Policy Research Working Paper 9723

Is Investment in Preprimary Education Too Low?

Lessons from (Quasi) Experimental Evidence across Countries

Alaka Holla
Magdalena Bendini
Lelys Dinarte
Iva Trako

WORLD BANK GROUP
Development Research Group
Education Global Practice
&
Development Impact Evaluation Group
June 2021

Policy Research Working Paper 9723

Abstract

A large body of evidence suggests sizeable improvements in cognitive and social-emotional skills and subsequent educational attainment following preprimary education interventions as well as increases in earnings later in life. Yet, while the world has nearly reached universal primary education, coverage of early childhood education is still low in many countries. This study uses a novel global dataset of effect sizes from more than 50 studies conducted in 19 countries to examine measures of school participation, cognitive skills, social-emotional skills, and behavior, both during and after preprimary ages. Estimates from meta-regression analysis suggest both strong demand for preprimary services when offered and significant improvements in children's cognitive skills (0.15 sd) and their executive functioning, social-emotional learning, and behavior (0.12 sd) during the pre-primary period. Moreover, our meta-analytic results indicate statistically significant persistent advantages (0.07 sd) in each type of skill beyond the preprimary period. Pooled heterogeneous treatment effects also suggest higher gains for disadvantaged children. Lastly, cost-benefit analysis using studies from low- and middle-income countries implies benefit-to-cost ratios ranging between 1.7 and 14.2, suggesting high returns to preprimary investments even in contexts with limited state capacity.

This paper is a product of the Development Research Group, the Education Global Practice, and the Development Impact Evaluation Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at aholla@worldbank.org.

Is investment in preprimary education too low? Lessons from more than 50 (quasi) experimental studies around the world*

Alaka Holla† Magdalena Bendini‡ Lelys Dinarte-Diaz§ Iva Trako**

Originally published in the Policy Research Working Paper Series on *June 2021.* This version is updated on *March 2022.*
To obtain the originally published version, please email prwp@worldbank.org.

Keywords: Preprimary education, preschool, meta-regression, cognitive and non-cognitive skills

JEL Classification: I20, I26, I28

* We would like to thank Kathleen Beegle, David Evans, Emanuela Galasso, Amer Hasan, Marcus Holmlund, Katherine King, Arianna Legovini, Florencia Lopez Boo, Dana McCoy, David McKenzie, Emma Näslund-Hadley, Berk Ozler, Sharon Wolf, and Hirokazu Yoshikawa for their valuable comments and suggestions. We are also grateful to Steffanny Romero Esteban, Yilin Pan, Luiza Cardoso de Andrade for their assistance in the meta-regression and benefit-cost calculations. This research benefitted from funding from the World Bank's Strategic Impact Evaluation Fund. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

† Senior Economist. Human Development Vice Presidency. The World Bank. email: aholla@worldbank.org

‡ Senior Economist. Education Global Practice. The World Bank. email: mbendini@worldbank.org

§ Economist. Development Research Group. The World Bank. email: ldinartediaz@worldbank.org

** Economist. Development Impact Evaluation Group (DIME). The World Bank. email: itrako@worldbank.org

1. Introduction

Studies in high-income countries have shown preprimary education provided to children between the ages of 3 and 6 has high returns both in small-scale interventions like the Perry Preschool Project (Heckman et al., 2010) and larger-scale government programs like childcare expansions in Norway (Havnes & Mogstad, 2011) and the Head Start program in the United States (Carneiro & Ginja, 2014; Kline & Walters, 2016; Bailey et al., 2020b). Studies have also estimated gains in cognitive and social emotional skills following modifications to curricula and pedagogy in existing preprimary classrooms (Barnett et al, 2008 and Clements et al, 2013) or changes in teacher quality (Araujo et al, 2016). In a comparative welfare analysis of 133 policies in the United States, Hendren and Sprung-Keyser (2020) find that investments in health and education for low-income children pay for themselves in the long-run.

Recent global evidence also shows a strong positive association between participation in preprimary education and child development measures, and this relationship is even stronger in low- and middle-income countries (McCoy et al, 2021). Evaluations in low- and middle-income countries also demonstrate sizeable improvements in skills and subsequent educational attainment following preschool interventions that expand access to preprimary education or improve the quality of existing programs (Berlinski et al., 2008; Martinez et al., 2017; Ganimian et al., 2021; Dean & Jayachandran, 2020). Yet, while the world has nearly reached universal primary education, coverage of early childhood education is not universal in many high and middle-income countries and falls below 20 percent on average in low-income countries (UIS, 2018).

Determining whether preprimary education investments are falling short needs to factor in both the benefits as well as the costs of such programs. To date, there has been no global meta-analytical study that aggregates estimates of causal impact for preprimary education or consistently calculates benefit-cost ratios of preprimary interventions in low- and middle-income countries. This review uses a multidisciplinary body of experimental and quasi-experimental evidence on preprimary education to estimate (i) the effect of expanding access or improving quality of preprimary education, separately for studies in high-income countries and studies in low- and middle-income countries, and (ii) benefit-cost ratios based on studies in low- and middle-income countries that report the per-child costs of the interventions under evaluation. Drawing on 798 effect sizes from more than 50 studies conducted in 19 countries, we examine measures of school participation, cognitive skills, social-emotional skills, and behavior, both during and after preprimary ages.

As Elango et al (2016) argue, evaluating preprimary programs solely on their potential to improve cognitive skills risks overlooking important drivers of the long-term impacts observed in adulthood, such as social-emotional skills (Heckman, Pinto, and Savelyev, 2013). For this study, we extracted all outcomes related to school participation and progression and skill formation from studies in our sample. School participation impacts of expanding access to preprimary education not only provide an indication of the strength of an evaluation's first stage, but they also reflect demand for the services under evaluation, which may be low if parents do not value the services more than their status quo arrangements, which might include care in the home or private daycare (Bouguen et al., 2018).

During and after the preprimary years studies tend to focus on cognitive skills (such as literacy and numeracy) and noncognitive skills (such perseverance, attention, and motivation). There is little uniformity, however, in how studies measure these skills. Moreover, most studies tend to report estimates for multiple measures of the same type of skill (for example, they may use two different instruments to measure early literacy). To handle this heterogeneity and to exploit all the information in studies that estimate effects using multiple outcome measures with varying levels of precision, we standardize the outcomes we extracted from studies following Lipsey and Wilson (2001). Then, we use meta-regression methods to aggregate evidence across studies and estimate an average effect size for each class of outcomes, effectively weighting study-specific average treatment effects by their precision. Specifically, we use robust variance meta-regression, which adjusts the standard error of the aggregate average effect size for potential dependence among outcomes coming from the same study (Hedges, et al., 2010; Tanner-Smith & Tipton, 2014; Tanner-Smith, et al., 2016; Tipton, 2015; Tipton et al., 2019).

Although we rely on our own judgment to screen studies for inclusion in our sample, meta-regression offers a way to assess whether an intervention (or class of interventions) works across contexts that does not require the analyst to engage in subjective weighting of positive, negative, and statistically insignificant effects as is done in more narrative reviews. It also does not require the analyst to engage in heuristics like "vote counting," which tallies the number of positive, negative, statistically significant, and statistically insignificant results but ignores information contained in the size and precision of estimated effects.

The average effect sizes we estimate for outcomes related to school participation and progression and children's skills during the preprimary years and beyond suggest high returns to increasing investment in preprimary education and to targeting disadvantaged populations for investment. When experimentally offered access to preprimary education, children are on average 1.4 standard deviations (sd) more likely to participate in the preprimary program than their control group counterparts, suggesting strong demand for services. This is approximately equivalent to a 30 percentage point increase in preprimary school participation. During the pre-primary period, children's cognitive skills (in language, literacy, math, and other general skills) improve by an average of 0.15 sd, with larger improvements in literacy (0.22 sd) and math (0.22 sd) compared with language (0.10 sd) and general skills/indices capturing multiple domains (0.08 sd). During this period, children's non-cognitive skills (their executive functions, social-emotional skills, and behavior) improve by an average of 0.12 sd. Effects are similar (and statistically indistinguishable) across high- and low- & middle-income countries.

The average effects pool effects from studies examining impacts relative to different counterfactual scenarios. Sometimes the intervention under evaluation introduced preprimary education where the only other option was care in the home; sometimes the intervention expanded the set of facility-based options or attempted to improve the quality of existing services. We find larger average effects among evaluations assessing interventions that aimed to improve the quality of existing preprimary education services than among evaluations of expansions in services. The average effect on cognitive skills, for example, was approximately 0.2 sd among evaluations of quality improvements compared to 0.1 sd among evaluations of expansions.

Our estimates of average effect sizes also reflect average effects across small-scale pilots implemented by non-governmental organizations and nationally scaled government programs; they also average effects across contexts where teachers with formal qualifications teach small class sizes for a full day and contexts where teachers with only secondary education and a couple of weeks of training manage large classes for half a day. Insufficient reporting of program implementation features and contexts, as well as the sample size of studies, preclude testing for heterogeneous average effects along these dimensions.

