0.105)
Difference	−0.145	−0.087	−0.160	−0.010	−0.201	−0.213	−0.165	−0.135
	(0.082)	(0.104)	(0.102)	(0.101)	(0.118)	(0.160)	(0.137)	(0.084)
p-value	0.098	0.420	0.140	0.340	0.097	0.206	0.248	0.132
Wild cluster	0.127	0.431	0.150	0.419	0.109	0.241	0.263	0.138
RI	0.133	0.433	0.150	0.419	0.072	0.093	0.188	0.278
N	652	650	652	648	661	661	660	659
Adj. R^2	0.61	0.44	0.49	0.53	0.55	0.40	0.41	0.52

Note: Each column in each panel presents the regression coefficient of treated (standard error) interacted with gender using ordinary least squares. The coefficient for the interaction gives the total treatment effect on that gender, so that the difference between them represents the marginal effect. The *Difference* panel reports coefficients and errors for tests on significant differences between the gender specific estimates. Below we report three sets of *p*-values computed using clustering on block, the Wild T bootstrap procedure, and randomization inference respectively. For both assessment periods we regress outcome on the treatment–gender interaction, controlling for baseline test scores, gender, birth month, parental characteristics (mother and father's education level, earnings, and indicator for non-Western country of birth), and indicators for late consent and not having participated in the T1 assessment. All regressions are clustered on, and control for, randomization block.

While our data do not allow us to investigate the extent to which the centers *cause* these gender gaps, we find little evidence that business-as-usual mitigates them. Indeed, from T1 to T3 the gender gap in the *sum score* actually increases from 0.29σ to 0.34σ in the control group, although the difference is not statistically significant. This is driven by growing gender gaps in executive functioning and language, while the mathematics gap decreases, particularly once the children start formal schooling.

6.2. Main results

We report the main results from estimating Eq. (1) in Table 2. In the leftmost panel we find large and positive point estimates for the treatment effect on boys at the T2 assessment, with effects of meaningful magnitude in mathematics and executive functions in particular. For the *sum score* measure we estimate that the intervention improved boys' overall school readiness by 0.19σ. For an intuitive comparison, this effect size is equivalent to 4 months of development for boys in the control group. For girls, we find much smaller and statistically insignificant estimates. This pattern of large discrepancies is found consistently across all our outcome measures. While the point estimates are positive for the most part, we do not find strong evidence that girls in the treated group score better than the girls in the control group post-intervention. At most, our results suggest that it is unlikely that there were large negative effects for the girls. For example, the lower bound of the 90% confidence interval the sum score measure is -0.067σ, suggesting that if the true effect on girls is in fact negative, the effect is likely very small.[11] This pattern of large discrepancies between the genders in the estimated effect is found consistently across all our outcome measures.

In row 3 we report the results of testing for significant differences between the two gender-specific estimates. While the difference in point estimates for all measures is substantial – ranging from 0.087σ to 0.16σ – we often lack the precision necessary to attain significance at conventional levels for outcomes other than the *sum score* measure.[12] Even so, we argue that the meaningful and consistent differences found between the estimates strongly suggest that boys are the primary beneficiaries of our intervention, and also that our results provide moderately strong evidence that the difference is not zero (Romer, 2020).

In the rightmost panel we find similar patterns for treatment effects at the one-year follow-up. In fact, T3 estimates exceed T2 ones. For boys, we find large, positive effects, while for girls we find small effects that are both statistically and substantively insignificant. The gender difference in the *sum score* measure is 0.20σ. The largest effect is on boys' mathematical skills, for which we estimate a treatment effect of 0.33σ, which also reflects a substantial increase from the T2 assessment.

6.3. Treatment effect heterogeneity by baseline skill

Policy discussions often have particular focus on low-achieving boys, who are deemed to be at greatest risk and the most vulnerable (see, e.g., Chetty et al., 2016, and Autor et al., 2020). Against this backdrop, we investigate whether treatment effects are heterogeneous to baseline skill level. To do so, we plot the results from estimating (2) in Fig. 3.[13] Presented in the figure are five

[11] Moreover, Romer (2020) points out that in the case of the 90% confidence interval the point estimate is roughly four times more likely as the value at the boundary.

