Do Increases in National Level Preschool Enrollment Increase Student Achievement? Evidence from International Assessments

Rolf Strietholt

TU Dortmund University, Germany

University of Gothenburg, Sweden

Nina Hogrebe

University of Münster, Germany

Henrik Daae Zachrisson

University of Oslo, Norway

Dr. Rolf Strietholt , TU Dortmund University, Center for Research on Education and School Development, Vogelpothsweg 78, 44227 Dortmund, Germany, Email: rolf.strietholt@tu-dortmund.de, Office: 0049.231.755.5510

Dr. Nina Hogrebe, HAW Hamburg, Faculty of Business & Social Sciences, 20099 Hamburg. Germany, Email: nina.hogrebe@haw-hamburg.de, Office: 0049.40.428757083

Dr. Henrik Daae Zachrisson, Department of Special Needs Education, University of Oslo, PB 1140, Blindern, 0318 Oslo, Norway, Email: h.d.zachrisson@isp.uio.no, Office: 0047.22858066

Correspondence concerning this article should be addressed to Rolf Strietholt. The manuscript has not been published elsewhere and has not been submitted for publication elsewhere.

Abstract

The main purpose of the study is to analyze whether globally observed trends towards preschool expansion have impacted student achievement in primary and secondary school. We use data from multiple study cycles of two international large-scale assessments that have a longitudinal component at the country level—PIRLS and PISA—and combine these data with a country-level measure of preschool enrollment rates as the main explanatory variable. Employing a multilevel regression with fixed effects for countries and years, we have found that changes in preschool enrollment are unrelated to changes in average student achievement. Even after controlling for covariates on the individual and country levels, we do not find any support for the policy expectation that expanding preschool enrollment per se leads to better student achievement on the country level.

Keywords: academic achievement, comparative analysis, early childhood policies, preschool enrollment, international panel data

Do Increases in National-Level Preschool Enrollment Increase Student Achievement? Evidence from International Assessments

1. Introduction

An integral element in UNESCO's post-2015 development agenda is to have at least one year of pre-primary education[i] for all (UNESCO, 2014). This reflects growing attention towards early childhood education from organizations like UNESCO (2006; see also Marope & Kaga, 2015), the OECD (2001a, 2006), and the World Bank (Denboba et al., 2014; World Bank, 2011, 2016), all of which cite their promise for improving educational gains and ultimately children's life chances. This attention has been accompanied by increased enrollment in early childhood education and care (ECEC) across the globe. Of course, there are still large differences between countries, with 2014 enrollment rates of 17%, 44%, and 83% in low, middle, and high-income countries, respectively (UNESCO Institute for Statistics, n.d.). And even among high income countries, rates vary significantly. For example, enrollment rates for four-year-olds in the OECD range from 32% in Turkey to 100% in France; the United States, for comparison, has an enrollment rate of 68%, which is below the OECD average of 86% (OECD, 2016). Despite these differences, however, there is an evident expansion of preschool education underway worldwide, with a total increase in pre-primary education enrollment rates from 30% to 44% between 1999 and 2014 (UNESCO Institute for Statistics, n.d.).

From a policy evaluation perspective, it is important to examine the outcomes of such expansion efforts. In this paper, therefore, we tested whether within-country changes in the proportion of children attending preschool was associated with medium-term within-country changes in academic achievement using a country panel fixed-effects approach (Hanushek, Link, & Woessmann, 2013). We created a panel using data on national-level preschool coverage over time from UNESCO Institute for Statistics and achievement measures (of reading) from multiple cycles of two large-scale international assessments— Progress in International Reading Literacy Study (PIRLS) (Mullis, Martin, Foy, & Drucker, 2012) and Program for International Student Assessment (PISA)(OECD, 2014). Our main finding was that we see no association between changes in preschool coverage and achievement scores in primary and secondary school within countries over time. This was also the case for students from lower socioeconomic backgrounds. We conditioned our models on both national- and individual-level covariates. We found our results to be robust across different re-specifications of the sample (e.g., to include OECD countries only), across alternative outcomes from PISA (math and science achievement). Finally, using individual level, parent-reported data from surveys on preschool suggests that the national-level estimates of coverage were fairly accurate. We conclude that expansion of preschool in and of itself does not improve national-level test scores through elementary and middle school age.

2. Background

2.1 The Inspiration: Effects of Preschool Programs

The enthusiasm among both policy makers and researchers for early education and preschool programs builds, to a great extent, on a set of theoretical notions about children's development and consequences thereof. The first is that brain growth, and consequently child development, is at its most rapid, but also most malleable, during the early years (e.g., Knudsen et al., 2006; Shonkoff & Phillips, 2000; Engel et al., 2011). Hence, the early years is considered a time window where children are the most receptive to contextual interventions, as early development lays the foundation for future development. Another guiding theoretical notion has been one of compensation. Children growing up in disadvantage are consequently at risk for less nurturing home environments than their more affluent peers (due to higher levels of stress and less investments in the home (e.g., Dearing, 2007). An enriching environment in an early education setting has been hypothesized to compensate for such lack of nurturing environment (Leseman & Slot, 2014). These two notions form the theoretical foundation for economic theories and analyses. Economic returns to human capital formation are greater when investments are made through educational programs in the early years, as opposed to later (e.g., Heckman, 2006).

While many policy initiatives to extend and improve preschool programs come from international organizations (OECD, World Bank, UNICEF, EU) and are directed towards national policy making of countries worldwide, the overwhelming majority of empirical evidence fueling these initiatives comes from small program evaluations on a local or regional levels, yet in some cases also state- or country-wide evaluations.

Meta-analyses of small-scale randomized trials of early childhood interventions in the United States and internationally have largely justified this enthusiasm by providing evidence for positive effects (Camilli, Vargas, Ryan, & Barnett, 2010; Duncan & Magnuson, 2013; McCoy et al., 2017; Nores & Barnett, 2010). Yet, there are (at least) three lingering concerns as to whether this enthusiasm might be overly optimistic. These concerns relate to the fadeout of effects, scale up of programs, and the yet fairly strong US-centricity of the evidence as well as a lack of high-quality international comparisons.

The question of whether initial preschool effects persist as students continue through school, especially with regard to cognitive and achievement outcomes, was recently addressed in a meta-analysis. Using a database of randomized studies in the US, Bailey et al. found diminishing longer-term effects, despite end of program gains (Bailey, Duncan, Odgers, & Yu, 2017).

Moreover, as pointed out by Barnett (2011), the policy relevance of such interventions is not only limited by possible fade-out effects but also by the challenges of taking successful model interventions to a larger scale. Some evaluations of program scale up and state-wide programs in the US have been promising. For example, in Oklahoma and Georgia, positive impacts on math scores were still evident in eighth grade, albeit to a lesser extent than at earlier ages (Cascio & Schanzenbach, 2013). In North Carolina, the scale up two flagship early childhood programs—*Smart Start* for all children from birth to age four and the pre-kindergarten program *More at Four* targeted at high-risk four-year-olds—led to higher math and reading test scores, reductions in special education placement rates, and better grade retention rates through grades three to five for all children living in a county regardless of their actual participation in any of the programs, indicating spill-over effects (Dodge, Bai, Ladd, & Muschkin, 2016). Yet, some more recent studies of large-scale preschool programs in the US have also shown fade-out of initially promising program effects (Lipsey, Farran, & Durkin, 2018; Weiland et al., 2019), consistent with the Bailey et al. (2017) meta-analysis.

Although a wealth of evidence from preschool evaluations stems from the US, there is some evidence from other countries on longer-term outcomes into school age. This evidence base is growing and includes studies with positive effects. The expansion of preschool to nearly full coverage for three- and four-year-olds in France in the 1960s and 1970s resulted in fewer grade repetitions as well as higher test scores and high-school graduation rates (Dumas & Lefranc, 2012). In the 1990s, Spain expanded publicly subsidized full-time high-quality childcare for three-year-olds across the country at different speeds. The expansion of preschool led to an improvement of PISA test scores in the regions where the reform was implemented first compared to the regions where the reform was implemented later (Felfe, Nollenberger, &

Rodríguez-Planas, 2015). Such effects of scale-ups are not restricted to Western, high-income countries; in Latin America, large scale-ups resulted in higher school enrollment and grade completion in Uruguay (Berlinski, Galiani, & Manacorda, 2008) and in better math and language test scores in Argentina (Berlinski, Galiani, & Gertler, 2009). In Indonesia, a large study of combined playgroup and kindergarten programs demonstrated positive effects on test scores in early primary school (Nakajima, et al., 2019).

Importantly, in some studies, mixed effects were found for particular subgroups. For example, Leuven, Lindahl, Oosterbeek, and Webbink (2010) reported that expanding enrollment opportunities by one month in the Netherlands increased performance among disadvantaged students on both the language and arithmetic test; yet the authors did not find any impact on the whole sample or on non-disadvantaged groups. Similar effects that were restricted to disadvantaged children (or children with immigrant backgrounds) have been reported in Sweden (Fredriksson, Hall, Johansson, & Johansson, 2010), the United States (Fitzpatrick, 2008), and Germany (Felfe & Lalive, 2012).

Yet, there are well-designed studies failing to find such positive medium-term effects. In Quebec, Canada, a scale-up of subsidized child care even led to a decrease in early language skills as well as in socio-emotional and motor development (Baker, Gruber, & Milligan, 2008). While longer-term negative consequences were evident for socio-emotional development, there were no long-term consequences (positive or negative) for test scores (Baker, Gruber, & Milligan, 2019). Similarly, the increase in preschool enrollment for children aged one to five in Sweden between 1967 and 1982 was not associated with any main effects on language and inductive skills at age 13 or with improvements in long-term educational attainment (Fredriksson, et al. 2010).

In sum, evidence for effects of preschool lasting into school age is mixed, especially with regard to larger scale or national-level programs. This implies that effects of preschool scale-ups might differ depending on what systems were previously in place, what alternative options were available, and who benefitted most from a given expansion. In some countries, like Germany, children from privileged families are more likely to be enrolled in universal childcare if only few slots are available; here, expansion efforts have created more care options for disadvantaged children (Felfe & Lalive, 2012). In other contexts, such as the USA, targeted and universal public programs coexist with private preschools; hence, scaling up publicly funded universal programs might prompt more privileged families to substitute private for public programs (Cascio & Schanzenbach, 2013) as well as crowding out enrollment from targeted programs for disadvantaged children (Cascio, 2009a). There are also countries where increases in preschool enrollment did not increase the proportion of disadvantaged children in early education programs or only did so marginally (e.g., Sibley, Dearing, Toppelberg, Mykletun, & Zachrisson, 2015) or where the eligibility criteria changed from favoring disadvantaged children to favoring rather privileged families (e.g., Sweden, Fredriksson et al., 2010). Assuming that disadvantaged children drive the effects of preschool expansion, it would not be surprising to find no major effects when non-disadvantaged children are primarily affected by expansion efforts. Some authors have therefore suggested that policies should be mainly focused on disadvantaged children (Fitzpatrick, 2008; Magnuson, Ruhm, & Waldfogel, 2007a). There is some evidence that if scale-ups of such targeted programs reach a certain level, population-level outcomes can improve as a whole, possibly by exerting an impact not only on program participants but also on their peers through spillover effects (Artz & Welsch, 2016; Dodge et al., 2016).