A focus solely on impacts observed during the preprimary period could underestimate the full benefits of improving skills during the early childhood. Investments in preprimary education could lead to improvements that extend beyond the pre-primary period (Kline & Walters, 2016)

– even if some impacts may fade out during primary school only to materialize later in life (Chetty et al, 2011; Elango et al, 2016). Only a subset of studies in our sample had longitudinal designs, but this more limited sample shows persistence of these skill advantages beyond the preprimary period, with significant average gains of 0.07 sd in both cognitive and non-cognitive skills. These smaller effects highlight an advantage of meta-analyses: they can offer greater statistical power to detect impacts than individual studies, particularly small-sample studies with longitudinal designs. An even smaller subset of studies, all in high-income contexts, can track children into adulthood, and while some individual studies do find large and statistically significant advantages in both adult health and labor market outcomes (for example, Carneiro and Ginja, 2014; and Bailey et al., *forthcoming*), we lack sufficient statistical power to estimate average effects with precision.

To consistently estimate longer-term returns for a subset of studies from low- and middle-income countries that report unit costs for the interventions under evaluation, we combine study-specific estimates of impacts on skills and unit costs with data on average wages and real wage growth and estimates from the literature on the extent to which improvements in childhood skills translate into earnings. Our most conservative estimates suggest benefit-to-cost ratios above 1, ranging from 1.7 to 14.2. Less conservative estimates, where we assume a discount rate and return to cognitive skills for low- and middle-income countries similar to values typically used in economic evaluations of programs in high-income countries, suggest benefit-to-cost ratios ranging from 3.5 to 103.5. Both sets of ranges can be considered lower bounds for the returns to preprimary investments as they focus on benefits solely related to the future earnings of children. They do not consider any "fiscal externalities" in the sense of Hendren and Sprung-Keyser (2020), such as the increases in tax revenues and decreases in transfer payments (public assistance) that might accompany increases in children's lifetime earnings. Our estimates of the rate of return also ignore any contemporaneous benefits of preprimary education that might arise from increases in female labor force participation (Evans et al., 2021).

While very few studies permit a comparison of effects across subpopulations, the limited set of studies that report disaggregated effects is sufficient to detect statistically meaningful differences between populations of high and low socioeconomic status. Children from more disadvantaged backgrounds who are exposed to the preprimary interventions evaluated in the sample show significantly higher responses when we pool outcomes.

Our results are consistent with other reviews of subsets of the preprimary experimental and quasi-experimental literature that find significant positive impacts on cognitive skills during early childhood (Nores & Barnett, 2010; Duncan & Magnuson, 2013; Yoshikawa et al., 2013; McCoy et al., 2017; van Huizen & Plantenga, 2018) and larger impacts for disadvantaged children (Yoshikawa et al., 2013; van Huizen & Plantenga, 2018) by using narrative, vote-tallying, and meta-analytic techniques to aggregate evidence. Our review is unique in that we limit our focus to center-based preprimary education programs (rather than more holistic packages of services that intervene at earlier stages and often include intensive, home-based coaching for parents).

In contrast to other studies and reviews that document fade-out (Yoshikawa et al., 2013; Duncan & Magnuson, 2013), we find persistence of an advantage in cognitive skills after the preprimary period, although it is possible that individual studies (and therefore more narrative or vote-tallying reviews) lack sufficient statistical power to detect effects as small as 0.07 sd. Persistence in an advantage in executive function and social-emotional skills, however, is consistent with the hypothesized mechanism behind "sleeper effects," or gains in health and earnings observed in adulthood when test score advantages dissipate or disappear after the preprimary period (Heckman et al., 2010; Bailey et al., 2020a; Cascio & Schanzenbach, 2014).

Finally, our results, along with current coverage rates, suggest that levels of preprimary investment may be suboptimal, particularly in low- and middle-income countries. When offered services, families do elect to send their children to preprimary education centers. The significant average effects for cognitive and social-emotional skills and the lack of any differences across high and low- and middle-income countries show that preprimary education does provide better conditions for the development of skills than children's counterfactual options of remaining at home or in informal care settings, even in contexts with low state capacity. This improvement in skills could help counter the inequities in skill development that children bring with them when they enter primary school (Fernald et al., 2011; Naudeau et al., 2011; Duncan and Magnuson, 2013; Schady et al., 2014). The high benefit-to-cost ratios indicate high returns to preprimary investments, and the persistence of impacts beyond the preprimary period suggests that investments in preschool can play a role in improving the effectiveness of primary education (Johnson & Jackson, 2019), possibly by enhancing executive functions and social-emotional skills that help children learn (Cunha and Heckman, 2007) or by reducing heterogeneity in the classroom and thus making it easier to teach foundational skills (Banerjee et al., 2007; Dupas et al., 2011).

The rest of this paper is organized as follows. The next section (Section 2) provides background, describing skill development during the preprimary period, theories on when preprimary education should improve skills and well-being, and preprimary coverage and quality in low- and middle-income countries. Section 3 describes the methods we used to search for studies, screen them for inclusion, and extract data on effect sizes. This section also describes the studies, interventions, and outcomes included in the analysis, as well as the methods we use to aggregate evidence across studies and to test our research questions related to the demand for preprimary education, skill improvements during the preprimary years, the persistence of any advantages beyond the preprimary years, and the returns to preprimary investments. Section 4 presents the main results and those by socioeconomic status. Section 5 discusses the implications of these results for preprimary investments and concludes. Various appendixes provide more detailed information on our methods and studies and present study-specific estimates of the standardized average treatment effects that we use in our main analyses.

2. Background and context

2.1 Learning during the preprimary years

Research in neuroscience, psychology, and developmental cognitive science has established that, due to higher brain malleability at younger ages, children learn fast in their first six years of life compared to later stages of development (Shonkoff et al., 2000; Knudsen, 2004). They rapidly acquire skills in areas of knowledge that include numbers (DeWind, 2019) and language (Wang et al., 2020; Yuan & Fisher, 2009) as well as social interactions (Hamlin et al., 2013; Tamis-LeMonda et al., 2008; Tomasello, et al., 2005). Before primary school, children also develop skills including executive functions (inhibition, working memory, and cognitive flexibility) that allow them to “learn to learn,” manage emotions, and relate to others (Diamond, 2013; Zelazo et al., 2003). As a result, most research on the development of children during the preprimary period measures skills in “domains” of development that include cognitive skills (like language, literacy, and numeracy), executive functioning, and social-emotional skills (Fernald et al., 2017).

Evidence from neuroscience and developmental psychology also demonstrates that exposure to learning opportunities promotes children’s innate learning abilities (Jara-Ettinger et al., 2016; Wang et al., 2020). Learning appears to be sequential and cumulative as early, more basic skills provide the basis to master later and more complex skills (Knudsen et al., 2006; Spelke & Shutts, forthcoming). Learning is also interdependent; learning in specific domains promotes learning in other domains and results in complex skills such as literacy (Dehaene 2010; Bailey et al., 2020a). Cunha and Heckman (2007) argue that due to the cumulative and interdependent nature of learning, the establishment of sound foundational skills early in life can lead to a virtuous cycle of skill acquisition as children develop.

Whether preprimary education facilitates this process in practice, however, depends on the quality of the learning environment provided through pre-primary education and how it compares to what children would have experienced without formal preprimary education—both during and after the pre-primary period. If children receive more cognitive and psychosocial stimulation from their interactions with caregivers in the home or informal care settings (for example, the homes of relatives or friends) than informal pre-primary education, then an expansion of formal services is unlikely to improve average skill development and may even set children back. That is what researchers suggest happened when Canada's province of Quebec extended subsidized coverage of childcare to less needy families (Baker et al., 2008; Baker et al., 2015). If, on the other hand, the learning environment offered through pre-primary education services provides more stimulation to children than their counterfactual situation, then coverage expansions or quality improvements will likely enhance skill development (Cascio, 2015; Cascio & Schanzenbach, 2014).

Similarly, if available preprimary education substitutes for other services that children would otherwise access, then investments in preprimary education may not translate into improvements in children's skills. For example, in Denmark, children who had already benefited as infants and toddlers from a nurse home visiting program exhibited smaller gains in adulthood on an index of human capital from an expansion in preprimary education than children who did not have the home visiting program (Rossin-Slater & Wüst, 2020). Similarly, a preprimary school construction program in Cambodia induced parents to switch their underage children out of formal primary school and into the new community-based preprimary schools (Bouguen, et al., 2018).[1] The federal Head Start program in the United States also induced switching out of private preprimary schools (Kline & Walters, 2016).

The inadequacy of stimulation in their counterfactual environment and lower access to substitute services may partially explain why children from more disadvantaged backgrounds tend to benefit more from preprimary education (Cascio and Schanzenbach, 2014; Currie, 2001).