[12] We test different approaches and specifications for combining the skill measures in a sum score in Appendix E.

[13] Similar figures for the skill-specific measures are available in Appendix B.

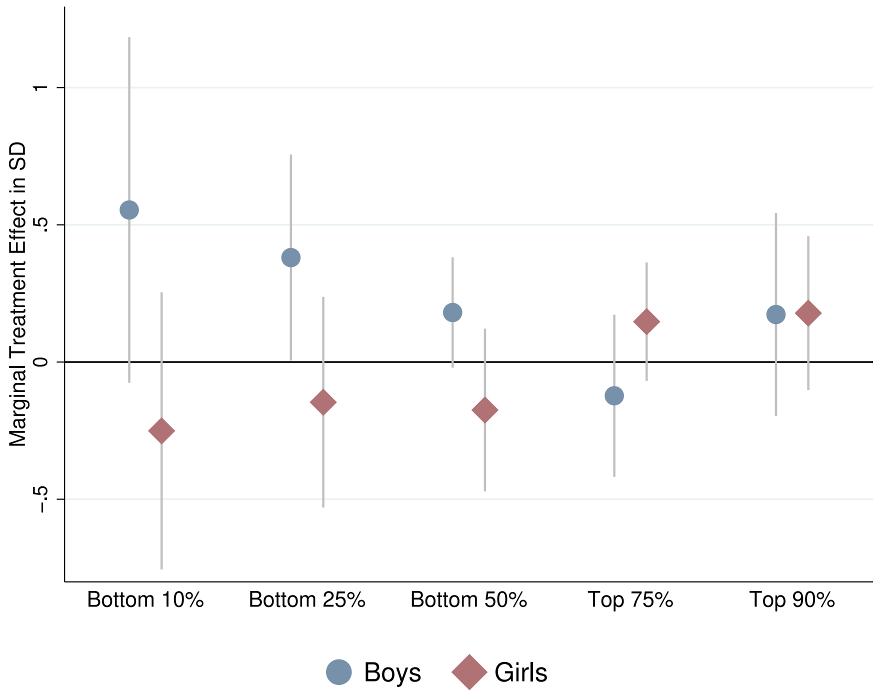

Fig. 3. Heterogeneous effects by baseline skills.
Note: In this figure we plot estimates of the marginal treatment effect of placing in a specific segment of the *Sum Score* distribution for your gender at baseline. Each of the five circle/diamond pairs represents a separate regression, where the plot gives the estimated coefficient of the three-way gender×treatment×baseline score segment interaction. For example, the left-most circle gives the marginal treatment effect for boys scoring in the lowest 10% of boys at baseline, relative to the treatment effect for all other boys. The gray lines indicate 90% confidence bands.

pairs of plots representing results for boys and girls, respectively, stemming from separate regressions. As an example, the leftmost circle (diamond) represents the marginal treatment effect for boys (girls) scoring in the lowest 10 percent of boys (girls) at baseline, relative to the effect for all other treated boys (girls). Moving along the horizontal axis implies moving upward in the baseline skill distribution.

We see a clear downward trend in the marginal effects for boys. Starting with those scoring in the lowest 10 percent at baseline, we find a marginal treatment effect of over 0.5σ, although imprecisely estimated. It should be noted that this marginal effect is in addition to the average treatment effect on treated boys, meaning that the total effect sums up to about 0.7σ. The marginal effect then declines rapidly as we move to higher-achieving boys, converging to zero for the top of the distribution. This pattern suggests that the majority of the treatment effect is concentrated in the group of initially low-achieving boys.

For girls, we do not find a similar pattern. If anything, we find some indication that the highest-achieving girls may have benefited the most from the intervention. Overall, however, the small (if any) effect on girls seems fairly homogeneous across the distribution.

We also investigated whether we could find similar heterogeneity across socioeconomic status, to check whether the pattern observed in Fig. 3 could be explained by the low-achieving population overlapping with the disadvantaged population. However, we did not find any indication that our treatment effect varies across socioeconomic status, neither overall nor within gender. This suggest that the pattern is driven by initial skill level, and not other forms of disadvantage, such as low income or low parental education. The full heterogeneity analysis with regard to socioeconomic background is available in Appendix C.