2.2 International Comparisons

While evaluations of local or national ECEC expansions are informative about effects of particular programs and policies, and meta-analyses are useful summaries of these findings, they differ in various ways in terms of the design, the type of outcome, and the length of follow-up. To inform international policy decisions and initiatives, they must therefore be supplemented with studies that adopt a more overarching, cross-national perspective. So far, however, only a few studies have compared associations between ECEC attendance and outcomes across countries. Moreover, studies that have used such an approach have generated mixed results. While some reported a strong relationship between participation in pre-primary education and student achievement in primary and secondary school (Burger, 2016; Mullis, Martin, Foy, & Drucker, 2012), the results of other studies are more ambiguous, especially regarding effects on disadvantaged children. The positive effects in some countries contrast with the fadeout effects found elsewhere or no effects at all.

For example, a comparison of Denmark and the USA found that enrollment in center-based early childhood education at age three was associated with higher cognitive scores at age eleven, with the larger effects for the lowest-income children in Denmark; by contrast, initially visible beneficial impacts faded out in the United States, especially for disadvantaged children (Esping-Andersen et al., 2012). Hogrebe and Strietholt (2016) used PIRLS 2011 data to investigate the impact of disadvantaged children's nonparticipation in preschool in eight countries and found that efforts to include these children in early childhood education programs would not have made any differences in their reading skills in school. Similarly, Burger (2016) did not find any linear effect of the preschool enrollment rate on the gap relating to socio-economic status using PISA 2012 data from 31 European countries. Using TIMSS data from 1995 and 1999, Schütz, Ursprung, & Woessmann (2008) tested whether preschool enrollment reduced the association between family background and later test scores. They found an inverted U-shaped association, indicating that when enrollment rates were lower than 60%, there was an increasing association between family background (measured as books at home) and test scores. Beyond this point, increases in enrollment were associated with lower disparities in test scores due to family background. The authors interpreted this finding to mean that national enrollment needs to reach levels where a substantial proportion of children from disadvantaged backgrounds are covered before the effect is evident at national level provided further evidence that *high* enrollment rates reduced social inequality (Schlicht, Stadelmann-Steffen, & Freitag, 2010). They compared countries where either less or more than 75% of the children were enrolled in preschool using PISA 2006 data from 25 European countries and observed smaller achievement gaps in countries with high enrollment rates.

All the findings from these cross-country comparisons, however, still suffer from methodological limitations. For instance, they are conditional on idiosyncrasies of the datasets included. In the previously mentioned two-country comparison by Esping-Andersen et al. (2012), the outcome variables in the United States data were Item-Response-Theory (IRT) scores from tests of the Early Childhood Longitudinal Study (ECLS-K); the Danish data contained sum scores from another test (see also footnote 5 in Esping-Andersen et al. 2012). For researchers who wish to circumvent technical issues related to the comparability of measures or samples, international large-scale assessments provide an alternative approach that includes standardized measures and comparable samples. However, previous research using international assessments has been subject to other limitations. Importantly, some comparative studies using these data replicated the analyses for different countries (e.g., Hogrebe & Strietholt, 2016; OECD, 2013). This approach uses the within-country variation in preschool participation to estimate the effect on performance separately for each country. Yet, even though this design has its merits, it is subject to serious selection effects and fails to exploit the full international variance in the treatment variable in pooled international data, which is typically much larger than the variance in a single country (Hanushek & Woessmann, 2011). Studies that make use of pooled international data have exploited this variation but only simple regression estimates have been reported to date (e.g., Burger, 2016; Mullis, Martin, Foy, & Drucker, 2012; Schlicht et al., 2010). Additionally, such cross-sectional associations are probably biased by unobserved between-country heterogeneity because educational systems differ not only in enrollment rates in preprimary education but also in terms of other important institutional features.

2.3 The Present Study

Altogether research findings are still inconsistent, international policy making is unequivocally promoting preschool expansion worldwide. It is not only due to the lack of adequate international comparative research but also against the background of the existing evidence base that the international organizations' generalist argument might be called into question: "[C]hild care expansion […] may in some cases be considered as promising policy instruments but in others a form of costly, ineffective (or even counterproductive) public policy" (van Huizen, & Plantenga, 2015, p. 2). It thus remains to be seen whether these globally observed trends towards preschool expansion have impacted country-level school achievement.

Has preschool enrollment had an effect on later student achievement in school internationally? This paper uses longitudinal achievement data from two large-scale assessments and combines this data with information on preschool enrollment to shed light on this issue. We relate changes in preschool coverage to changes in student achievement within countries over time in order to estimate the average effect of this within-country association across a large number of countries. Although we use data from almost two million individuals at the student level, the longitudinal component of our analyses is at the country level.

3. Method

3.1 Sample

We use data from multiple study cycles of two international large-scale assessments, PIRLS (Progress in International Reading Literacy Study; Mullis, Martin, Foy, & Drucker, 2012) and PISA (Program for International Student Assessment; OECD, 2014)[ii]. PIRLS tests nationally representative samples of students at the end of primary school (grade four in most countries), regardless of age and is conducted by the International Association for the Evaluation of Educational Achievement (IEA). PISA tests nationally representative samples of 15-year-olds who are enrolled in school, regardless of grade, and is conducted by the Organization for Economic Cooperation and Development (OECD). Both studies are repeated every third to fifth year, and researchers can study trends in achievement for countries that participate repeatedly because the repeated studies draw the samples from the same populations and the achievement tests are linked on the same scale. We used PIRLS data from the study cycles in 2001, 2006, and 2011, and PISA data from 2000, 2003, 2006, 2009, and 2012. It should be noted that the design of the assessments is longitudinal at a macrolevel but not at microlevel.

We supplemented the test data with additional data. Background questionnaires were administered along with the assessment material to collect further information on students and their homes; we used these data as covariates. Furthermore, we combined the student-level data with country-level data taken from the UNESCO Institute for Statistics and the World Bank. These extensive databases provided internationally comparable statistics that included our main explanatory variable—the preschool enrollment rate—and some country-level covariates.

We restricted our sample to countries with data on achievement and preschool enrollment data for at least two time points because our approach relied on the longitudinal variation within countries. Combining the available data led to a PIRLS dataset with 389,641 students from 80 country-by-year observations that referred to 32 countries and a PISA dataset that contained 1,646,345 students from 234 country-by-year observations from 59 countries (Appendix A).

All statistics reported below use student weights to generalize our analyses to the student population within each country-by-year observation. As the sample sizes varied across countries and over time, we divided the weight of each student by the sum of the weights within the same country-by-year observation. This weight was used so that the contribution of each country-by-year observation would be the same, regardless of the size of the population.

3.2 Instruments

3.2.1 *Student achievement*

The main outcome variables were the reading achievement scores in PIRLS and PISA. The assessment framework of PIRLS defines reading literacy as “the ability to understand and use those written language forms required by society and/or valued by the individual. Young readers can construct meaning from a variety of texts. They read to learn, to participate in communities of readers, and for enjoyment” (Campbell, Kelly, Mullis, Martin, & Sainsbury, 2001, p. 3). In a similar vein, reading literacy in PISA is defined “as the ability to understand, use and reflect on written texts in order to achieve one’s goals, to develop one’s knowledge and potential, and to participate effectively in society” (OECD, 2001b, p. 37). The assessment material of both studies contained text passages and corresponding items that the students responded to after reading. Some passages and items from earlier study cycles were integrated in subsequent cycles to establish links over time. Due to these overlaps, scores that refer to the same study can be understood to have a common metric, making it possible to investigate trends over time. The PIRLS reading scores were standardized such that the scale had a mean of 500 and a standard deviation of 100 across all countries participating in 2001. The PISA scale had a mean of 500 and a standard deviation of 100 for the OECD countries participating in 2000. Despite the similarities in how PIRLS and PISA assessed and defined reading literacy, their scores were not comparable.[iii]

Table 1 Descriptive statistics by country

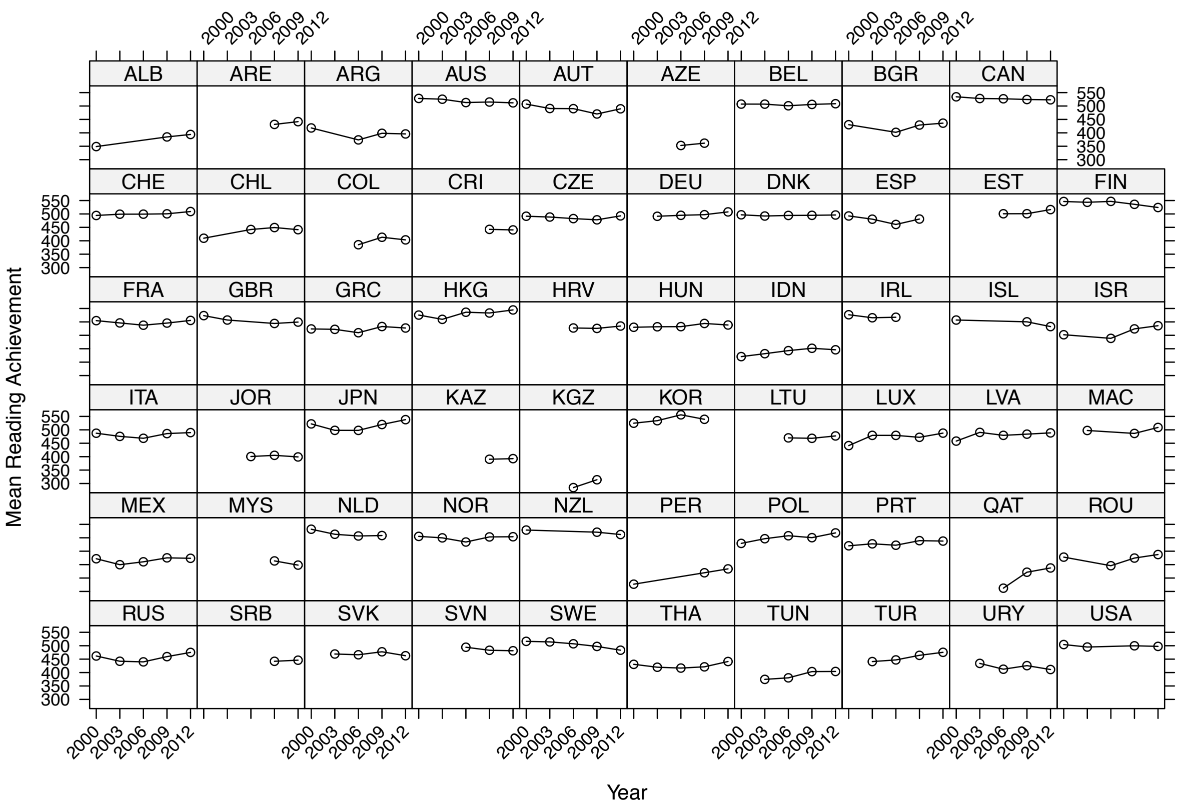
		PIRLS				PISA			
		Preschool Enrollment		Reading Achievement		Preschool Enrollment		Reading Achievement	
		min	max	min	max	min	max	min	max
ALB	Albania					0.393	0.513	348.85	393.96
ARE	United Arab Emirates					0.630	0.651	431.42	441.70
ARG	Argentina					0.512	0.611	373.72	418.25
AUS	Australia					0.704	1.016	511.80	528.28
AUT	Austria	0.806	0.903	528.88	538.30	0.683	0.806	470.28	507.13
AZE	Azerbaijan					0.181	0.183	352.89	361.52
BEL	Belgium					1.029	1.171	500.90	508.62
BGR	Bulgaria	0.644	0.813	531.83	550.50	0.644	0.730	401.93	436.13
CAN	Canada	0.638	0.690	544.15	548.42	0.601	0.646	523.12	534.31
CHE	Switzerland					0.907	0.969	494.37	509.04
CHL	Chile					0.745	0.929	409.56	449.37
COL	Colombia	0.313	0.411	422.43	447.68	0.289	0.398	385.31	413.18
CRI	Costa Rica					0.453	0.604	440.55	442.58
CZE	Czech Republic	0.889	1.161	536.88	545.49	0.884	1.050	478.19	492.89
DEU	Germany	0.856	1.027	539.09	547.59	0.856	0.983	491.36	507.68
DNK	Denmark	0.898	0.952	546.35	553.99	0.816	0.967	492.32	496.87
ESP	Spain					0.717	0.978	460.83	492.55
EST	Estonia					0.674	1.036	500.75	516.29
FIN	Finland					0.331	0.555	524.02	546.87
FRA	France	1.120	1.147	520.00	525.17	1.050	1.133	487.71	505.48
GBR	United Kingdom					0.477	0.796	494.18	523.44
GEO	Georgia	0.401	0.594	470.84	487.76				
GRC	Greece					0.563	0.688	459.71	482.78
HKG	Hong Kong	0.818	0.945	527.87	570.54	0.758	0.855	509.54	544.60
HRV	Croatia					0.373	0.443	475.75	484.57
HUN	Hungary	0.795	0.877	539.27	550.89	0.795	0.845	479.97	494.18
IDN	Indonesia	0.247	0.360	404.74	428.48	0.177	0.247	370.61	401.71
IRN	Iran	0.098	0.515	413.83	457.36				
IRL	Ireland					1.027	1.093	515.48	526.67
ISL	Iceland					0.866	1.036	482.52	506.93
ISR	Israel	0.764	0.949	508.94	540.92	0.795	0.922	438.67	485.80
ITA	Italy	0.964	1.033	540.73	551.47	0.901	0.979	468.52	489.75
JOR	Jordan					0.257	0.320	399.03	405.01
JPN	Japan					0.467	0.852	497.96	538.05
KAZ	Kazakhstan					0.171	0.285	390.41	392.74