2.2 Preprimary education in low- and middle-income countries

According to UNESCO's Institute of Statistics, preprimary enrollment has increased considerably around the world over the last two decades, from an average of 30 percent of children in 2000 to 50 percent in 2018.[2] Access to preprimary education in low- and middle-income countries is still low, with 19 percent of preprimary-aged children in low-income countries enrolled (UIS, 2018), a coverage rate less than half of what was observed in high-income countries fifty years ago. Beyond these averages, household survey data using UNICEF's Multiple Indicator Cluster Surveys (MICS) suggest substantial variation in preprimary enrollment across and within countries that is associated with socioeconomic status, with the largest differences in enrollment in the poorest countries (Figure 1). This unequal access to preprimary education can exacerbate learning inequalities, as children in families from lower socioeconomic groups tend to also have limited learning opportunities and stimulation at home and in their communities (McCoy et al., 2018) (Figure 2).

Domestic financing for preprimary education has increased over the past decade, amounting to 6.6 percent of domestic education budgets globally. Low-income countries allocate substantially less, with less than 2 percent of their education budgets going toward preprimary education (UIS, 2018). In these countries, standards and quality assurance systems are often nonexistent or under-resourced and learning spaces often do not meet minimum safety and

[1] In contrast to the practice of "red-shirting" in higher-income countries, in many low-income countries, parents try to enroll their children in primary school before the children are age eligible.

[2] UIS use a variety of data sources to measure enrolment: administrative data from schools or household survey data on enrolment; population censuses and surveys for population estimates by single year of age (if using administrative data on enrolment); administrative data from ministries of education on the structure (entrance age and duration) of early childhood education. Data on provision are limited but suggest that around 38 percent of countries offer some free preprimary education, with most of these countries offering between one and three years of service. There is also wide variation of private provision across regions, ranging from 7 percent in Europe and Central Asia to 72 percent in the Middle East and North Africa (UIS, 2018).

sanitation conditions (World Bank, 2013; UNICEF, 2019). Due to an acute shortage of preprimary teachers and staff, child-teacher ratios remain high even when enrollment rates are low (UNICEF, 2019). The preprimary education workforce also has lower renumeration and higher attrition rates than what is found among the workforce in primary education (UIS, 2018).

Thus, on the one hand, the low levels of stimulation occurring at home and very limited access to preprimary education in low-income countries suggests that offered preprimary education services must exceed a low bar in terms of cognitive stimulation and psychosocial support to promote children's skill development and learning. On the other hand, countries with limited state capacity to provide and regulate school quality may struggle to meet even this low bar.

3. Methods

To assess the average effects of preprimary education interventions, we conducted a global systematic review of the experimental and quasi-experimental literature focused on preprimary education. We then used meta-regression methods to aggregate evidence across studies. This section describes our process for selecting studies to use in our analyses and for extracting "raw data" from the studies. We also lay out our empirical specifications and describe how we standardize extracted coefficients across studies and map study-specific outcomes into broader categories that we can use in meta-regressions.

3.1 Data

To identify studies to review, we followed standard practice for systematic reviews and proceeded in three iterative stages: (i) search and application of inclusion criteria, (ii) screening (two rounds), and (iii) data extraction. Figure 3 summarizes the entire process and indicates the number of studies that remained in the review sample after each stage.

3.1.1 Search strategy and selection criteria

To start, we conducted a systematic search for studies using relevant keywords and terms in several search engines and databases, as well as known portfolios of experimental and quasi-experimental studies, such as the portfolio of the World Bank's Strategic Impact Evaluation

Fund.[3] Table 1 Panel A lists the databases and specific search terms we used. In this phase, we looked for studies published before 2021 that stated an aim to estimate the effects of preprimary interventions. Keywords related to experimental and quasi-experimental evaluation designs effectively limit our attention to studies published after 2007.

In addition to database searches, we found studies through two other ways. We searched bibliographies of papers found through the database search. We also contacted experts and researchers who have frequently published on the impacts of preprimary programs in peer-reviewed journals and asked for other relevant published studies to incorporate in our review.

This search process yielded a total of 270 studies, with 183 studies from database searches and 87 additional studies from other search methods. These studies included systematic reviews, experimental and quasi-experimental studies, and studies that did not employ experimental or quasi-experimental methods (for example, cost analysis, qualitative reports, etc.).

Table 1 Panel B describes the inclusion criteria we used to identify studies for review. We sought to include experimental and quasi-experimental studies of preprimary programs that provided group-based childcare in a center setting with a developmental or educational focus for children between the ages of 3 and 6 years. This definition of preprimary education encompasses the provision of formal preprimary, community-based preprimary, kindergarten, pre-k, and daycare with an educational component, as well as interventions that make preprimary education more affordable (such as subsidies) or increase its quality (such as teacher training, the introduction of a new curriculum or pedagogical approach, and provision of materials). We did not include programs that also provided services outside the preprimary period, such as childcare during infancy or classroom quality improvements after kindergarten, as it would be impossible to distinguish the impacts of interventions during the preprimary years from those of interventions that spanned a longer period. This restriction ruled out seminal studies of programs like the Abecedarian Project (Campbell et al., 2014; Conti et al., 2016) and Project STAR (Chetty

---

[3] The Strategic Impact Evaluation Fund supports research that estimates the effects of programs and policies to improve education, health, access to quality water and sanitation, and early childhood development in low- and middle-income countries. See here for more information on the fund and a list of evaluations in its portfolio.

et al., 2011; Krueger, 1999), as well as large-scale expansions in childcare that also covered infants (Baker et al., 2008; Baker et al., 2015; Bernal, 2019; Bernal and Ramirez, 2019).[4]

For our purposes, experimental (randomized controlled trials or RCTs) or quasi-experimental designs had to isolate the causal impact of preprimary programs or policies on children's outcomes. Children's outcomes had to include either a measure of school participation or progression, or a measure of children's skills or development. This measurement could take place either when children were still in preprimary education or after they had progressed to higher levels of education. We also extracted outcomes observed when the beneficiaries of preprimary programs were adults, although as we will discuss in later sections, the number of outcomes and studies reporting wellbeing in adulthood was too small to estimate average effects with precision.

Included studies also had to be published in peer-reviewed journal articles or in a formal working paper series, such as the working papers of the National Bureau of Economic Research or the Policy Research Working Paper series of the World Bank. We included technical reports only if they included a suggested formal institutional citation.[5]

3.1.2 Screening

We then screened these studies to verify that they met our inclusion criteria, focusing on the credible estimation of causal impacts of preprimary interventions and the inclusion of outcomes that captured preprimary coverage and/or children's skills or wellbeing.

Figure 3 describes the screening process. We first eliminated studies based solely on a review of their citations, abstracts, and introductions. Of the 270 studies identified in the initial search, we excluded 111 studies during this stage because (i) they were not published in peer-reviewed journals or did not appear as part of a working paper series; (ii) they were not experimental or quasi-experimental studies; (iii) they assessed early childhood interventions that targeted children outside of the 3 to 6 year age range, (iv) they measured outcomes unrelated to

[4] We did include a subset of effects from Baker et al., (2008) – specifically those restricted to 4-year-olds in the post-period survey – who were unlikely to have access to subsidized care before age 3 since the childcare expansion started with older children.

[5] One implication of not including unpublished studies is the potential for publication bias—that is, a bias toward positive and significant results. On the other hand, Brodeur et al. (2020) find that randomized control trials exhibit less of this bias than methods like instrumental variables and differences-in-differences, and more than 70 percent of the studies we eventually included were randomized control trials.

learning and well-being for affected children at some point in their life cycle; (v) they were duplicates (for example, working-paper versions of published studies); or (vi) they had no relevance to our research questions. This first stage of screening left us with a total of 159 studies.

In the second stage of screening, at least two authors of the current study reviewed each identified study, reading the full text and assessing features meant to proxy for study quality, such as the use of an evaluation design that would generate causal impacts (RCT, regression discontinuity design, differences-in-differences estimation, matching, or an instrumental variable strategy) and the presentation of information on issues that could compromise evaluation design.[6] Specifically, to be included in our final sample, studies needed to (i) isolate the impact of the program using some sort of comparison group; (ii) present sufficient evidence that the experimental groups in an RCT were balanced on a set of characteristics prior to the intervention; (iii) report sample attrition and compliance to the treatment assignment for the case of RCTs; (iv) experience less than 30 percent attrition in either the treatment or comparison group between the initiation of treatment and measurement; (v) report the precision of the estimated effect size by providing either standard errors, confidence intervals, t-statistics, standard deviations, sample size and/or p-values; and (vi) use well-known or established outcome variables and tools to measure them. For issues like attrition and compliance, we could not exclude studies based on how authors addressed them, as approaches vary across disciplines.[7] Some studies omitted key design and estimation details such as balance checks and the presentation of the exact precision of estimated impacts. When related analyses already published in earlier work provided this information, we kept the study in the review if it met our other criteria.