7. Implications

Overall, we find substantial gender gaps in early-life skills crucial for future learning.[14] This implies that Norwegian girls start school with a significant skill advantage over boys. This difference in school readiness may explain, at least partly, why boys might fall behind as they progress through the education system.

The structured curriculum investigated in the present study boosts school readiness on average, and we find that these positive effects appear to be driven by boys. Moreover, we present suggestive evidence that the lowest-achieving boys benefit the most from this curriculum. We note that our estimates in several cases are imprecise, and therefore urge for caution when interpreting our

[14] In a related paper, Thijssen et al. (2023) show that the boost in executive function stemming from the intervention is particularly crucial as it also bolsters the development of language and mathematical skills once in primary school.

results. Still, the overall results of the intervention suggest that introducing more structured activities could be beneficial for boys in contexts where ECEC practice is primarily centered around free play.

Overall, we find little evidence that the curriculum affected the school readiness of the girls in the treated sample. That is, at the post-intervention assessments, the scores of the girls in the treatment group are statistically indistinguishable from those of the girls in the control group. Yet, the point estimates, although imprecise, are in most cases positive, suggesting that we at least can rule out that the curriculum had a large negative effect on girls. Our results therefore suggest that the improved school readiness among boys resulting from the change in instruction style does not come at the expense of the girls, but rather is closer to a pareto improvement.

We stress, however, that there could also be longer-term effects that our present study is unable to identify. For example, if a policy shift toward a more structured curriculum in preschool greatly improves the developmental trajectory of mathematical skills for boys, but not girls, there could be implications for the gender composition at competitive programs and institutions where math ability is an important determinant for access. In the Norwegian context such concerns may not be warranted, as girls outperform boys in all subjects in school as well as greatly outnumbering them both the rate at which they enroll in higher education and gain access to the most prestigious programs (NOU 2019: 3, 2019). Depending on the trends in gender gaps in education outcomes in other contexts, such concerns might however be more pressing. While uncovering such effects is beyond the scope of the present study, assessing the long-term impact of more structured curricula in preschools will be important for determining the desirability of implementing such policies at scale.

We believe our findings are therefore most relevant for the discussion of how the design of early-childhood curricula and pedagogical practice contribute to ECEC quality and child development. The extant literature has emphasized the need for a better understanding of the process and heterogeneous impacts of early-childhood education. Our study suggests that preschools can mitigate gender gaps in early learning by providing boys with appropriate scaffolding and support to stimulate school readiness. In general, opening further the black box that is ECEC quality, so as to better understand how preschool practice and curricula affect child subgroups, is an important task for future research. To that end, it will be particularly important to replicate and validate the findings from our convenience sample in more diverse and representative contexts.

Data availability

The authors do not have permission to share data.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.euroecorev.2023.104413.