KGZ	Kyrgyzstan					0.093	0.100	284.71	314.02
KOR	Korea, Republic of					0.539	0.857	524.75	556.02
KWT	Kuwait	0.808	0.920	330.30	396.47				
LTU	Lithuania	0.373	0.699	528.23	543.39	0.373	0.561	468.44	477.31
LUX	Luxembourg					0.804	0.966	441.25	487.81
LVA	Latvia	0.435	0.584	540.91	544.61	0.322	0.646	458.07	490.56
MAC	Macao					0.859	0.952	486.64	508.95
MAR	Morocco	0.583	0.608	322.58	349.51				
MDA	Moldova	0.447	0.496	491.74	499.88				
MEX	Mexico					0.604	0.704	399.72	425.27
MKD	Macedonia	0.255	0.289	441.59	442.40				
MYS	Malaysia					0.520	0.546	398.20	413.81
NOR	Norway	0.745	0.898	498.01	507.05	0.581	0.778	484.29	505.28
NLD	Netherlands					0.976	1.005	506.75	531.91
NZL	New Zealand	0.864	0.912	531.02	531.72	0.607	0.855	512.19	528.80
PER	Peru					0.301	0.600	327.08	384.15
POL	Poland	0.490	0.577	519.39	525.57	0.432	0.499	479.12	518.19
PRT	Portugal					0.327	0.731	470.15	489.33
QAT	Qatar	0.305	0.372	353.44	424.85	0.237	0.305	312.21	387.50
ROU	Romania	0.515	0.756	489.47	511.71	0.515	0.757	395.93	437.60
RUS	Russian Federation	0.626	0.878	526.16	568.42	0.620	0.825	439.86	475.15
SRB	Serbia					0.527	0.550	442.02	446.13
SVK	Slovakia	0.734	0.949	518.09	535.08	0.733	0.821	462.77	477.44
SVN	Slovenia	0.609	0.797	501.52	530.32	0.588	0.743	481.32	494.41
SWE	Sweden	0.699	0.957	541.67	561.01	0.636	0.771	483.34	516.33
THA	Thailand					0.424	0.936	416.75	441.22
TTO	Trinidad and Tobago	0.614	0.854	435.59	470.85				
TUN	Tunisia	0.618	0.713	539.92	556.37	0.091	0.172	374.62	404.08
TUR	Turkey					0.049	0.069	440.97	475.49
URY	Uruguay					0.419	0.649	411.35	434.15
USA	United States	0.618	0.713	539.92	556.37	0.591	0.643	495.19	504.42

Note. PIRLS and PISA reading achievement data (weighted by sampling probabilities); UNESCO preschool enrollment data.

Table 1 summarizes the lowest and highest average achievement for each country. It shows that countries differed in their performance levels. Most importantly, the table illustrates the change within countries. The mean range between the lowest and highest performance level was 20.0 in PIRLS and 24.5 in PISA; this corresponded to between a fifth and a quarter of the international standard deviation in test scores. Figure 1 visualizes the change within countries for the PISA sample (see Appendix B for PIRLS). Some countries managed to increase their achievement level over time consistently—for example, Luxembourg and Turkey—whereas others, including Iceland and Sweden, experienced a downward trend. Most countries showed no clear trends and some were mostly flat.

Figure 1 Mean reading achievement in PISA, 2000-2012



Note. PISA reading achievement data (weighted by sampling probabilities).

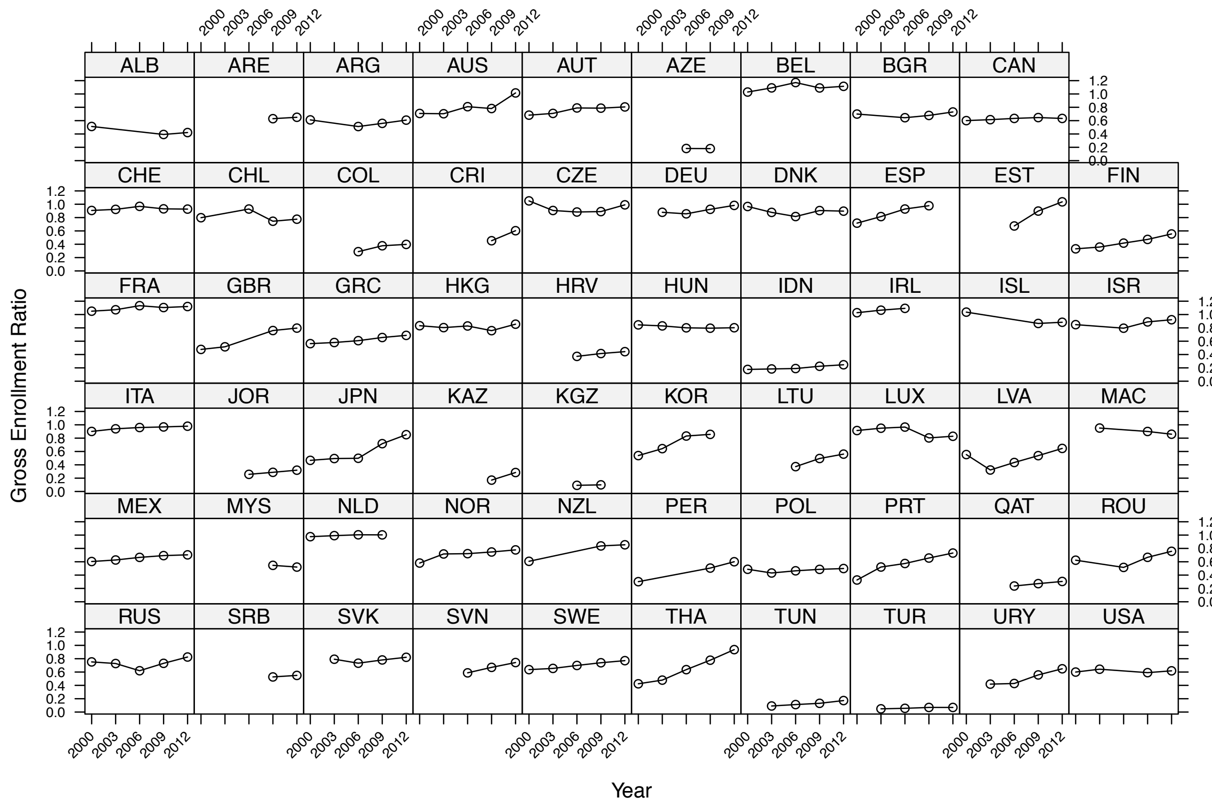
3.2.2 Preschool Enrollment

International comparative measures of educational attainment compiled by UNESCO or Eurostat are frequently used in social science and analysis (e.g., Barro & Lee, 2001). Most previous comparative research used the number of years of schooling. Our main explanatory variable was a country-level measure of the gross enrollment rate in preschool provided by the UNESCO Institute for Statistics. Here, preschool was defined according to the International Standard Classification of Education (ISCED) level 0 (see Endnote 1 for more details). More precisely, we used the gross enrollment in preprimary education for children between the age of three and the start of primary school. Such programs aim to prepare children for entry into primary education and equip them with the necessary academic skills (UNESCO, 2012).

The gross enrollment rate was the number of children enrolled, regardless of their actual age, divided by the total number of children in the official preschool age group. Due to early preschool starts or late school starts, this ratio may exceed 1. We used the preschool participation rates from the years before the respective student cohort started school. For instance, student achievement data for fourth graders in PIRLS 2001 was combined with preschool enrollment rates from 1997.[iv] For most countries, this information is available annually. For the rare cases in which values were missing, we inferred the average gross enrollment rate based on data for the two adjacent years. With correlations of r=.83, the UNESCO enrollment data is highly correlated with the survey data from PIRLS which asked parents how long their children had attended preschool (see Robustness Checks below).

Table 1 summarizes the lowest and highest gross enrollment rate observed in each country. For the PISA data, the mean enrollment across countries and over time was 0.67 with a standard deviation of 0.26. The table not only shows international variation in enrollment levels but also highlights considerable within-country variation over time. The average within-country standard deviation is 0.07, that is, the preschool coverage within countries varies by 7% from the country's mean. Figure 2 illustrates the variation over time for the PISA countries (see Appendix C for PIRLS). Besides a general increase in gross enrollment rates, there is evidence that countries increased preschool participation at a different speed. The largest changes were observed in Australia, Estonia, the United Kingdom, Portugal, Romania, and Thailand. The countries with relatively stable enrollment rates include Canada, Hungary, Indonesia, Poland, Turkey, and the United States. It must be noted that the years on the abscissa refer to the PISA assessment, i.e., they refer to preschool enrollment in the 1990s and early 2000s when PISA students were at preschool age. We observed a similar distribution of the preschool enrollment for the PIRLS data. The mean enrollment was 0.71 with a standard deviation of 0.24, and an average within-country (across years) standard deviation was 0.08.

Figure 2 Gross enrollment rate in preschool for the PISA samples, 2000-2012



Note. UNESCO preschool enrollment data.

3.2.3 *Covariates*

Individual-level covariates were gender, socioeconomic background, and language spoken at home. The teacher-reported gender was used in PIRLS, and student-reported gender was used in PISA. To measure socio-economic background, the PISA data includes the so-called index of economic, social, and cultural status (ESCS). This index was created on the basis of student-reported information on their parents' occupational status, level of education, wealth, and possession of educational and cultural resources; it is comparable across study years (OECD, 2014, Chapter 16). As such an index was only available for PISA, we simply used the number of books as an indicator of socio-economic background in PIRLS. Furthermore, students were asked whether they spoke the language of the reading tests at home in both studies. Besides the language spoken at home, we did not use any further information on migratory background because migration has different meanings across countries while the language use at home has a similar meaning and is more proximal to reading literacy. Further, social differences between immigrant and native students were taken into account by controlling for socio-economic background.