Taking all screening criteria into account, each reviewer assigned studies to two categories—those to include in the review and those to exclude. When reviewers' ratings matched, a study was either automatically included or removed from consideration. All four authors discussed ratings discrepancies to arrive at a conclusion. Of the 159 studies that

[6] We did not include studies that used sibling fixed effects to isolate the impact of a program on children's outcomes, which excluded studies of the Head Start program such as Garces et al. (2002) and Deming et al., (2009).

[7] For example, papers published by economists typically document attrition, note whether it is differential across treatment and comparison groups, and address it by calculating bounds for average treatment effects using the methods of either Lee (2009) or Horowitz and Manski (2000). Papers written by developmental psychologists, however, tend to address sample attrition through multiple imputation methods and then calculate average treatment effects.

underwent this more intense screening, we excluded 105 studies largely because they failed to credibly identify causal impacts or to report sufficient information for us to calculate an effect size or its precision. Thus, this second phase of screening yielded a final sample of 54 studies that moved to the next phase of data extraction.

3.1.3 Data extraction

We extracted three levels of information from each study. *Studies* are defined as independent publications concerning preprimary programs. *Interventions* refer to the studies' different experimental arms (in an RCT) or implied treatment and comparison groups (in a quasi-experimental study). For example, a study might provide one group of participants with a preprimary program, a second group of participants with the preprimary program and an accompanying program targeting parents, and a third group with no service. During data extraction, we treated these as three different intervention groups or contrasts. *Outcomes* refer to the estimated coefficients corresponding to the average treatment effects we extracted for each outcome for each intervention. The resulting dataset includes 54 studies, 141 interventions or contrasts, and 798 outcomes.

3.1.3.1 ***Studies***

Figure 4 maps the locations of studies, demonstrating coverage across 19 different countries with a concentration of research in the United States. Table 2 presents characteristics of the studies in our sample.[8] We see that 85 percent of the studies we reviewed appeared in peer-reviewed journals, and around half appeared in journals that publish research in the field of Economics. More than 70 percent of studies were RCTs. Close to half evaluated a program that expanded coverage of preprimary education, while the other half focused on programs that aimed to improve the quality of existing preprimary education services.

3.1.3.2 Interventions

We also extracted information about the evaluated interventions. Table 3 summarizes characteristics of the treatment arms of the included studies for our full sample and separately for high-income countries and low- and middle-income countries.[9]

---

[8] Appendix Table 1 presents additional details of each study, including intervention components and evaluation design

[9] Appendix Table 2 presents additional characteristics of the programs evaluated in each country.

We assigned interventions to *at least* one of 12 intervention categories.[10] Around 60 percent of programs included a component focused on teachers' professional development, with more of a focus on this kind of component in low- and middle-income countries. This is not surprising as many interventions that aim to improve the quality of preprimary education—such as a change in curriculum or pedagogy—will require teachers to receive some training prior to program implementation. Similarly, offering a new service will also initially entail training. Close to half of all programs provided subsidized or free access in studies from high-income countries, compared to only 3 percent among studies from low- and middle-income countries. In both contexts, around 70 percent aimed to target a disadvantaged population.

These interventions include both small-scale pilots implemented by non-governmental organizations (NGOs) and nationally scaled government programs. Daycare programs with education components represented very few evaluated programs; formal preschool was the dominant program in both high-income and low- and middle-income contexts, although community preschool accounted for more than 40 percent of programs in low- and middle-income contexts. While a majority of programs in both contexts took place in publicly managed schools, a sizeable fraction had a combination of public and private management. A higher fraction of programs in studies from high-income countries took place in classrooms with a teacher who was formally qualified to teach preprimary students than in studies from low- and middle-income countries, where teachers were formally qualified to teach in preprimary settings in only half of the studies in the sample.

3.1.3.3 Outcomes

We focused on outcomes related to school participation and child development, extracting average treatment effects and their precision for each intervention (Table 4). When available, we also captured outcomes of teachers, such as classroom practice or self-reported measures of wellbeing. We extracted a total of 798 outcomes at the child and teacher levels. In one study that relied on a very large administrative dataset with a panel structure (Rossin-Slater & Wüst, 2020), outcomes were aggregated to combinations of geographic unit and birth cohort. Sample sizes

[10] These categories included (i) teacher professional development, (ii) subsidized or free access, (iii) change in curriculum, (iv) change in pedagogy, (v) provision of materials, (vi) provision of new staff, (vii) provision of health and nutrition services, (viii) preschool construction, (ix) parental engagement, (x) community outreach, (xi) preschool day extension, and (xii) teacher payments.

associated with extracted outcomes ranged from 123 to 22,480,000 individuals, with a median of 2,449.

Measures of precision collected from the studies included standard errors, confidence intervals, and p-values. When only stars or other symbols indicated precision in a table or chart and the study provided no other metric of precision, we assumed p-values corresponding to the highest p-value of the interval indicated by the symbol.[11] If a paper presented no standard measure of precision, we extracted information on the sample sizes and standard deviations of the outcome for the treatment and control groups to calculate the standard error of the average treatment effect.

For each outcome, we included both average treatment effects measured on the full study sample and those measured separately for different socioeconomic groups and different age groups, as these tend to be the characteristics of children used to target preprimary programs when universal coverage is not an option.[12] We aimed to extract intention-to-treat estimates of impact, although we extracted local average treatment effects or treatment-on-the-treated effects when intention-to-treat effects were unavailable or less relevant.[13] Three-quarters of extracted outcomes were not disaggregated by socioeconomic status. Appendix Table 3 lists the different subpopulations for which we extracted separate coefficients plus a mapping of these subpopulations to a more aggregate indicator meant to capture high and low socioeconomic status.

We classified outcomes according to the timing of their measurement, tagging them as occurring during the preprimary period (below age 6), during post-preprimary education (age 6 to 18), or during adulthood (after age 18). The majority (63 percent) of extracted outcomes were measured when children were still in the preprimary period, while 37 percent, all from high-income countries, were measured afterwards.

---

[11] For example, if ** indicated $p<0.05$, we assumed a p-value of 0.04. We could not use the exact threshold of 0.05 as that led us to misclassify effects that were statistically significant in their respective studies as insignificant when standardized.

[12] For two studies (Carneiro & Ginja, 2014; Heckman et al., 2010), we had to extract effects disaggregated by gender as full sample estimates were not reported.

[13] For example, in Ganimian et al. (2021), the evaluated intervention entailed an improvement in existing services. Thus, while all children living around the preprimary education center could access the services, only those regularly attending the services would be exposed to the treatment.

We scored each outcome as *positive* or *negative* depending on whether increases represented improvements or reductions in welfare. We also mapped the specific outcomes measured in each study (for example, receptive vocabulary using the Peabody Picture Vocabulary Test) to more aggregate domains that could be used in our subsequent analysis (for example, language), following standard domains observed in the early childhood education literature (Fernald et al., 2017). Specifically, we classified all outcomes that measured attendance, enrollment, years of education, and schooling attainment into a single category meant to capture *school participation and progression*.

We mapped all outcomes that measured children's skills and knowledge about specific content areas (for example, mathematics, language, and literacy), as well as dispositions and skills that help children to think about and understand the world around them (for example, general intellectual ability) into one of the following categories: *literacy, language, math,* and *general* cognition. In our analyses, we classify these domains as cognitive skills.

We classified constructs such as attention, inhibition, and working memory as *executive functions;* outcomes such as social cognition, social competence, and emotional regulation as *social-emotional skills*; and reported or observed measures of aggression, internalizing and externalizing behavior problems, conduct, and disciplinary actions, as well as incarceration as *behaviors*.[14] We refer to executive functioning, social-emotional skills, and behaviors as non-cognitive skills.

Health outcomes encompassed *physical health* and *motor development.* Lastly, we classified teachers' responses in the classroom, such as the extent of emotional and pedagogical support they provide their students, as *teacher practices*.

One-to-one mapping of outcomes to domains was not always straightforward since some measures used in studies were composite scores that included items from more than one related domain – for example, both language and literacy or both social-emotional skills and behavior. To address such difficulties with classification and to manage potential multiple inference problems that could arise as the number of outcome categories gets larger, we focus on three main categories: school participation and progression, cognitive skills, and non-cognitive skills. We

[14] Appendix 1 describes the contents of each skill category in more detail.

report effects on teacher practices and health outcomes in an appendix. Though the sample sizes for these outcomes are adequate, far fewer studies include outcomes from these categories.

When a study estimated treatment effects using more than one measure for a particular domain of development (for example, the authors used both a measure of separation anxiety and a measure of physical aggression and opposition to assess social-emotional skills), we extracted coefficients for each mode of measurement, unless one measure represented only the aggregation of all the other measures. We omitted effect sizes for outcomes where reported measurement did not conform to accepted practice—for example, asking about nonacute, nonchronic illnesses with a 12-month (instead of 2–4 week) recall period—and for outcomes with very low response rates.[15] We also omitted outcomes when it was not clear if an increase in the measured outcome represented an improvement or decrement to well-being. Take, for example, outcomes such as grade retention in kindergarten or body mass index. Grade retention in kindergarten could occur when a child's learning suffers in the classroom (a bad outcome) or when teachers become more attuned to children's readiness to progress to first grade (a good outcome). An increase in body mass index for an underweight child in low-income countries would represent progress, but a decrease in body mass index for an overweight poor child in high-income countries would also be considered an improvement.