References

Abadie, A., Athey, S., Imbens, G.W., Wooldridge, J.M., 2020. Sampling-based versus design-based uncertainty in regression analysis. Econometrica 88 (1), 265–296.
Anderson, M.L., 2008. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry preschool, and early training projects. J. Amer. Statist. Assoc. 103 (484), 1481–1495.
Athey, S., Imbens, G., 2017. The econometrics of randomized experiments. In: Banerjee, A.V., Duflo, E. (Eds.), Handbook of Economic Field Experiments. In: Handbooks in Economics, vol. 1, Elsevier, Amsterdam, Netherlands, pp. 73–140 (Chapter 3).
Autor, D., Figlio, D.N., Karbownik, K., Roth, J., Wasserman, M., 2020. Males at the tails: How socioeconomic status shapes the gender gap. In: NBER Working Paper 27196. National Bureau of Economic Research.
Autor, D., Wasserman, M., 2013. Wayward sons: The emerging gender gap in education and labor markets. NEXT Series.
Bedard, K., Cho, I., 2010. Early Gender Test Score Gaps Across OECD Countries. Econ. Educ. Rev. 29 (3), 348–363.
Bennett, J., Tayler, C., 2006. Starting Strong II: Early Childhood Education and Care. OECD Publishing, Paris, France.
Berlinski, S., Galiani, S., Manacorda, M., 2008. Giving children a better start: Preschool attendance and school-age profiles. J. Public Econ. 92 (5), 1416–1440.
Bertram, R.M., Blase, K.A., Fixsen, D.L., 2015. Improving programs and outcomes: Implementation frameworks and organization change. Res. Soc. Work. Pract. 25 (4), 477–487.
Blau, F.D., Kahn, L.M., 2007. The gender pay gap. Acad. Manag. Perspect. 21 (1), 7–23.
Brandlistuen, R.E., Flatø, M., Stoltenberg, C., Helland, S.S., Wang, M.V., 2020. Gender gaps in preschool age: A study of behavior, neurodevelopment and pre-academic skills. Scand. J. Public Health 140349482094474.
Cameron, A., Miller, D., 2015. A pactitioner's guide to cluster-robust inference. J. Hum. Resour. 50 (2), 317–372.
Chetty, R., Hendren, N., Lin, F., Majerovitz, J., Scuderi, B., 2016. Childhood environment and gender gaps in adulthood. Amer. Econ. Rev. 106 (5), 282–288.
Clements, D.H., Sarama, J., 2011. Early childhood mathematics intervention. Science 333 (6045), 968–970.
Cornelissen, T., Dustmann, C., Raute, A., Schönberg, U., 2018. Who benefits from universal child care? Estimating marginal returns to early child care attendance. J. Political Econ. 126 (6), 2356–2409.
Cunha, F., Heckman, J., 2007. The technology of skill formation. Amer. Econ. Rev. 97 (2), 31–47.
Davidson, M.C., Amso, D., Anderson, L.C., Diamond, A., 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, Advances in Developmental Cognitive Neuroscience.
Deming, D., Dynarski, S., 2008. The lengthening of childhood. J. Econ. Perspect. 22 (3), 71–92.
Diamond, A., Barnett, W.S., Thomas, J., Munro, S., 2007. Preschool program improves cognitive control. Science 318 (5855), 1387–1388.
Dillon, M.R., Kannan, H., Dean, J.T., Spelke, E.S., Duflo, E., 2017. Cognitive science in the field: A preschool intervention durably enhances intuitive but not formal mathematics. Science 357 (6346), 47–55.
DiPrete, T.A., Buchmann, C., 2013. Rise of Women, the: The Growing Gender Gap in Education and What It Means for American Schools. Russell Sage Foundation.
DiPrete, T.A., Jennings, J.L., 2012. Social and behavioral skills and the gender gap in early educational achievement. Soc. Sci. Res. 41 (1), 1–15.
Domond, P., Orri, M., Algan, Y., Findlay, L., Kohen, D., Vitaro, F., Tremblay, R.E., Côté, S.M., 2020. Child care attendance and educational and economic outcomes in adulthood. Pediatrics 146 (1).