We combined the student data with additional country-level covariates that were available for different years. Information on economic performance, educational spending, and pupil-teacher ratios is provided by the UNESCO Institute for Statistics and the World Bank. We used the log GDP (gross domestic product) per capita (current US$) as a general measure of economic performance. The annual government expenditure per student (% of GDP per capita) in primary and secondary education was used to measure educational spending. The pupil-teacher ratio (headcount basis) in primary and secondary education was used as a proxy for human resources at different educational stages. One factor we had to consider was that countries may have changed the school entry age, which would have affected the comparability of the country samples over time in terms of age (maturation) and grade (schooling). For this reason, we used the aggregated student age as a covariate for the grade-based PIRLS samples and the aggregated student grade for the age-based PISA samples. Both were measured on the student level but we controlled for them on the country level because early/late school entries and repeating/skipping grades would have biased their association with achievement on the student level (see Strietholt, Rosén, & Bos, 2013).[v] We list the means, standard deviations, and amount of missing data for all covariates in Appendix D for PIRLS and Appendix E for PISA.

3.3 Empirical strategy

Our empirical strategy is inspired by Hanushek, Link, and Woessmann (2013), using within-country fixed effects panel models. To describe how we sought to identify the effect of preschool on student achievement in primary and secondary school, it is useful to distinguish between variables on individual and institutional levels. Conventional research regards preschool participation as a variable on individual level, but it is also possible to measure the preschool enrollment ratio as an institutional feature on the country level. In the present study, we focused on the international variation in the preschool enrollment ratio over time using the combined data from multiple cycles (years) of international assessments (PIRLS and PISA) and panel information on enrollment in preschool on country-level from the UNESCO Institute of Statistics.[vi] The empirical issues related to this approach can be most easily seen from a simple hierarchical model, which we extended by adding a time dimension. On the student level, achievement *Y* in country *c* at time *t* for student *i* is a function of student features *I* and an error term, r_{cti}:

$$Y_{cti} = \beta_{0ct} + \beta_1 I_{cti} + r_{cti} \tag{1}$$

We assumed that each country has a different intercept at each time point, β_{0ct}. On the country level, this intercept is a function of country features *C* (here, the enrollment ratio for preschool) and an error term, u_{0ct}:

$$\beta_{0ct} = \gamma_{00} + \gamma_{01} C_{ct} + u_{0ct} \tag{2}$$

As we used international data from different years, it was useful to expand the error term on country level into time-invariant (v_{0c}) and time-varying (v_{0ct}) components:

$$u_{0ct} = v_{0c} + v_{0ct} \tag{3}$$

Our model with individual-level and country-level predictors could be written as a single model by stepwise substituting equation (3) into (2), and then (2) into equation (1):

$$Y_{cti} = \gamma_{00} + \gamma_{01}C_{ct} + \beta_1 I_{cti} + v_{0c} + v_{0ct} + r_{cti} \tag{4}$$

We were mainly interested in estimating γ_{01}, the impact of preschool on later student achievement holding other predictors of student performance constant. To identify this parameter, the error term had to be orthogonal to the observed explanatory variables and, in particular, to the measure of preschool. Unlike conventional random intercept models for cross-sectional data, the error term in equation (4) has three components, where v_{0c} stands for time-invariant institutional features in country *c* (e.g., country size, the organizational structure of the school system, Confucian heritage), v_{0ct} stands for time-varying institutional features in country *c* (e.g., changes in spending on education, economic development, pupil-teacher ratios), and r_{cti} is time-varying individual error term. We will now elaborate on how our approach addressed the issues related to the three parts of the error term.

At the individual level, unobserved child characteristics and preschool participation may be correlated. If, for instance, children from disadvantaged families participate less frequently in preschool, the estimation of the effect of preschool would be biased because r_{cti} would be correlated with the explanatory variable. We circumvented bias from any selection mechanisms at the individual level because our explanatory variable was the preschool participation rate observed at country level. Another advantage of modeling preschool on country level relates to remedial education in primary and secondary school. On individual level remedial education may bias the estimation of the *long-term* effects of preschool on student achievement downwards: If preschool helps children to develop certain academic competences, participation in preschool and participation in later remedial education may be negatively correlated because more education efforts are to be assigned to children in need (those who did not attend preschool). Country-level analyses encompass both direct effects on children who attend preschool as well as spillovers to other children.

At country level, unobserved institutional factors may be correlated with preschool participation rates. Using the repeated measures from international assessments along with the longitudinal data on change in preschool participation rates allowed us to eliminate bias emerging from any time-invariant institutional confounders in v_{0c}. For this purpose, we estimated a fixed-effects model, where we added sets of dummies for countries, μ_c, and study years, μ_t:

$$Y_{cti} = \gamma_{00} + \gamma_{01}C_{ct} + \beta_1 I_{cti} + \mu_c + \mu_t + v_{0ct} + r_{cti} \quad (5)$$

In this model, we estimated γ_{01} based upon changes in the enrollment in preschool over time because all stable country characteristics were absorbed into the country fixed effects. The fact that our approach effectively circumvented bias from any stable country features is reflected in equation (5) because it does not contain an error term for time-invariant institutional features. With cross-sectional data, it is usually not possible to disentangle the bias from time-variant and stable confounding variables. The key feature of our analytical approach was that we exploited the within-country variation over time for the estimation of the preschool effect.

Possible issues remaining in the identification of the effect of preschool included the time-varying confounders in v_{0ct}. The estimation of γ_{01} would have been biased if changes in the preschool enrollment ratio and changes in other institutional features were correlated. We attempted to address this by including a set of demographic, economic, and educational covariates in our analyses. For this purpose, we used the rich data from the background questionnaires of the international assessments (such as data on gender, social background) along with data on international databases (such as the pupil-teacher ratio in the school system, spending on education). We controlled for student characteristics at individual level to address possible changes in the demographic makeup of countries (e.g., due to migration) and included institutional covariates at country level. The main assumption of our approach was that there was no hidden bias from unobserved time-varying confounders that are unrelated to the observed covariates.[vii]

We were also interested in how the expansion of preschool affected children from different family backgrounds. If preschool had a positive effect for disadvantaged children but no effect or even a negative one for privileged children, this would indicate that increasing preschool enrollment rates had been effective in closing the social achievement gap. To test the impact of the extension of preschool enrollment rates on social achievement inequality, we split the samples into children from privileged and disadvantaged backgrounds and replicated the main analyses for each subsample.[viii] The "number of books" variable in PIRLS and the index of economic, social, and cultural status (ESCS) in PISA were used to divide the full samples into subsamples.

3.4 Missing Data

As we restricted our samples to countries with achievement and preschool enrollment data, there were no missing data for these variables. The student-level covariates had low levels of missingness, ranging from 0.1% to 7.9% in PIRLS and 0.1% to 4.5% in PISA (Appendices D and E lists % the percentage of missing data for each variable). The range of missingness is higher for the country-level data, from 0.0% to 20.4% in PIRLS and 0.0% to 46.6% in PISA. As missing data can bias results, and in order to take advantage of all available information—that is, to avoid removing cases with incomplete data—we used the full information maximum likelihood (FIML) estimator in Mplus 7 (Muthén & Muthén, 1998–2015).

4 Results

4.1 Main Results

As a baseline for our panel analyses below, we pooled the data from the various study waves and regressed student achievement on the county-level preschool enrollment measure. Such simple analyses estimated the association between preschool enrollment rates and later school achievement from the cross-sectional variation. The results for reading in primary (PIRLS) and secondary (PISA) school showed statistically significant positive associations and are reported in column 1 in Table 2 and Table 3. As the achievement scales in both studies were created such that the international standard deviation was 100 in the year 2000, the observed associations meant that an increase from 0 to 1 (i.e., 100%) in the preschool variable corresponded to an increase of more than one standard deviation on the achievement scales. However, although measuring preschool enrollment on the country level circumvented selection bias on the individual level, a serious concern with these associations was that possible confounders on the country level may have been driving them.

The main model with country-by-year fixed effects effectively controlled for any time-invariant confounding factors because they exploit longitudinal variation on country level. As this strategy basically relates change in preschool enrollment to change in the outcome, constant confounders cannot bias the estimation of the preschool effect. Furthermore, we added a set of control variables that may be correlated with change in preschool enrollment to the model with country-by-year fixed effects. In particular, we included student covariates to capture change in the demographic composition and time-varying country covariates to capture change in institutional features such as other educational reforms that were possibly correlated with the expansion of preschool enrollment rates. The results of these analyses revealed that the cross-sectional associations completely vanished for the PIRLS (column 2 in Table 2; Appendix D shows parameter estimates for the full models including covariates) as well as for the PISA data (column 2 in Table 3; Appendix E). This finding suggests that the previously reported cross-sectional associations do not reflect the effect of preschool on later school achievement but rather bias from between-country confounding factors. Our main analyses provide no evidence for an effect of preschool on school achievement in the medium term.

Yet, the average effect of an expansion of preschool participation may be hiding heterogeneity in the effects for children from different social backgrounds. It can be hypothesized that preschool is more effective for children from socially disadvantaged families who grew up in less stimulating learning environments than for children from socially privileged backgrounds. To test for this, we split the samples into children from socially privileged and disadvantaged backgrounds and replicated the analyses for the respective subsamples. However, we found no support for this hypothesis because the effects remained nonsignificant for both social groups of students (columns 4 and 5 in Table 2 and 3).

Table 2 Effects of preschool enrollment rates on reading achievement in primary school (PIRLS)

	Cross-sectional	Panel fixed effects		
	All students	All students	Subsample of disadvantaged students	Subsample of privileged students
	1	2	3	4
Preschool enrollment rate	124.9**	7.1	8.7	3.1
	(20.1)	(21.7)	(23.3)	(16.7)
Country fixed effects	No	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes

Covariates	No	Yes	Yes	Yes
Student observations	389,641	389,641	138,490	228,680
Country observations	32	32	32	32
Country-by-year observations	80	80	80	80

Note. Each column represents separate two-level regressions. The dependent variable is the PIRLS reading score (five plausible values). Full information maximum likelihood estimation weighted by sampling weights for equally weighted country-by-year observations. Student covariates are sex, language at home, and books at home. Country covariates are average student age, pupil-teacher ratio (primary school), GDP per capita and government expenditure per student (primary school). Descriptive statistics and full results of the model in column 2 are reported in Appendix D. The variable on the number of books at home was used to split the full sample (students with missing data on this variable were excluded) into subsamples of "disadvantaged" (up to 25 books at home, column 3) and "privileged" (more than 25 books at home, column 4) children.
** 1% significance level (two-tailed).
* 5% significance level (two-tailed).

Table 3 Effects of preschool enrollment rates on reading achievement in secondary school (PISA)

	Cross-sectional	Panel fixed-effects		
	All students	All students	Subsample of disadvantaged students	Subsample of privileged students
	1	2	3	4
Preschool enrollment rate	111.4**	12.6	1.5	21.1
	(11.6)	(12.4)	(14.5)	(12.0)
Country fixed effects	No	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes
Covariates	No	Yes	Yes	Yes
Student observations	1,646,345	1,646,345	645,897	968,748
Country observations	59	59	59	59
Country-by-year observations	234	234	232	232

Note. Each column represents separate two-level regressions. Dependent variable is the PIRLS reading score (five plausible values). Full information maximum likelihood estimation weighted by sampling weights for equally weighted country-by-year observations. Student covariates are sex, language at home, and the index of economic, social and cultural status (ESCS). Country covariates are average grade, pupil-teacher ratio (primary and secondary school), GDP per capita and government expenditure per student (primary and secondary school). Descriptive statistics and full results of the model in column 2 are reported in Appendix E. The variable ESCS was used to split the full sample (students with missing data on this variable were excluded) into subsamples of "disadvantaged" (ESCS ≤ -.434) and "privileged" (ESCS > -.434). The samples from Macedonia 2000 and Japan 2000 contain no information for the variable ESCS and were excluded for the estimation of the models in columns 3 and 4.
** 1% significance level (two-tailed).
* 5% significance level (two-tailed).