Not all studies measured outcomes covering all domains of development, nor did studies routinely measure outcomes like teachers' behavior in the classroom. Thus, sample sizes and the composition of studies vary by aggregated outcome.

3.2 Analytical strategy

We aggregate all extracted outcomes to measure an average treatment effect across studies, using meta-regression methods that account for the potential correlation among outcomes measured from the same study and that penalize precision appropriately when the sample size of either studies or outcomes is low. We also combine our estimates of the average effects of preprimary interventions with information on the costs of these interventions and their estimated rates of return to estimate benefit-to-cost ratios.

[15] We also did not include outcomes for all reported age groups in Berlinski et al. (2008), as primary school enrollment rates suggested little room for improvement for some ages. For this study, we selected age groups that matched with the age composition of other studies in our sample that reported similar outcomes.

3.2.1 Robust variance meta-regression

Many reviews of evidence, particularly those focused on low- and middle-income countries, do not estimate an average effect across studies or incorporate the uncertainty associated with estimated effects in their assessments of the benefits of a certain intervention or class of interventions. For example, authors may plot effect sizes in a bar chart and use their expert judgement to determine whether an intervention does or does not improve outcomes (Kremer & Holla, 2009; Evans & Popova, 2016; Evans & Yuan, 2021). Others may employ a vote counting method, tallying the number of positive and significant coefficients. These methods, however, do not fully consider the fact that estimated coefficients come with confidence intervals of different sizes that inform us about the precision of estimated impacts. That is, even if two estimated coefficients are both significantly different from zero, they are likely to have different levels of precision. Moreover, when a literature does include many statistically insignificant effects, we might worry about inter-rater reliability when human judgment adjudicates whether the average impact of an intervention appears positive (or negative) overall.

The education, health, and psychology literatures try to solve these issues through meta-regression, where estimated coefficients y_j for studies j = 1, … J are the units of observation and the goal of the analysis is to estimate the average effect size, β_0, in a model such as

$$y_j = \beta_0 + \boldsymbol{x_j}\boldsymbol{\beta} + u_j + e_j \qquad (1)$$

where e_j is the study-level residual, u_j is a study-level random effect, and β_0 and any other coefficients in vector $\boldsymbol{\beta}$ associated with covariates in vector $\boldsymbol{x_j}$ are estimated through weighted least squares, with weights coming from the (inverse of) standard errors s_j associated with each coefficient y_j (Tanner-Smith et al., 2016; Tipton et al., 2019).[16] The precision of the average effect size β_0 is measured under the assumption that all estimated coefficients included in the regression are independent draws from the potential distribution of outcomes.

[16] Some meta-analyses are not only interested in measuring the average effect β_0 across studies but also (i) the effects $\boldsymbol{\beta}$ of any study characteristics $\boldsymbol{x_j}$ that might determine the size of a study's average effect and (ii) heterogeneity in effects observed across studies that is not due to sampling variation. Meager (2019), for example, uses hierarchical linear modeling to jointly estimate both this sampling variance and heterogeneity for studies of microcredit interventions to assess external validity of estimates and the extent to which we can generalize treatment effects from one context to another. Vivalt (2020) has a similar goal for a larger set of randomized controlled trials with multiple objectives.

Since we had no reason a priori to select one outcome over another when studies reported effects using multiple measures, we extracted all outcomes. Thus, this assumption of independence is unlikely to hold. We could have selected one effect size per study or averaged all effects within a study to create a single synthetic effect size, as in Baird et al. (2014) or Hidrobo et al. (2018). However, this could result in loss of information, as outcomes within a study are rarely perfectly correlated (and sometimes not correlated), and we would like to use as many extracted effect sizes as possible.

To deal with the potential dependence among outcomes measuring the same outcome category extracted from the same study, we use robust variance meta-regression (Hedges et al., 2010; Tanner-Smith & Tipton, 2014; Tanner-Smith et al., 2016; Tipton, 2015). This method of meta-regression uses a working model of the structure of dependence of outcomes within a study (that is, the variance-covariance structure) and does not require assumptions on the exact distribution of effect size estimates, unlike approaches such as hierarchical linear modeling that nest estimated effects within clusters. The robust variance regression method also provides unbiased estimates of the variance for the average effect size across studies even with sample sizes as low as 10 studies (Tanner-Smith et al., 2016) and adjusts the robust variance estimation (RVE) estimator and the degrees of freedom when the number of studies is small (Tipton, 2015). In particular, we use a working model of the variance-covariance structure characterized by correlated effects, assuming that study average effect size varies across studies, the effect sizes within studies are equally correlated (that is, they all have the same intracluster correlation), and this correlation arises from sampling error when multiple outcome measures are collected on the same units.[17]

Thus, we modify Equation (1) to account for multiple outcomes coming from a single study in Equation (2),

---

[17] We use the *robumeta* command in Stata developed by Hedberg (2011) and the default assumption about the correlation among outcomes of the same study (that is, $\rho = 0.80$). Another option for the working model of the variance-covariance structure is the hierarchical effects model, in which the observed effect size estimates are nested within studies that are nested within clusters, where clusters may correspond to countries, research groups, or interventions (when there are multiple studies drawn from the same intervention). We have no reason to believe that our country-income groups correspond to clusters, and while we do have some studies focused on a single intervention (for example, the Head Start experiment in the United States), this is not our dominant driver of potential dependence among outcomes. We follow Tanner-Smith and Tipton (2014) and use the model of variance that likely describes most of our data when both correlated effects and hierarchical effects may describe the dependence of outcomes of our sample of estimated coefficients.

$$y_{ij}^k = \beta^k + u_j + \varepsilon_{ij} \qquad (2)$$

where y_{ij}^k is the effect size i in study j of outcome category k; β^k is the average effect size of outcome category k; u_j is the study-level random effect such that $var(u_j)$ captures the between-study variance component; and ε_{ij} is the residual for effect size i in study j.

For the benefits of preprimary education to outweigh its costs, families must send their children to preprimary programs when they are made available and these programs must, at a minimum, improve children's skills relative to their counterfactual situations. We use Equation (2) and measures of school participation in the preprimary years to assess demand for preprimary education. We use measures of children's knowledge and skills plus teacher practices measured during the preprimary years to assess whether preprimary services can be delivered with sufficient quality to provide children a better learning environment than they would otherwise receive. In both cases, we test whether $\beta^k > 0$.

Finally, to test whether certain subsamples benefit more from preprimary interventions, we modify Equation (2) to add a covariate D_{ij}, which indicates if effect size i in study j was measured for a disadvantaged population, and we limit our sample to the set of effect sizes separately estimated for subpopulations defined by their socioeconomic status. Because so few studies reported effects for different subpopulations, we group the outcomes y_{ij}^k into two main aggregate categories of outcomes: school participation and progression, and skills (cognitive and noncognitive). For the pooled heterogeneity analysis, we regress Equation (3) as follows:

$$y_{ij}^z = \beta^z + \delta D_{ij} + u_j + \varepsilon_{ij} \qquad (3)$$

where y_{ij}^z is the effect size i in study j of the aggregated outcome category z and β^z is the average effect size of the aggregated outcome category z. Our estimated coefficient of interest in this case is $\hat{\delta}$, which indicates if the average effect size of preprimary interventions differs for disadvantaged populations relative to populations that are less disadvantaged.

3.2.2 Standardization of outcomes

Even within the outcome categories that we created, studies in our sample used different measures, each with their own scales, to assess school participation and progression, children's skills and behavior, the responses of teachers, and well-being in adulthood. Of all extracted outcomes, 62.6 percent were reported in terms of standard deviations by the researchers. Thus, before we could use treatment effects as dependent variables in our meta-regression, we needed to standardize the remaining outcomes, using other information we extracted from each study, such as sample sizes, standard errors, and the means of the treatment and control groups at follow-up. Appendix 2 details our process for standardizing both continuous and binary outcomes and for imputing values when information (like the separate sample sizes of the treatment and control group) was missing. We were able to standardize an additional 30.5 percent of all outcomes.[18]

Sometimes information required for standardization was missing – for example, separate sample sizes for the treatment and control groups or the mean of the control group at endline for each sub-population when heterogeneous treatment effects are presented. To address these issues, we relied on information provided in the text or appendices of papers as much as possible to approximate sample allocations across treatment and control groups and assumed an even allocation when papers provided no guidance. We used control means of the full sample to represent the means for each sub-population when this information was missing. Given that disadvantaged sub-populations tend to be minorities and given that their average outcomes tend to be lower than less disadvantaged sub-populations, this type of imputation should bias us against finding significant differences between children coming from high socio-economic backgrounds and those coming frow low socioeconomic backgrounds.