Drange, N., Rønning, M., 2020. Child care center quality and early child development. J. Public Econ. 188, 104204.
Duncan, G.J., Magnuson, K., 2013. Investing in preschool programs. J. Econ. Perspect. 27 (2), 109–132.
Early, D.M., Iruka, I.U., Ritchie, S., Barbarin, O.A., Winn, D.-M.C., Crawford, G.M., Frome, P.M., Clifford, R.M., Burchinal, M., Howes, C., Bryant, D.M., Pianta, R.C., 2010. How do pre-kindergarteners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms. Early Child. Res. Q. 25 (2), 177–193.
Engel, A., Barnett, W.S., Anders, Y., Taguma, M., 2015. Early Childhood Education and Care Policy Review. OECD Publishing, Paris, France.
Felfe, C., Lalive, R., 2018. Does early child care affect children's development? J. Public Econ. 159, 33–53.
Felfe, C., Nollenberger, N., Rodríguez-Planas, N., 2015. Can't buy mommy's Love? Universal childcare and children's long-term cognitive development. J. Popul. Econ. 28 (2), 393–422.
Fessler, P., Schneebaum, A., 2019. The educational and labor market returns to preschool attendance in Austria. Appl. Econ. 51 (32), 3531–3550.
Fortin, N.M., Oreopoulos, P., Phipps, S., 2015. Leaving boys behind gender disparities in high academic achievement. J. Hum. Resour. 50 (3), 549–579.
Goldin, C., Katz, L.F., Kuziemko, I., 2006. The homecoming of American college women: The reversal of the college gender gap. J. Econ. Perspect. 20 (4), 133–156.
Goodman, A., Sianesi, B., 2005. Early education and children's outcomes: How long do the impacts last? Fiscal Stud. 26 (4), 513–548.
Gørtz, M., Rye Johansen, E., Simonsen, M., 2018. Academic achievement and the gender composition of preschool staff. Labour Econ. 55, 241–258.
Gray-Lobe, G., Pathak, P.A., Walters, C.R., 2023. The long-term effects of universal preschool in Boston. Quart. J. Econom. 138, 363–411.
Havnes, T., Mogstad, M., 2015. Is universal child care leveling the playing field? J. Public Econ. 127, 100–114, The Nordic Model.
Heckman, J.J., Moon, S.H., Pinto, R., Savelyev, P.A., Yavitz, A., 2010. The rate of return to the HighScope perry preschool program. J. Public Econ. 94 (1), 114–128.
Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., Barbarin, O., 2008. Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. Early Child. Res. Q. 23 (1), 27–50.
Jenkins, J.M., Duncan, G.J., Auger, A., Bitler, M., Domina, T., Burchinal, M., 2018. Boosting school readiness: Should preschool teachers target skills or the whole child? Econ. Educ. Rev. 65, 107–125.
Karlsen, L., Lekhal, R., 2019. Practitioner involvement and support in children's learning during free play in two Norwegian Kindergartens. J. Early Child. Res. 17 (3), 233–246.
Kenney-Benson, G.A., Pomerantz, E.M., Ryan, A.M., Patrick, H., 2006. Sex differences in math performance: The role of children's approach to schoolwork. Dev. Psychol. 42 (1), 11.
Lenes, R., Gonzales, C.R., Størksen, I., McClelland, M.M., 2020. Children's self-regulation in Norway and the United States: The role of mother's education and child gender across cultural contexts. Front. Psychol. 11, 2563.
Lundberg, I., Larsman, P., Strid, A., 2012. Development of phonological awareness during the preschool year: The influence of gender and socio-economic status. Read. Writ. 25, 305–320.
Magnuson, K., Duncan, G.J., 2016. Can Early Childhood Interventions Decrease Inequality of Economic Opportunity? RSF: Russell Sage Found. J. Soc. Sci. 2 (2), 123–141.
Magnuson, K.A., Kelchen, R., Duncan, G.J., Schindler, H.S., Shager, H., Yoshikawa, H., 2016. Do the effects of early childhood education programs differ by gender? A meta-analysis. Early Child. Res. Q. 36, 521–536.
Mashburn, A.J., Pianta, R.C., Hamre, B.K., Downer, J.T., Barbarin, O.A., Bryant, D., Burchinal, M., Early, D.M., Howes, C., 2008. Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. Child. Dev. 79 (3), 732–749.
Matthews, J.S., Ponitz, C.C., Morrison, F.J., 2009. Early gender differences in self-regulation and academic achievement. J. Educ. Psychol. 101 (3), 689.
McClelland, M.M., Cameron, C.E., Duncan, R., Bowles, R.P., Acock, A.C., Miao, A., Pratt, M.E., 2014. Predictors of early growth in academic achievement: The head-toes-knees-shoulders task. Front. Psychol. 5, 599.
Melhuish, E.C., 2011. Preschool matters. Science 333 (6040), 299–300.
Norwegian Directorate for Education and Training, 2019. Utdanningsspeilet [The Education Mirror]. Oslo, Norway.