4.2 Robustness Checks

Several further analyses confirmed the robustness of our main findings. In the following sections, we present additional checks using alternative measures and more homogenous samples to assess the sensitivity of our main results. Further, we present evidence that shows that our study is powered to identify a policy-relevant effect of preschool enrollment on later student achievement.

4.2.1 OECD country samples

A wide range of countries participated in the international student assessments we considered in the present study. From an analytical perspective, the inclusion of all countries generally implies greater statistical power because the sample size is greater and the variation in the explanatory variable is higher in a diverse sample. However, there may be issues regarding the comparability of very heterogeneous countries. We addressed this concern by replicating the country-by-year fixed effects analyses including covariates with a restricted set of OECD countries; these countries tended to be high-income countries with a very high human development index. This restriction reduced the analyses from 32 to 16 countries for PIRLS and 59 to 34 for PISA. The overall pattern of results was virtually the same: Preschool had no significant effect on achievement in primary and secondary school (Table 4, columns 1 and 2).

Table 4 Robustness Checks

Data	PIRLS	PISA	PISA	PISA	PIRLS
Modification	OECD countries	OECD countries	Math as outcome	Science as outcome	Alternative preschool measure
	1	2	3	4	5
Preschool enrollment rate	-10.2	2.6	-5.3	-7.7	3.4
	(19.4)	(15.7)	(10.6)	(12.4)	(2.3)
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes
Student observations	223,160	1,180,544	1,646,345	1,646,345	357,683
Country observations	16	34	59	59	31
Country-by-year observations	42	151	234	234	75

Note. Each column represents separate two-level regressions. Dependent variables are the PIRLS or PISA achievement scores. Full information maximum likelihood estimation weighted by sampling weights for equally

weighted country-by-year observations. Student covariates are sex, language at home, and the index of economic, social and cultural status (ESCS). Country covariates are average grade, pupil-teacher ratio (primary and secondary school), GDP per capita and government expenditure per student (primary and secondary school). For the model in column (5), data from Kuwait 2001, Morocco 2001, and the US 2001/2006/2011 were excluded because no survey information on preschool participation is available.
** 1% significance level (two-tailed).
* 5% significance level (two-tailed).

4.2.2 *Mathematics and science achievement as outcomes in PISA*

PISA tests students in reading, math, and science. So far, we have presented results for reading outcomes because the reading achievement scores of the various study cycles have a common scale, which is not true for the PISA math and science scales (OECD, 2014). Following Hanushek, Link, and Woessmann (2013), we replicated the analyses with alternative outcome measures because the fixed effects for years (at least partly) absorbed differences in the scales of the achievement scores. The results of the additional analyses were qualitatively the same; preschool had no significant effect on achievement in math and science in primary and secondary school (Table 4, columns 3 and 4).

4.2.3 *Measuring preschool enrollment with survey data*

The information on preschool enrollment we used came from official national statistics that government agencies reported to UNESCO. Although UNESCO has considerable experience in defining and collecting international comparative educational statistics, national agencies may misreport their data or provide unreliable information. To validate our findings in this regard, we also used country-level aggregated survey information on preschool participation. Such information was collected in PIRLS. Parents were asked, in retrospect, whether their children had attended preschool and for how long. The wordings of the response scales were somewhat different across study cycles, but it was possible to reduce the impact of this by recoding the data (see Appendix F). The fixed effects for years ameliorated further problems related to changes in the item wording. The survey preschool measure had a mean of 3.93 with a standard deviation of 0.57. The country-level correlation between the UNESCO measure and the aggregated survey measure of preschool enrollment was r=0.83 (SE=.06, p<0.01). Replicating our main analysis with the aggregated survey measures of preschool enrollment confirmed the previous finding that preschool had no significant effect on later student achievement (Table 4, columns 5).

Certain groups of students may have been more affected by preschool expansion than others. For example, if countries only expanded preschool participation among disadvantaged students, we could not have reasonably expected such an expansion to have affected the performance of privileged students. To address this possibility, we computed the survey preschool measure for the subsamples of “disadvantaged” (up to 25 books at home) and “privileged” (more than 25 books at home) students and replicated the analysis for the respective subsamples. Again, the results were qualitatively the same and provided no support for the idea that preschool participation impacted later student achievement (not presented in Table 4).

Another approach addressing possible measurement issues related to the preschool measure involved restricting the analyses to a set of countries for which we were confident that UNESCO had reliable data. Following Heyneman (1999), OECD members have generally higher standards of quality control when reporting data. From this perspective, the previously presented robustness check on the subset of OECD countries also provides further evidence for the credibility of the UNECSO preschool enrollment measure.

4.2.4 *Post hoc statistical power*

The minimum detectable effect (MDE) was calculated as a sensitivity test to verify that the present study is powered to detect the effect of a policy-relevant change in preschool. Bloom (1995) suggests multiplying the standard errors of a policy-relevant effect estimate by 2.8 to receive the MDE that has an 80 percent power to detect estimates that is statistically significant at .05 level (two-sided).

What it a policy-relevant increase in preschool enrollment? The reported effects and standard errors for a one-unit change in the treatment variables correspond with an increase in preschool from zero to complete enrollment. However, no country experienced such a huge increase, but the average within-country standard deviation was 7 percent for PISA and 8 percent for PIRLS (see Section 3.3.3). Based on these values, we consider a 10 percent increase a more realistic change in preschool for our power calculation.

The standard errors for a 10 percent increase is 2.17 for PIRLS and 1.24 for PISA (the standard errors of the main specifications were 21.7 and 12.4 and they were divided by 10; see Table 2, column 2, Table 3, column 2). Multiplying these values by the previously mentioned factor 2.8 gives the MDE=6.08 for PIRLS and the MDE=3.47 for PISA data at 80 percent power and 0.05 significance.

The reading scales in PIRLS and PISA have an international standard deviation of 100 points. We can use this information and divide the MDEs by 100 to achieve standardized MDEs of 0.06 (PIRLS) and 0.03 (PISA). In other words, our study is powered to detect policy-relevant effect as small as only 0.06 (PIRLS) and 0.03 (PISA) standard deviations of the outcome variable. To put these effects in context, the average effects sizes for the participation in the Perry Preschool program and the Abecedarian study on academic achievement at school age is about ten times larger (0.3 to 0.5 standard deviations; see Barnett, 2011).

5 Discussion

For quite some time now, international organizations have been arguing for preschool expansion worldwide and are continuing to do so. Policy makers all around the globe are investing in increasing preschool enrollment rates in the hope that this is an effective strategy to improve educational outcomes amongst other things. However, there is growing evidence at least advising caution and suggesting that such expansion efforts of early childhood education may not ultimately improve students' medium-term academic skills internationally. Issues that are discussed in this context include difficulties in scaling-up successful programs and the observation of fade-out effects. Although there is much research focusing on the efficacy of scaled-up preschool education, most studies are local, regional, or at best national in their scope. Taken as a whole, the evidence is very mixed and does not give much guidance for international policy makers. By taking an international comparative perspective, our study aimed to add to the existing evidence base by offering a more overarching, cross-national perspective. We could not find any support for the idea that expanding preschool enrollment necessarily leads to better student achievement in primary and secondary school on country level.

A key contribution of our study is that we did not only consider direct effects on children who attended preschool but also spillovers to other children. It has been proposed that children who do not attend preschool catch up in school, potentially because more support is offered to children who have not mastered basic skills (e.g., Magnuson, Ruhm, & Waldfogel, 2007b). Following this line of argumentation, a fair evaluation of early education expansion efforts should not only consider the effects on children who participated in preschool but also possible positive externalities that emerge. A key advantage of conceptualizing early education as enrollment rates at the country level is that we were able to capture both direct effects on children who participated in preschool as well as spillovers on other children.

5.1 Possible Explanations for the Null Effect

Why didn't the expansion of preschool lead to higher student achievement in the medium run? First, it is important to bear in mind that the present study did not evaluate a carefully designed small-scale program, but that we used international data to evaluate the global expansion of early childhood education in order to inform international policy making. Our results may be tentatively interpreted as suggesting that the global expansions of preschools during the 1990s and early 2000s were on average designed in such a way that they did not impact longer term student achievement. It is not within the scope of our study to say whether this is due to ineffective ECEC policies and programs per se or due to an ineffective of ECEC and subsequent schooling combination. There is some evidence suggesting that a positive impact of early education can only be sustained if it is followed by a school system that is designed in a way to retain this effect (e.g., Magnuson et al., 2007b).

It may well be that certain programs in certain countries had positive effects. Given that we found a neutral overall effect, however, this would imply that participation in preschool had a negative effect in other countries and that the positive and negative effects cancelled each other out. Following this argument, the calls for a worldwide expansion of preschool systems by the OECD, The World Bank, UNICEF, and the EU do not satisfy expectations, at least not overall in relation to our outcome of interest, i.e. country-level test scores through school age.

Another aspect relates to children who did not attend any form of early education. The central question here is what kinds of nonformal or informal learning environments they experienced and what they were doing at home or in informal care contexts (Duncan & Magnuson, 2013). What is the added value of preschool if children grow up in stimulating home learning environments? As there is a growing awareness of the importance of early education for child development, parents may be providing more stimulating home learning environments than they did decades earlier, which is when some seminal and influential studies and programs like the Perry Preschool Program were conducted. For example, in the United States, there was an increase in parents' investment in their children's home learning environments, most strongly so in low-income families, between 1998 and 2010 (Bassok, Finch, Lee, Reardon, & Waldfogel, 2016). Meanwhile, other measures such as Sesame Street or family interventions were also implemented so that children who did not participate in preschool and did not receive enough parental support could benefit (e.g., Hannon, 2003).

Importantly, null effects of preschool expansions on medium-term achievement scores do not necessarily mean that there are no other positive effects of preschool in the long run. As pointed out by Phillips et al. (2017), cumulative evidence suggests sleeper effects, meaning that the fade-out in academic domains can be followed by gains in other domains, like education, income, and employment. For example, a recent meta-analysis including quasi-experimental studies of universal early education programs in the United States, Canada, and Western Europe (van Huizen & Plantenga, 2018) reports insignificant medium-term effects but mostly positive long-term impacts with regard to completed education and labor market success. An important follow up of our current analyses will be to address this issue also in international comparative data.

5.2 Limitations of the Study

Finally, it is important to note the inherent difficulties in constructing comparable measures for international comparisons and to bear in mind that the present study's scope is limited to specific medium-term educational outcomes. First, the school entry age varies across countries and there is hence some variation in the meaning of the gross preschool enrollment rate across countries. In this regard, we followed UNESCO's approach and used their preschool measures. Second, although we considered key measures of student achievement at different educational stages and in different domains, there are other important outcomes, such as noncognitive and behavioral measures we did not consider. Third, it should be noted that we consider data from only roughly one third of all countries worldwide (66 countries). The sample of countries is heterogenous to a certain extent and, to check the sensitivity of our results, we repeated our analysis for a more homogenous sample of high-income OECD-countries with a very high human development index. Consequently, we believe our results to be quite robust for these countries but we were not able to conduct similar robustness checks for sets of countries with other characteristics (i.e. low-income countries). Also, certain regions such as Africa and large parts of Asia are not well represented. Therefore, our results cannot be generalized beyond the countries included in our study. Moreover, although preschool education programs for children have an educational remit, this is not the only motivation to increase participation rates. There are other possible outcomes that need to be considered (i.e. behavioral outcomes and social competences) For example, public daycare services and mothers' labor market participation are closely related. There is some international evidence that the availability of public childcare has positive effects on maternal employment (e.g., Bauernschuster & Schlotter, 2015; Cascio, 2009b).