3.2.3 Economic evaluation

Very few studies in general can track children into adulthood to assess the full impact of preprimary education and thus appropriately capture the benefits in a benefit-to-cost ratio. Moreover, no study from a low- or middle-income country in our sample follows children into adulthood. To estimate a benefit-to-cost ratio for these countries, we need to infer what eventual benefits might be from the short-term effects that we can observe. For the subset of studies in low- and middle-income countries that do report significant improvements in children's cognitive skills or social-emotional skills as well as sufficient information to infer per-child costs for the duration of the evaluated intervention, we follow the strategy of Galasso and Wagstaff (2018) and

[18] A total of 6.9 percent of extracted outcomes were not standardized because some inputs for the standardization process were not reported.

Ganimian et al. (2021) to project what benefits might be and assess whether these projected benefits exceed reported costs.

For each study, we gather information about the country's labor force participation, average wages, and real wage growth to calculate the total expected lifetime earnings of a child.[19] We assume children will work between the ages of 22 and 65, which should be a conservative assumption for the number of active years in the labor market, given that labor market participation tends to start much earlier than age 22 in low- and middle-income countries. Because most studies reported program costs in US dollars, we first convert wage data into US dollars using official exchange rates from the International Financial Statistics of the International Monetary Fund.[20] We discount this lifetime stream of earnings to the year that preprimary program costs were incurred, using two different discount rates – 3 percent and 5 percent. While 3 percent reflects the standard in the literature on economic evaluation in high-income countries, some researchers have argued that the higher economic growth rates in low- and middle-income countries would make rates as high as 5 percent more appropriate (Haacker et al, 2020).

Next, we rescale the present value of earnings to account for the benefits of preprimary interventions, combining study-specific estimates of the improvement in cognitive skills following a preprimary intervention with estimates from the literature on the extent to which improvements in cognitive skills observed during the preprimary period translate into improvements in earnings. To be conservative, we select each study's lowest estimated treatment effect that is significantly different from zero. To translate average effects on cognitive skills into increases in earnings, Kline and Walters (2016) use a value of 13 percent per standard deviation increase in cognitive skills in their study of the Head Start program in the United States, while Galasso and Wagstaff (2018), focusing on low-income countries, use a value of 4.3 percent. Thus, for example, we can multiply the present value of earnings by (0.13 x study-specific average

[19] We omit this data collection and calculation for Ganimian et al., (2021) which directly reports benefit to cost ratios. Because of lack of information from the general population on monthly wages in Mozambique, we also cannot calculate benefit-to-cost ratios corresponding to Martinez et al., (2017) which found significant impacts on children's cognitive skills from an expansion of community preschools and did report per child costs.

[20] As studies did not indicate whether they had used official exchange rates or exchange rates adjusted for purchasing power parity to convert program costs to US dollars, we assume they used official exchange rates and therefore use official exchange rates to convert wages to US dollars.

treatment effect expressed in standard deviations) to calculate the lifetime gain in earnings from an increase in cognitive skills.

When studies also include effect sizes for social-emotional outcomes, we again extracted the lowest significant average treatment effect and augmented the estimate of individual gains in earnings with the gains implied by improvements in social-emotional skills, again using estimates from the literature of the returns to social-emotional skills (Belfield et al., 2015).

These estimates of benefit-to-cost ratios, even when using a discount rate of 3 percent and the higher value of the labor market returns to cognitive skills, should be considered conservative, as they solely capture earnings benefits accruing to the individual child. They do not capture any intergenerational transmission of human capital as documented in Rossin-Slater and Wust (2020), nor do they include any social externalities that could arise from an increase in tax revenues following an earnings increase, from a lower reliance on public assistance, from improvements in health (Carneiro and Ginja, 2014; Brotman et al., 2016) and the resulting lower burden on health systems, or from decreases in crime (Heckman et al., 2010).

4. Results

In this section, we present average effect sizes from robust variance meta-regressions. We show results for our full sample of studies and for high-income countries and low- and middle-income countries separately. We also present separate estimates for interventions that expanded preprimary education services and interventions that aimed to improve quality in existing preprimary classrooms.

We could not compute average effects when the sample size was too small to obtain trustworthy p-values, when the degrees of freedom fell below 4, or when all extracted outcomes came from a single study; cells corresponding to these situations are blank in our tables. Appendix 4 presents the figures with the corresponding study-specific standardized effect sizes by outcome category.[21]

4.1 Demand for preprimary education programs

In our review, 10 studies - equally distributed between high-income countries and low- and middle-income countries - report impacts on children's take-up of offered preprimary education

[21] Figure 1 in the Appendix summarizes the results presented in sections 4.1 and 4.2.

services. Table 5, Panel A presents the average effect sizes of the impact of preprimary interventions on preprimary school participation and progression for the full sample of studies. On average, the odds of children being enrolled in or attending preprimary education are 1.4 sd higher for children who were given access to preprimary education programs compared to those who did not have access. This is equivalent to a 30 percentage point increase in school participation.[22] Although most studies only reported participation in the program under evaluation, some also reported participation in *any* preprimary education services. Results from these studies indicate that preprimary interventions improve overall enrolment and attendance, suggesting that the improvements in program-specific participation do not solely reflect substitution away from existing services (Kline & Walters, 2016; Brinkman et al., 2017; Berkes et al., 2019; Spier et al., 2020).

4.2 Impacts on children's skills in the preprimary period

Panel B of Table 5 presents robust variance meta-regression results for cognitive skills related to language, literacy, and math, as well as for outcomes that either relate to general cognition or represent indices that combine multiple cognitive skills. Panel C of Table 5 presents results for non-cognitive skills: executive functioning, social-emotional skills, and behavior. Taken together, our results suggest that preprimary interventions improve children's skills in both high-income countries and low- and middle-income countries, with higher impacts estimated among studies evaluating quality improvements in existing preprimary programs compared to studies evaluating expansions in services.[23]

---

[22] See Appendix Figure 1 for study-specific estimates on school participation and progression during the preprimary years.

[23] Studies in the sample inconsistently reported detailed program features of the evaluated interventions or measured the quality of classroom instruction. Thus, it is not possible to identify exactly what underlies program success or failure or the persistence of impacts. Given the importance of teachers to children's skill development during the preprimary period, Appendix Table 8 presents average effect sizes of the impact of preprimary interventions on teacher practices in the classroom. Average effect sizes pooled across country-income groups suggest that preprimary interventions can improve teachers' practices in the classroom by an average of 0.473 sd (column 1). Similarly, preprimary programs, particularly if they include school meals and nutrition counselling for parents, can improve children's health during and after the preprimary period. Average effect sizes for health outcomes during the preprimary period are small and statistically insignificant (effect size: 0.032; standard error: 0.029; and N=27).

4.2.1 Cognitive skills

When we pool outcomes across all domains of cognitive skills, across country-income contexts, and across intervention goals, preprimary education interventions raise children's cognitive skills by an average of 0.15 sd during the pre-primary period (Table 5 Panel B Column 1). While the estimated average effect size is higher for studies from high-income countries in Column 2 (0.18 sd) than for studies from low- and middle-income countries in Column 3 (0.12 sd), these estimates are statistically indistinguishable (see Appendix Table 7 for the difference between the two estimates and its associated standard error). On the other hand, studies evaluating interventions that aimed to improve the quality of existing services generated significantly higher impacts on average (0.20 sd in Column 5) than studies of expansions in services (0.10 sd in Column 4) (again see Appendix Table 7 for a formal test of this difference).

We have sufficient sample size to test whether these gains are present separately for all the developmental domains we pooled together as cognitive skills.[24] For language, literacy, and math, the average effect sizes for the full sample of extracted outcomes are positive and statistically significant, suggesting gains of 0.108 sd in language, 0.216 in literacy, 0.224 in math, and 0.084 in general skills (Table 5, Panel B, column (1)). This advantage over children in comparison groups who either did not attend preprimary education or did not experience increased investment in their preprimary classrooms is also evident for the general cognition category. The larger gains in literacy and math may be explained by the fact that children are less likely to engage in activities in the home that build these skills compared to language skills. In studies from high-income countries, preprimary interventions lead to a significant 0.115 sd increase in language scores, a 0.238 sd increase in literacy scores, and a 0.360 sd increase in math scores. In low- and middle-income countries, preprimary education interventions raise math scores by a significant 0.164 sd, on average. The estimated average effect size for language in studies from these countries is positive but fails to reach conventional levels of statistical significance. Our sample size of observations and studies measuring literacy in these countries was too small to estimate an average effect size.

[24] Appendix Figures 2–6 present study-specific effect sizes for cognitive skills (language, literacy, math, and general).