NOU 2019: 3, 2019. Nye sjanser – bedre læring: Kjønnsforskjeller i skoleprestasjoner og utdanningsløp. Oslo: The Ministry of Knowledge.
OECD, 2015. The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence. OECD Publishing, Paris, France.
Phillips, D., Lipsey, M., Dodge, K., Haskins, R., Bassok, D., Burchinal, M., Duncan, G., Dynarski, M., Magnuson, K., Weiland, C., 2017. Puzzling it out: The current state of scientific knowledge on Pre-Kindergarten effects—A consensus statement. Issues Pre-Kindergarten Programs Policy 19–30.
Rege, M., McClelland, M., Kalil, A., Solli, I.F., Størksen, I., 2017. Leveling the playing field: An intervention to promote school readiness and human potential. AEA RCT Regist.
Rege, M., Solli, I.F., Størksen, I., Votruba, M., 2018. Variation in center quality in a universal publicly subsidized and regulated childcare system. Labour Econ. 55, 230–240.
Rege, M., Størksen, I., Solli, I.F., Kalil, A., McClelland, M., ten Braak, D., Lenes, R., Lunde, S., Breive, S., Carlsen, M., Erfjord, I., Hundeland, P.S., 2021. The effects of a structured curriculum on preschool effectiveness: A field experiment. J. Hum. Resour. 0220-10749R3.
Robinson, J.P., Lubienski, S.T., 2011. The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings. Am. Educ. Res. J. 48 (2), 268–302.
Romer, D., 2020. In praise of confidence intervals. AEA Pap. Proc. 110, 55–60.
Ruble, D.N., Martin, C.L., Berenbaum, S.A., 2006. Gender development. In: Damon, W., Lerner, R.M., Eisenberg, N. (Eds.), Handbook of Child Psychology: Social, Emotional, and Personality Development, Vol. 3, sixth ed. John Wiley & Sons, Inc., Hoboken, NJ, pp. 858–932 (Chapter 14).
Schmitt, S.A., McClelland, M.M., Tominey, S.L., Acock, A.C., 2015. Strengthening school readiness for head start children: Evaluation of a self-regulation intervention. Early Child. Res. Q. 30, 20–31.
Spelke, E.S., 2005. Sex differences in intrinsic aptitude for mathematics and science?: A critical review. Am. Psychol. 60 (9), 950–958.
Stangeland, E.B., Lundetræ, K., Reikerås, E., 2018. Gender differences in toddlers' language and participation in language activities in Norwegian ECEC institutions. Eur. Early Child. Educ. Res. J. 26 (3), 375–392.
Størksen, I., Ellingsen, I.T., Tvedt, M.S., Idsøe, E.M., 2013. Norsk vokabulartest (NVT) for barn i overgangen mellom barnehage og skole: Psykometrisk vurdering av en nettbrettbasert test. Spesialpedagogikks forskningsdel 4 (13), 40–54.
Størksen, I., Ellingsen, I.T., Wanless, S.B., McClelland, M.M., 2015. The influence of parental socioeconomic background and gender on self-regulation among 5-year-old children in Norway. Early Educ. Dev. 26 (5–6), 663–684.
Størksen, I., Ertesvåg, S.K., Rege, M., 2021. Implementing implementation science in a randomized controlled trial in Norwegian early childhood education and care. Int. J. Educ. Res. 108, 101782.
Størksen, I., ten Braak, D., Breive, S., Lenes, R., Lunde, S., Carlsen, M., Erfjord, I., Hundeland, P.S., Rege, M., 2018. Playful Learning - A Research Based Preschool Curriculum from the Agder Project. GAN Aschehoug.
Sumsion, J., 2005. Male teachers in early childhood education: Issues and case study. Early Child. Res. Q. 20 (1), 109–123.
Synodi, E., 2010. Play in the kindergarten: The case of Norway, Sweden, New Zealand and Japan. Int. J. Early Years Educ. 18 (3), 185–200.
ten Braak, D., Størksen, I., 2021. Psychometric properties of the ani banani math test. Eur. J. Dev. Psychol. 18, 610–628.
Thijssen, M.W.P., Rege, M., Solli, I.F., Størksen, I., 2023. Cross-productivities of executive functions: Evidence from a field experiment. Working Paper.

Tonyan, H.A., Howes, C., 2003. Exploring patterns in time children spend in a variety of child care activities: Associations with environmental quality, ethnicity, and gender. Early Child. Res. Q. 18 (1), 121–142.

Vincent-Lancrin, S., 2008. The reversal of gender inequalities in higher educationn: An on-going trend. In: OECD (Ed.), Higher Education To 2030: Demography, Vol. 1. OECD Publishing, Paris, France.

Wechsler, D., 2003. Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV). The Psychological Corporation, San Antonio, TX.

Weiland, C., Yoshikawa, H., 2013. Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. Child. Dev. 84 (6), 2112–2130.

White, L.A., Prentice, S., Perlman, M., 2015. The evidence base for early childhood education and care programme investment: What we know, what we don't know. Evid. Policy 11 (4), 529–546.