We want to close our discussion by underscoring the need for caution when interpreting our results. Even with longitudinal data, it is difficult to identify the impact of large-scale educational interventions on educational outcomes with international data (Hanushek & Woessmann, 2011). Obviously, it is easier to identify the impact of a carefully implemented small-scale intervention. However, we believe that our research is an important complement to those rigorous interventions, because the external validity of these studies is unclear (Hanushek et al., 2013). Yet, there is always the possibility that the expansion of preschool could be correlated with other educational reforms or institutional changes. Our fixed-effects analyses effectively control for any (observed and unobserved) stable confounding variables, but time-varying confounders remain a possible issue. However, we controlled for several observed covariates, including student characteristics and various time-varying institutional measures. The only remaining sources of bias are time-varying covariates correlated with student achievement but uncorrelated with the observed covariates. While it is impossible to eliminate this risk, we think that our results should initiate policy discussions about the goals that are actually pursued by preschool expansion efforts and how to evaluate their achievement.

References

Artz, B., & Welsch, D. M. (2016). The impact of publicly provided early childhood education programs on district-level test scores. *Contemporary Economic Policy, 34*(1), 89-106. doi:10.1111/coep12128

Bailey, D., Duncan, G.J., Odgers, C.L. & Yu, W. (2017). Persistence and fadeout in the impacts of Child and Adolescent Interventions. *Journal of Research on Educational Effectiveness*, 10, 7-39. doi: 10.1080/19345747.2016.1232459

Baker, M., Gruber, J., & Milligan, J. (2008). Universal child care, maternal labor supply, and Family Well‐Being. *Journal of Political Economy, 116*(4), 709-745. doi:10.1086/591908

Baker, M., Gruber, J., & Milligan, K. (2019). The Long-Run Impacts of a Universal Child Care Program. *American Economic Journal: Economic Policy, 11*(3), 1-26. doi:10.1257/pol.20170603

Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science, 333*(6045), 975–978. doi:10.1126/science.1204534

Barro, R. & Lee, J. W. (2001). International Data on Educational Attainment: Updates and Implications. *Oxford Economic Papers 53*(3), 541–563. doi: 10.1093/oep/53.3.541

Bassok, D., Finch, J. E., Lee, R., Reardon, S. F., & Waldfogel, J. (2016). Socioeconomic gaps in early childhood experiences: 1998 to 2010. *AERA Open*, *2*. doi: 10.1177/2332858416653924

Bauernschuster, S. & Schlotter, M. (2015). Public child care and mothers‘ labor supply—Evidence from two quasi-experiments. *Journal of Public Economics, 123,* 1–16.

Berlinski, S., Galiani, S., & Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics, 93*(1–2), 219-234. doi:http://dx.doi.org/10.1016/j.jpubeco.2008.09.002

Berlinski, S., Galiani, S., & Manacorda, M. (2008). Giving children a better start: Preschool attendance and school-age profiles. *Journal of Public Economics, 92*(5–6), 1416-1440. doi:http://dx.doi.org/10.1016/j.jpubeco.2007.10.007

Brunello, G., & Rocco, L. (2013). The effect of immigration on the school performance of natives: Cross country evidence using PISA test scores. *Economics of Education Review, 32*. doi:10.1016/j.econedurev.2012.10.006

Burger, K. (2016). Intergenerational transmission of education in Europe: Do more comprehensive education systems reduce social gradients in student achievement? *Research in Social Stratification and Mobility, 44*, 54-67. doi:10.1016/j.rssm.2016.02.002

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112*(3), 579-620.

Campbell, J. R., Kelly, D. L., Mullis, I. V. S., Martin, M. O., & Sainsbury, M. (2001). *Framework and specifications for PIRLS assessment 2001*. Chestnut Hill, MA: International Study Center, Lynch School of Education, Boston College.

Cascio, E. U. (2009a). Do investments in universal early education pay off? Long-term effects of introducing kindergartens into public schools. NBER Working Paper 14951. Retrieved from http://www.nber.org/papers/w14951

Cascio, E. U. (2009b). Maternal labor supply and the introduction of kindergartens into American public schools. *Journal of Human Resources, 44*(1) 140-170. doi:http:// 10.1353/jhr.2009.0034

Cascio, E. U., & Schanzenbach, D.W. (2013). The impacts of expanding access to high-quality preschool education. *Brookings Paper on Economic Activity*, 127-192. doi:https://doi.org/10.1353/eca.2013.0012

Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. Handbook of the Economics of Education, 1, 697-812. doi:10.1016/S1574-0692(06)01012-9

Davier, M. v., Gonzales, E. J., & Mislevy, R. J. (2009). What are plausible values and why are they useful? In *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments* (Vol. 2, pp. 9-36). Hamburg/Princeton NJ: IEA-ETS Research Institute.

Dearing E. (2007). Psychological costs of growing up poor. *Annuals of the New York Academy of Sciences*. 1136: 324-332. doi:10.1196/annals.1425.006

Denboba, A. D., Sayre, R. K., Wodon, Q. T., Elder, L. K., Rawlings, L. B., & Lomabardi, J. (2014). Stepping up early childhood development. Investing in young children for high returns. Retrieved from http://documents.worldbank.org/curated/en/868571468321240018/pdf/92988-REVISED-PUBLIC-WB-ECD-Mar2016-ENG-v2-web.pdf

Dodge, K. A., Bai, Y., Ladd, H. F., & Muschkin, C. G. (2016). Impact of North Carolina's early childhood programs and policies on educational outcomes in elementary school. *Child Development*, n/a-n/a. doi:10.1111/cdev.12645

Dumas, C., & Lefranc, A. (2012). Early schooling and later outcomes: evidence from pre-school extension in France. THEMA Working Paper No 2010-07. Retrieved from http://thema.u-cergy.fr/recherche/documents-de-travail/?lang=fr

Duncan, G. J., & Magnuson, K. (2013). Investing in preschool programs. *Journal of Economic Perspectives, 27*(2), 109-132. doi:10.1257/jep.27.2.109

Engle, P. L., Fernald, L. C. H., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., . . . Iltus, S. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet, 378*(9799), 1339-1353. doi:10.1016/s0140-6736(11)60889-1

Esping-Andersen, G., Garfinkel, I., Han, W.-J., Magnuson, K., Wagner, S., & Waldfogel, J. (2012). Child care and school performance in Denmark and the United States. *Children and Youth Services Review, 34*(3), 576-589. doi:http://dx.doi.org/10.1016/j.childyouth.2011.10.010

Felfe, C., & Lalive, R. (2012). Early child care and child development: For whom it works and why. IZA Discussion Paper Series. No 2010. Retrieved from http://ftp.iza.org/dp7100.pdf

Felfe, C., Nollenberger, N., & Rodríguez-Planas, N. (2015). Can't buy mommy's love? Universal childcare and children's long-term cognitive development. *Journal of Population Economics, 28*(2), 393-422. doi:10.1007/s00148-014-0532-x

Fitzpatrick, M. D. (2008). Starting school at four: The effect of universal pre-kindergarten on children's academic achievement. SIEPR Discussion Paper No. 08-05. Retrieved from www-siepr.stanford.edu/Papers/pdf/08-05.pdf

Fredriksson, P., Hall, C., Johansson, E.-A. and Johansson, P. (2010), Do pre-school interventions further the integration of immigrants? Evidence from Sweden, in E.-A. Johansson (ed.), *Essays on Schooling, Gender, and Parental Leave, Economic Studies 121*, Department of Economics, Uppsala University.

Hannon, P. (2003). Family literacy programmes. In N. Hall, J. Larson, J. Marsh (Eds.), *Handbook of early childhood literacy.* (pp. 99-111) London: Sage.

Hanushek, E., Link, S., & Woessmann, L. (2013). Does school autonomy make sense everywhere? Panel estimates from PISA. *Journal of Development Economics, 104*, 212-232. doi:10.1016/j.jdeveco.2012.08.002

Hanushek, E. A., & Woessmann, L. (2011). The economics of international differences in educational achievement. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education*. (Vol. 3) (pp. 89-200). Amsterdam: Elsevier.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*(5782), 1900-1902. doi:10.1126/science.1128898

Heyneman, S. P. (1999). The sad story of UNESCO's education statistics. *International Journal of Educational Development, 19*(1), 65-74.

Hill, C. J., Gormley Jr., W. T., & Adelstein, S. (2015). Do the short-term effects of a high-quality preschool program persist? *Early Childhood Research Quarterly, 32*(3), 60-79. doi: http://dx.doi.org/10.1016/j.ecresq.2014.12.005

Hogrebe, N., & Strietholt, R. (2016). Does Non-participation in Preschool Affect Children's Reading Achievement? International Evidence from Propen- sity Score Analyses. In: Large-scale Assessments in Education. 4(2), 1-22. Doi:10.1186/s40536-016-0017-3.

Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proc Natl Acad Sci U S A, 103*(27), 10155-10162. doi:10.1073/pnas.0600888103

Leseman, P. P., & Slot, P. L. (2014). Breaking the cycle of poverty: challenges for European early childhood education and care. *European Early Childhood Education Research Journal, 22*(3), 314-326. doi: 10.1080/1350293X.2014.912894

Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2010). Expanding schooling opportunities for 4-year-olds. *Economics of Education Review, 29*(3), 319-328.

Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly, 45*, 155-176. doi:https://doi.org/10.1016/j.ecresq.2018.03.005

Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007a). Does prekindergarten improve school preparation and performance? *Economics of Education Review 26*(1), 33-51. doi: 10.1016/j.econedurev.2005.09.008

Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007b). The persistence of preschool effects: Do subsequent classroom experiences matter? *Early Childhood Research Quarterly 22*(1), 18-38. doi:10.1016/j.ecresq.2006.10.002

Marope, P. T. M., & Kaga, Y. (Eds.) (2015). *Investing against evidence. The global state of early childhood care and education.* Paris: UNESCO Publishing.

McCoy, D. C., Yoshikawa, H., Ziol-Guest, K. M., Duncan, G. J., Schindler, H. S., Magnuson, K., . . . Shonkoff, J. P. (2017). Impacts of Early Childhood Education on Medium- and Long-Term Educational Outcomes. *Educational Researcher, 46*(8), 474-487. doi:10.3102/0013189X17737739

Mullis, I. V. S., Martin, M. O., Foy, P., & Drucker, K. T. (2012). *PIRLS 2011 international results in reading*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Nakajima, N., Hasan, A., Jung, H., Brinkman, S., Pradhan, M., & Kinnell, A. (2019). Investing in school readiness: A comparison of different early childhood education pathways in rural Indonesia. *International Journal of Educational Development, 69*, 22-38. doi:https://doi.org/10.1016/j.ijedudev.2019.05.009

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus User's Guide* (7 ed.). Los Angeles, CA: Muthén & Muthén.