4.2.2 Non-cognitive skills

Table 5 Panel C presents average effect sizes for outcomes measuring non-cognitive skills: executive functioning, social-emotional skills, and behavior. When pooled across income levels (column 1) and across all outcomes, estimated effects suggest that preprimary education interventions lead to significant gains of 0.121 sd in non-cognitive skills. Again, we cannot statistically distinguish impacts estimated from studies in high-income countries from those estimated in studies from low- and middle-income countries (see Appendix Table 7 for the estimated difference between the two effects and its precision). In this case, however, estimated impacts from studies evaluating modifications to existing programs (0.159 sd in Column 5) largely drive the pooled average effect size, whereas the average impact is smaller and insignificant (0.048 sd) among the smaller sub-sample of studies evaluating preprimary expansions that also measured non-cognitive skills.

When we examine specific domains, we see that executive functions improve by an average of 0.095 sd ($p<0.10$). This average effect is driven largely by interventions evaluated in high-income countries, most of which directly aimed to improve these types of skills, and which increased these skills by an average of 0.169 sd ($p<0.01$). We lack sufficient sample size to estimate a separate average effect size for executive functions among studies from low- and middle-income countries.

Preprimary education interventions also significantly improved social-emotional skills by an average of 0.115 sd when we pool across both income groups, with high-income countries and low- and middle-income countries demonstrating similar significant effects of 0.094 and 0.130 sd, respectively (columns 2 and 3). The average effect size for outcomes related to child behavior in studies from high-income countries was a statistically insignificant 0.07 sd. Studies conducted in low- and middle-income countries did not collect this outcome.[25] [26]

[25] In high-income contexts, the largest effect sizes, both positive and negative, come from a US study reporting a statistically significant improvement in behavior of 1.06 sd to -0.77 for externalizing behavior following a teacher intervention to reduce children's behavioral problems (Raver et al., 2009). In low- and middle-income contexts, standardized effect sizes range from a significant 0.45 sd on a social composite index following exposure to a preprimary curriculum that promoted social and emotional abilities through play in India (Dillon et al., 2017), to an insignificant -0.08 sd in prosocial scores following a preprimary school expansion in Cambodia (Bouguen et al., 2018).

[26] Appendix Figures 7–10 graph study-specific effect sizes and visually demonstrate the overall positive but still mixed results for non-cognitive skills.

4.3 Persistence of effects beyond preprimary education

This section reports robust variance meta-regression results for the average effect size of preprimary interventions *after* the preprimary period. If preprimary education imparts skills that help children learn and makes subsequent education efforts more effective, then we should expect to see the benefits of preprimary interventions persist when studies track children into post-preprimary education or adulthood.

Unfortunately, fewer studies estimate impacts beyond the preprimary period and no studies in low- and middle-income countries report impacts measured in adulthood. Thus, we often lack statistical power for inference. Table 6 (Panels A–C) reports average effect sizes for outcomes measuring school participation and progression, cognitive skills, and non-cognitive skills when children are between the ages of 6 and 18 years. On average, the time elapsed between the preprimary intervention and outcome measurement for the post-preprimary period is around 5 years, ranging from 1 to 15 years. Overall, the results suggest that preprimary interventions can generate advantages that last beyond the preprimary period, both in cognitive skills and non-cognitive skills.

4.3.1 School participation and progression

The meta-regression results in Table 6 Panel A suggest a positive but statistically insignificant average effect size of 0.142 sd for the impact of preprimary interventions on subsequent school participation and progression in the post-preprimary period.[27] The study-specific effect sizes in Appendix Figures 12-13 suggest differences across income contexts. Preprimary interventions can improve outcomes like high-school graduation rates in high-income countries (Rossin-Slater & Wust, 2020; Gray-Lobe et al., 2021), but the overall effect of these interventions on school participation and progression appears to be low after the preprimary period. On the other hand, in low- and middle-income contexts, effect sizes are positive and significant in most studies, suggesting that preprimary interventions can improve school participation beyond the preprimary years.

[27] Though we extracted school participation and progression outcomes in adulthood from studies in our sample, we lack statistical power to estimate the average effect size and its precision.

4.3.2 Cognitive skills

When we pool effect sizes across high-income countries and low- and middle-income contexts and aggregate outcomes into a single category of cognitive skills, we find a significant persistent advantage of 0.071 sd in the post pre-primary period following interventions that either expanded or improved preprimary education (Table 6, Panel B). Sample sizes preclude a complete breakdown by skill, country-income context, and intervention goal, but average effect sizes are generally similar across these sub-samples.[28]

4.3.3 Non-cognitive skills

The estimated average effect size pooled across income contexts and across outcomes in Panel C of Table 6 suggests significant persistence of advantages in non-cognitive skills of 0.068 sd. In the post-preprimary period, children benefitting from preprimary interventions show a significant advantage of 0.094 sd in social-emotional skills. Again, estimated effect sizes are similar in magnitude and statistically indistinguishable across country-income contexts.[29]

4.4 Heterogeneity by socioeconomic status

This section reports average effect sizes estimated from Equation (5) when we limit our sample to coefficients estimated separately for subpopulations defined by their socioeconomic status. Given the infrequency of separate estimates for these subpopulations, we only report estimates for outcomes related to school participation and progression and for all skills pooled together, controlling for both the individual's level of education at the time of measurement and for country-income context whenever possible.

---

[28] Appendix Figures 13–16, which plot study-specific effect sizes for cognitive skills, suggest that the limited longitudinal evidence is mixed. Some studies exhibit economically and statistically meaningful impacts; others report negative but statistically insignificant results. In high-income contexts, for example, effect sizes—when standardized—could be as high as the significant 0.38 sd in first-grade math scores following a change in the preprimary math curriculum and teacher training/coaching (Clements et al., 2013), and as low as the significant -0.14 sd in a state achievement test following an offer of a slot in a subsidized pre-k program (Lipsey et al., 2018). Similarly, in low-income contexts, standardized effect sizes could be as high as a significant 0.23 sd in literacy scores measured in first grade when an additional year of preprimary education is offered to children (Spier et al., 2020), and as low as a statistically insignificant -0.05 sd in literacy scores following a teacher training and coaching program (Wolf et al., 2019b).

[29] As for outcomes related to cognitive skills, study-specific effect sizes in Appendix Figures 17–19 make apparent both low sample sizes for executive functions, social-emotional skills, and behaviors after the preprimary period and mixed results, as well as a high count of statistically insignificant effects.

The average effect sizes across these pooled outcome categories in Table 7 provide suggestive evidence that pre-primary interventions can have greater impacts on more disadvantaged populations. For instance, for outcomes related to school participation and progression, estimated effects are 0.096 sd higher on average for children coming from lower socioeconomic backgrounds, compared to their less disadvantaged peers, although this advantage in gains is not statistically significant. However, average effect sizes are 0.05 sd significantly higher for disadvantaged populations when it comes to skills.

4.5 Economic analysis

In Table 8, we present our findings from an economic evaluation of a subset of studies in low- and middle-income countries that reported sufficient information for us to compare costs per child to an estimate of the gain in lifetime earnings the child would receive from an improvement in cognitive skills. Panel A reports calculations for a discount rate of 3 percent, while Panel B presents results from a more conservative rate of 5 percent. Ganimian et al. (2021) report a range of benefit-to-cost ratios following the same method for a discount rate of 3 percent, which we directly report in Panel A.[30]

In Panel A, our smallest benefit-to-cost ratio is 3.5 when we use the smallest estimate of the returns to cognitive skills (4.3 percent per standard deviation) to translate the smallest estimate of gains in cognitive skills to gains in earnings for the Spier et al. (2020) study, which assessed the impact of adding an additional year of preprimary education for four-year old children on top of the mandatory year prior to the start of first grade in Bangladesh. Our largest benefit-to-cost ratio is 103.5 when we use the largest estimate of the returns to cognitive skills (13 percent per standard deviation) and the largest estimate of the returns to social-emotional skills (15 percent per standard deviation) to translate the smallest estimated average treatment effects into earnings for the Wolf et al. (2019a) study. This study assessed the impact of an in-service teacher training program in Ghana that sought to help teachers transition to a more holistic, child-centered curriculum.[31] Although both estimates of the benefit-to-cost ratio are quite high, it is

[30] Appendix Table 4 documents the assumptions underlying these calculations as well as data sources for parameters such as the labor force participation rate and growth in real wages.

[31] Though poor children did show positive and significant gains in cognitive skills in Brinkman et al. (2017), the average treatment effects for all cognitive skills in the full sample were not statistically distinguishable from zero. However, even in the full sample, there were average gains in social-emotional skills. If we

understandable that the ratio would be higher for a teaching training program implemented once with some refresher sessions than for the provision of an additional year of education.

Panel B uses the more conservative discount rate of 5 percent to translate lifetime earnings into a present value (Haacker et al., 2020). We still find benefit-to-cost ratios all above 1. The minimum value for this ratio is 14.2 for the Gallego et al. (2021) study, which assessed the impact of a change in the math curriculum. The maximum value, which is for the Wolf et al. (2019a) study, is still quite large at 49.7.