Nores, M., & Barnett, W. S. (2010). Benefits of early childhood interventions across the world: (Under) Investing in the very young. *Economics of Education Review, 29*(2), 271-282. doi:http://dx.doi.org/10.1016/j.econedurev.2009.09.001

Organization of Economic Co-operation and Development (OECD). (2001a). *Starting strong: Early childhood education and care.* Paris: OECD Publishing. doi:http://dx.doi.org/10.1787/9789264192829-en

Organization of Economic Co-operation and Development (OECD). (2001b). *Knowledge and skills for life: First results from the OECD Programme for International Student Assessment (PISA) 2000.* Paris: OECD Publishing.

Organization of Economic Co-operation and Development (OECD). (2006). *Starting strong II: Early childhood education and care.* Paris: OECD Publishing. doi:http://dx.doi.org/10.1787/9789264035461-en

Organization for Economic Development and Cooperation (OECD). (2013). *PISA 2012 Results: Excellence through equity: Giving every student the chance to succeed* (Vol. II). Paris: OECD Publishing.

Organization of Economic Co-operation and Development (OECD). (2014). *PISA 2012 Technical Report*. Paris: OECD Publishing.

Organization of Economic Co-operation and Development (OECD). (2016). Education at a glance. OECD indicators. Paris: OECD Publishing. doi:10.1787/eag-2016-en

Phillips, D., Lipsey, M., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., & Weiland, C. (2017). Puzzling it out: The current state of scientific knowledge on pre-kindergarten effects. Washington, DC: Brookings Institution.

Rosén, M., & Gustafsson, J.-E. (2016). Is computer availability at home causally related to reading achievement in grade 4? A longitudinal difference in differences approach to IEA data from 1991 to 2006. *Large-scale Asseement in Education, 4*(5), 1-19. doi:10.1186/s40536-016-0020-8

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley + Sons.

Schlicht, R., Stadelmann-Steffen, I., & Freitag, M. (2010). Educational Inequality in the EU The Effectiveness of the National Education Policy. *European Union Politics, 11*(1), 29-60. doi:10.1177/1465116509346387

Schütz, G., Ursprung, H. W., & Woessmann, L. (2008). Education policy and equality of opportunity. *Kyklos, 61*(2), 279-308. doi:10.1111/j.1467-6435.2008.00402.x

Shonkoff, J. P., & Phillips, D. A. (2000). *From neurons to neghborhoods. The science of early childhood development.* Washington, D.C.: National Academy Press.

Sibley, E., Dearing, E., Toppelberg, C., Mykletun, A., & Zachrisson, H. (2015). Do increased availability and reduced cost of early childhood care and education narrow social inequality gaps in utilization? Evidence from Norway. *International Journal of Child Care and Education Policy, 9*(1), 1.

Strietholt, R., Rosén, M., & Bos, W. (2013). A correction model for differences in the sample compositions: The degree of comparability as a function of age and schooling? In*: Large-scale Assessments in Education, 1*(1). doi:10.1186/2196-0739-1-1

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*(1), 1-21. doi:10.1214/09-STS313

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2006). *Strong foundations: Early childhood care and education.* Paris: UNESCO.

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2012). *International standard classification of education: ISCED 2011.* Retrieved from http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2014). *Position paper on education post-2015.* Retrieved from http://unesdoc.unesco.org/images/0022/002273/227336E.pdf

United Nations Educational, Scientific and Cultural Organization (UNESCO) Institute for Statistics. (n.d.). *Gross enrolment ratio by level of education*. Retrieved from http://data.uis.unesco.org.

van Huizen, T., & Plantenga, J. (2018). Do Children Benefit from Universal Early Childhood Education and Care? A Meta-Analysis of Evidence from Natural Experiments. *Economics of Education Review*. doi:https://doi.org/10.1016/j.econedurev.2018.08.001

Weiland, C., Unterman, R., Shapiro, A., Staszak, S., Rochester, S., & Martin, E. (2019). The Effects of Enrolling in Oversubscribed Prekindergarten Programs Through Third Grade. *Child Development, n/a*(n/a). doi:10.1111/cdev.13308

World Bank. (2011). Learning for all. Investing in people's knowledge and skills to promote development. World Bank Group Education Strategy 2020. Retrieved from http://documents.worldbank.org/curated/en/685531468337836407/pdf/644870WP0Learn00Box0361538B0PUBLIC0.pdf

World Bank. (2016). World Bank Group, UNICEF urge greater investment in early childhood development [Press release]. Retrieved from https://www.worldbank.org/en/news/press-release/2016/04/14/world-bank-group-unicef-urge-greater-investment-in-early-childhood-development

CRediT authorship contribution statement Rolf Strietholt: Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing. **Nina Hogrebe:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing. **Henrik Daae Zachrisson:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing.

Acknowledgement H.D. Zachrisson's work on this article was supported by funding from the European Research Council Consolidator grant ERC-CoG-2018 EQOP [grant number 818425]. R. Strietholt has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant number 765400.

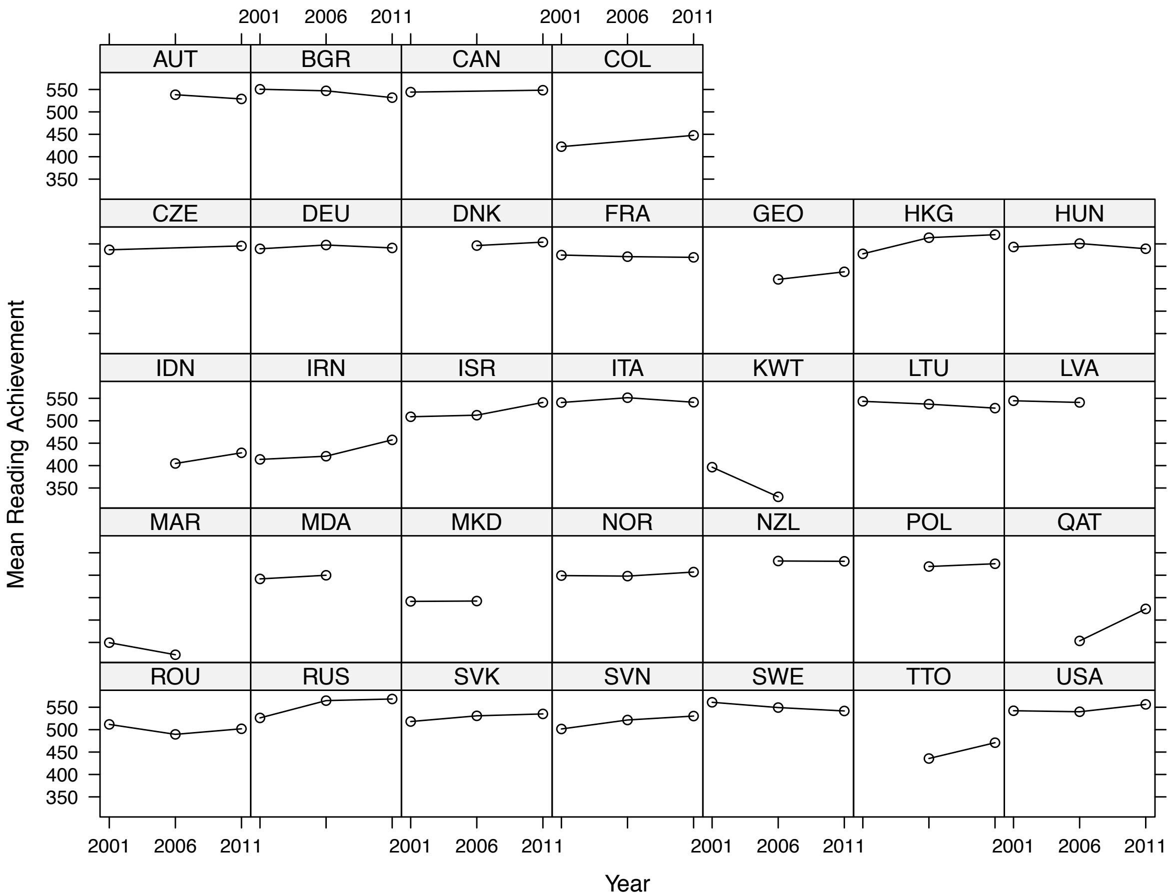
APPENDIX

Appendix A Sample sizes per country-by-year observation

		PIRLS			PISA				
		2001	2006	2011	2000	2003	2006	2009	2012
ALB	Albania				4,980			4,596	4,743
ARE	United Arab Emirates							10,867	11,500
ARG	Argentina				3,983		4,339	4,774	5,908
AUS	Australia				5,176	12,551	14,170	14,251	14,481
AUT	Austria		5,067	4,670	4,745	4,597	4,927	6,590	4,755
AZE	Azerbaijan						5,184	4,691	
BEL	Belgium				6,670	8796	8,857	8,501	8,597
BGR	Bulgaria	3,460	3,863	5,261	4,657		4,498	4,507	5,282
CAN	Canada	8,253		23,206	29,687	27,953	22,646	23,207	21,544
CHE	Switzerland				6,100	8,420	12,192	11,812	11,229
CHL	Chile				4,889		5,233	5,669	6,856
COL	Colombia	5,131		3,966			4,478	7,921	9,073
CRI	Costa Rica							4,578	4,602
CZE	Czech Republic	3,022		4,556	5,365	6,320	5,932	6,064	5,327
DEU	Germany	7,633	7,899	4,000		4,660	4,891	4,979	5,001
DNK	Denmark		4,001	4,594	4,235	4,218	4,532	5,924	7,481
ESP	Spain				6,214	10,791	19,604	25,887	
EST	Estonia						4,865	4,727	4,779
FIN	Finland				4,864	5,796	4,714	5,810	8,829
FRA	France	3,538	4,404	4,438	4,673	4,300	4,716	4,298	4,613
GBR	United Kingdom				9,340	9,535		12,179	12,659
GEO	Georgia		4,402	4,796					
GRC	Greece				4,672	4,627	4,873	4,969	5,125
HKG	Hong Kong	5,050	4,712	3,875	4,405	4,478	4,645	4,837	4,670
HRV	Croatia						5,213	4,994	5,008
HUN	Hungary	4,666	4,068	5,204	4,887	4,765	4,490	4,605	4,810
IDN	Indonesia		4,774	4,791	7,368	10,761	10,647	5,136	5,622
IRN	Iran	7,430	5,411	5,758					
IRL	Ireland				3,854	3,880	4,585		
ISL	Iceland				3,372			3,646	3,508
ISR	Israel	3,973	3,908	4,186	4,498		4,584	5,761	5,055
ITA	Italy	3,502	3,581	4,189	4,984	11,639	21,773	30,905	31,073
JOR	Jordan						6,509	6,486	7,038
JPN	Japan				5,256	4,707	5,952	6,088	6,351
KAZ	Kazakhstan							5,412	5,808
KGZ	Kyrgyzstan						5,904	4,986	

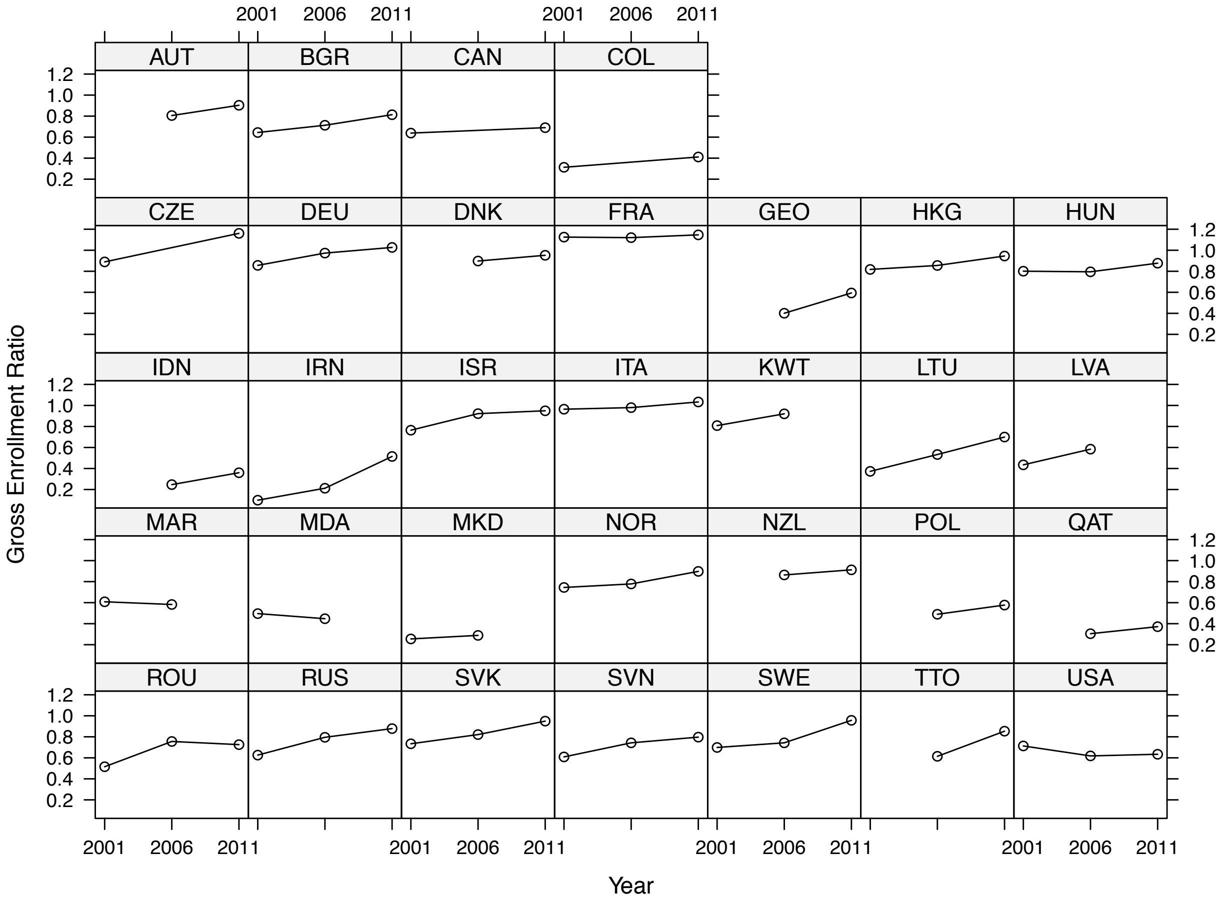
KOR	Korea, Republic of				4,982	5,444	5,176	4,989	
KWT	Kuwait	7,126	3,958						
LTU	Lithuania	2,567	4,701	4,661			4,744	4,528	4,618
LUX	Luxembourg				3,528	3,923	4,567	4,622	5,258
LVA	Latvia	3,019	4,162		3,893	4,627	4,719	4,502	4,306
MAC	Macao					1,250		5,952	5,335
MAR	Morocco	3,153	3,249						
MDA	Moldova	3,533	4,036						
MEX	Mexico				4,600	29,983	30,971	38,250	33,806
MKD	Macedonia	3,711	4,002						
MYS	Malaysia							4,999	5,197
NOR	Norway	3459	3,837	3,190	2,503	3,992	4,871	4,760	
NLD	Netherlands				4,147	4,064	4,692	4,660	4,686
NZL	New Zealand		6,256	5,644	3,667			4,643	4,291
PER	Peru				4,429			5,985	6,035
POL	Poland		4,854	5,005	3,654	4,383	5,547	4,917	4,607
PRT	Portugal				4,585	4,608	5,109	6,298	5,722
QAT	Qatar		6,680	4,120			6,265	9,078	10,966
ROU	Romania	3,625	4,273	4,665	4,829		5,118	4,776	5,074
RUS	Russian Federation	1,480	4,720	4,461	6,701	5,974	5,799	5,308	5,231
SRB	Serbia							5,523	4,684
SVK	Slovakia	3,807	5,380	5,630		7,346	4,731	4,555	4,678
SVN	Slovenia	2,952	5,337	4,512			6,595	6,155	5,911
SWE	Sweden	6,044	4,394	4,622	4,416	4,624	4,443	4,567	4,736
THA	Thailand				5,340	5,236	6,192	6,225	6,606
TTO	Trinidad and Tobago		3,951	3,948					
TUN	Tunisia					4,721	4,640	4,955	4,407
TUR	Turkey					4,855	4,942	4,996	4,848
URY	Uruguay					5,835	4,839	5,957	5,315
USA	United States	3,763	5,190	12,726	3,846	5,456	5,611	5,233	4,978

Appendix B Mean reading achievement in PIRLS, 2001-2011



Note. PIRLS reading achievement data (weighted by sampling probabilities).

Appendix C Gross enrollment rate in preschool for the PIRLS samples, 2001-2011



Note. UNESCO preschool enrollment data.

Appendix D Descriptive statistics and complete model of main specification for PIRLS data

	Descriptive statistics			Main Model	
	Mean	SD	% missing	Parameter	SE
Preschool enrollment rate	0.712	0.237	0.000	7.1	(21.7)
Student characteristics					
Male	0.507	0.500	0.001	-15.2**	(1.1)
Speak language of test at home					
Never	0.041	0.198	0.079		
Sometimes	0.204	0.402	0.079	30.7**	(6.2)
Always or almost always	0.751	0.432	0.079	42.0**	(6.4)
Books at home	2.861	1.267	0.058	16.9**	(0.6)
Country characteristics					
Average student age	10.363	0.397	0.000	108.9**	(15.2)
Pupil-teacher ratio (primary school)	15.407	5.154	0.119	0.0	(0.5)
GDP (log per capita)	9.337	1.261	0.000	4.7	(4.9)
Government expenditure per student (primary school)	0.109	0.253	0.204	-20.7	(24.5)
Country and year fixed effects				Yes	
Student observations	389,641			389,641	
Country observations	32			32	
Country-by-year observations	80			80	

Note. The sample statistics for student and country characteristics refer to the weighted full information maximum-likelihood (FIML) estimated within and between covariance matrices, respectively. Main Model: Full results of the model presented in Table 2, column 2.

** 1% significance level (two-tailed).

* 5% significance level (two-tailed).

Appendix E Descriptive statistics and complete model of main specification for PISA data

	Descriptive statistics			Main Model	
	Mean	SD	% missing	Parameter	SE
Preschool enrollment rate	0.668	0.255	0.000	12.6	(12.4)
Student characteristics					
Male	0.501	0.500	0.001	-37.6**	(0.9)
Language at home					
Other language	0.122	0.327	0.045		
Test language	0.878	0. 327	0.045	19.0**	(2.7)
ESCS (index of economic, social and cultural status)	-0.236	1.128	0.019	33.7**	(0.8)
Country characteristics					
Average Grade	9.531	0.458	0.000	1.0	(8.7)
Pupil-teacher ratio (primary school)	17.900	5.311	0.131	0.1	(1.2)
Pupil-teacher ratio (secondary school)	13.023	4.167	0.224	-0.7	(0.5)
GDP (log per capita)	9.383	1.122	0.000	18.5**	(5.4)
Government expenditure per student (primary school)	0.187	0.071	0.466	-21.4	(23.5)
Government expenditure per student (secondary school)	0.204	0.089	0.215	80.0**	(28.4)
Country and year fixed effects				Yes	
Student observations	1,646,345			1,646,345	
Country observations	59			59	
Country-by-year observations	234			234	

Note. The sample statistics for student and country characteristics refer to the weighted full information maximum-likelihood (FIML) estimated within and between covariance matrices, respectively. Main Model: Full results of the model presented in Table 3, column 2.
** 1% significance level (two-tailed).
* 5% significance level (two-tailed).

Appendix F Questionnaire items on preschool participation across PIRLS cycles.

Question wording in 2001	Question wording in 2006 and 2011	Assigned values
Did your child attend <ISCED Level 0>?	Did your child attend <ISCED Level 0>?	
“Yes”	“Yes“	-
“No“	“No“	1
If Yes... How long was he/she in <ISCED Level 0>?	If Yes... How long was he/she in <ISCED Level 0>?	
“less than 1 year“	„1 year or less“	2
“1 year“		2
“between 1 and 2 years“	„between 1 and 2 years“	3
“2 years“	„2 years“	4
„more than 2 years“	„between 2 and 3 years“	5
	„3 years or more“	5

Endnote

---

[i] In its International Standard Classification of Education (ISCED), UNESCO defines *preprimary education* as part of *early childhood education* (ISCED level 0). Specifically, the term refers to programmes targeted at children from the age of three to when they start primary education, while the category *early childhood education development* refers to children up to the age of two. Both have an intentional-education component and are meant to support children's early cognitive, language, physical, social, and emotional development through interaction with other children under the guidance of staff or educators in an institutional context (UNESCO Institute for Statistics, 2012). In our paper, we use the terms preschool education and early childhood education synonymously.

[ii] Another important international assessment is the Trends in International Mathematics and Science Study (TIMSS). We did not include it in the present study because the composition of participating countries in TIMSS varies more than in PIRLS and PISA.

[iii] PIRLS and PISA data do not contain single point estimates for student achievement; they have five "plausible values." Plausible values are random draws from the estimated student proficiency distribution, and they provide unbiased estimates for secondary analyses of international assessment data (Davier, Gonzales, & Mislevy, 2009). All analyses below were repeated for each plausible value achievement score, and the results were combined using the Rubin formula (Rubin, 1987).

iv PISA samples 15-year-old students from various grades. We used the modal grade in the respective country samples (ninth or tenth grade in most countries) to match achievement data with enrollment data from that year before this student cohort entered school.

v Grade and age are bad controls if preschool reduces rates of repeating/skipping grades. For this reason, we replicated the analyses without these covariates. The results are qualitatively the same.

[vi] Studies that used multiple waves of international student assessments and exploited the longitudinal variation across countries over time are still rare. Brunell and Rocco (2013) estimated models with country fixed effects using multiple PISA waves (2000–2009) to study the effect of immigration on the performance of native-born students, albeit using only country-level data, Hanushek, Link, and Woessmann (2013) used individual PISA data (2000–2009) to study how changes in school autonomy impacted change in student achievement, and Rosén and Gustafsson (2016) used data from PIRLS (2001–2006) and the Reading Literacy Study (1991–2001).

[vii] Controlling for observed covariates also controls (at least partly) for unobserved covariates insofar as they are correlated with observed covariates (Stuart, 2010). The most significant remaining issue is time-invariant confounders that are correlated with change in preschool enrollment ratio as well as with change in achievement but not with the change in the observed covariates.

[viii] An alternative approach to test for effect heterogeneity is a multilevel model with a cross-level interaction between the preschool and family background. The basic idea is to regress the (random) slope of the individual level regression of achievement on family background on the country-level preschool measure. However, estimating this parameter based on longitudinal within-country variation expands equation (5) not only by adding the interaction between family

background and preschool but also by adding additional sets of interactions between family background and fixed effects for countries and years. Obviously, adding the interactions for fixed effects almost doubles the number of model parameters on the country level. Due to the limited number of country-level observations (in PIRLS the number of parameters exceeds the number of country-by-year observations), we preferred the sample split.