5. Discussion and conclusion

This study investigates whether current investments in preprimary education are too low through a systematic review of quasi-experimental evidence from around the world. Using 798 outcomes extracted and standardized from 54 studies, we use robust variance meta-regression to assess the returns to investment in preprimary education by establishing whether on average there is demand for preprimary programs and the extent to which preprimary education improves children's skills during the preprimary period. To more concretely gauge returns, we also combine estimates from the literature on the relationship between improvements in cognitive skills and future earnings with study-specific estimates of improvements in these skills in a subset of studies from low- and middle-income countries that report per-child costs.

We find that interventions that expand access to preprimary education led to significant increases in the take-up of preprimary education services and school participation during the preprimary period. On average, preprimary education programs significantly improve children's cognitive and social-emotional skills and executive functions in the short run, suggesting that these kinds of services promote learning and skill development better than children's alternative care options during the preprimary period. A translation of these effects into gains in children's earnings suggest sizable benefit-to-cost ratios, ranging from 1.7 to 103.5.

Although fewer studies track children beyond the preprimary years, we find evidence of smaller but statistically significant persistent effects of preprimary interventions in both cognitive skills and non-cognitive skills. Not only do these estimates highlight the potential for preprimary education programs to generate lasting learning gains among young children, but they also

---

estimate wage returns only for these gains in social emotional skills with a discount rate of 3 percent, we still find sizable benefit-to-cost ratios of 8.1 (lower bound) to 24.5 (upper bound).

underscore the role that meta-analysis can play in overcoming estimation challenges related to statistical power that may face individual studies, particularly small-sample longitudinal studies.

Average effect sizes in studies from high-income countries could not be statistically distinguished from average effect sizes in low- and middle-income countries for both cognitive and non-cognitive skills, both during and after the preprimary period. Estimated impacts were also similar in magnitude across the income spectrum. This points to the potential for preprimary education programs to improve the school readiness of young children even in contexts with limited state capacity.

This set of results, together with the currently low coverage rates in low- and middle-income countries, suggest that current levels of investment in preprimary education may be suboptimal. That is, an increase in spending on preprimary education coverage and quality may improve the overall efficiency of education spending, particularly if investments first target children from lower socioeconomic backgrounds. Given that children from disadvantaged backgrounds tend to have very low access to preprimary education and alternative learning opportunities (McCoy et al., 2018), minimum quality thresholds for services may depend on local conditions. Thus, even preprimary interventions that do not cover every aspect of quality could still improve learning outcomes for very disadvantaged children (Cascio & Schanzenbach, 2014). Our results suggest that existing preprimary programs have on average promoted higher skill development compared to what children would have experienced in the absence of these programs (either at home or in alternative programs).

While our sample of 798 estimated effect sizes is large, we recognize some limitations in our study design that may affect the magnitude and significance of the average effect sizes we estimate for school participation and progression and skills. For example, we restricted our search to studies that had been published in peer-reviewed journals or through formal working paper series. Given levels of publication-bias estimated multiple disciplines (Ioannidis, 2008; Ioannidis et al., 2017), this decision might bias us toward a positive and significant average effect.

On the other hand, we also aimed to extract all relevant outcomes from each included study, even if they might not be considered the best metrics to assess our outcomes of interest. This decision, and the attendant measurement error (attenuation bias) in many extracted outcomes, could bias us toward extracting multiple insignificant effects.

Nevertheless, from our large sample of published studies and formal working papers, we find positive and often persistent average effects of preprimary interventions on school participation and progression, as well as on cognitive and non-cognitive skills, during the preprimary period and beyond. Future research that aggregates evidence across studies would ideally try to understand variation in effectiveness based on program and child characteristics. The identification of what makes programs most effective could help make the returns and cost-effectiveness of preprimary education programs even higher.

References

Bailey, D.H., G.J. Duncan, F. Cunha, B.R. Foorman, and D.S., and Yeager. 2020a. "Persistence and Fade-Out of Educational-Intervention Effects: Mechanisms and Potential Solutions." *Psychological Science in the Public Interest* 21(2): 55-97.

Bailey, M.J., B.D. Timpe, and S. Sun. 2020b. *Prep School for poor kids: The long-run impacts of Head Start on Human capital and economic self-sufficiency*. (No. w28268): National Bureau of Economic Research.

Baird, S., F.H. Ferreira, B. Özler, and M. Woolcock. 2014. "Conditional, unconditional and everything in between: a systematic review of the effects of cash transfer programmes on schooling outcomes." *Journal of Development Effectiveness* 6 (1): 1-43.

Baker, M., J. Gruber, and K. Milligan. 2015. *Non-cognitive deficits and young adult outcomes: The long-run impacts of a universal child care program.* (No. w21571): National Bureau of Economic Research.

Baker, M., J. Gruber, and K. Milligan. 2008. "Universal child care, maternal labor supply, and family well-being." *Journal of Political Economy* 116 (4): 709-745.

Banerjee, A,, Sh. Cole, E. Duflo, and L. Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics,* 122 (3): 1235-1264.

Belfield, C., A.B. Bowden, A. Klapp, H. Levin, R. Shand, and S. Zander. 2015. "The economic value of social and emotional learning." *Journal of Benefit-Cost Analysis* 6 (3): 508-544.

Berkes, J., A. Bouguen, D. Filmer, and T. Fukao. 2019. *Improving Preschool Provision and Encouraging Demand: Heterogeneous Impacts of a Large-Scale Program.* Washington D.C.: The World Bank Working Paper Series.

Berlinski, S., S. Galiani, and M. Manacorda. 2008. "Giving children a better start: Preschool attendance and school-age profiles." *Journal of Public Economics* 92 ((5-6)): 1416-1440.

Bernal, R., and S.M. Ramírez. 2019. "Improving the quality of early childhood care at scale: The effects of “From Zero to Forever”." *World Development* 118: 91-105.

Bernal, R., O. Attanasio, X. Pena, and M. Vera-Hernández. 2019. "The effects of the transition from home-based childcare to childcare centers on children’s health and development in Colombia." *Early Childhood Research Quarterly* 47: 418-431.

Bouguen, A., D. Filmer, K. Macours, and S. Naudeau. 2018. "Preschool and parental response in a second best world evidence from a school construction experiment." *Journal of Human Resources* (The World Bank Working Paper Series) 53 (2): 474-512.

Brinkman, S.A., A. Hasan, H. Jung, A. Kinnell, and M. Pradhan. 2017. "The impact of expanding access to early childhood education services in rural Indonesia." *Journal of Labor Economics* 35 (S1): S305-S335.

Brodeur, A., N. Cook, and A. Heyes. 2020. "Methods matter: P-hacking and publication bias in causal analysis in economics." *American Economic Review* 110 (11): 3634-3660.

Brotman, L.M., S. Dawson-McClure, D. Kamboukos, K.Y. Huang, E.J. Calzada, K. Goldfeld, and E. Petkova. 2016. "Effects of ParentCorps in prekindergarten on child mental health and academic performance: Follow-up of a randomized clinical trial through 8 years of age." *JAMA Pediatrics* 170 (12): 1149-1155.

Campbell, F., G. Conti, J.J. Heckman, S.H. Moon, R. Pinto, E. Pungello, and Y. Pan. 2014. "Early childhood investments substantially boost adult health." *Science* 343 (6178): 1478-1485.

Carneiro, P., and R. Ginja. 2014. "Long-term impacts of compensatory preschool on health and behavior: Evidence from Head Start." *American Economic Journal: Economic Policy* 6 (4): 135-73.

Cascio, E.U. 2015. *The promises and pitfalls of universal early education.* IZA World of Labor.

Cascio, E.U., and D.W. Schanzenbach. 2014. *Proposal 1: Expanding preschool access for disadvantaged children.* Policies to address poverty in America, p.19.

Cascio, E.U., and D.W., and Schanzenbach. 2013. "The impacts of expanding access to high-quality preschool education ." *National Bureau of Economic Research* (No. w19735).

Chetty, R., J.N. Friedman, N. Hilger, E. Saez, D.W. Schanzenbach, and D. Yagan. 2011. "How does your kindergarten classroom affect your earnings? Evidence from Project STAR." *The Quarterly Journal of Economics* 126 (4): 1593-1660.

Clements, D.H., J., Sarama, C.B. Wolfe, and M.E. Spitler. 2013. "Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies: Persistence of effects in the third year." *American Educational Research Journal* 50 (4): 812-850.

Conti, G., J.J. Heckman, and R. Pinto. 2016. "The effects of two influential early childhood interventions on health and healthy behaviour." *The Economic Journal* 126 (596): F28-F65.

Cunha, F., and J., Heckman. 2007. "The technology of skill formation." *American Economic Review* 97 (2): 31-47.

Currie, J. M. 2001. *Early childhood intervention programs: what do we know?* Brookings Institution, Brookings Roundtable on Children.

Dean, J.T., and S. Jayachandran. 2020. "Attending kindergarten improves cognitive development in India, but all kindergartens are not equal."

Dehaene, S. 2010. *Reading in the brain: The new science of how we read.* Penguin Group USA.

Deming, D. 2009. "Early childhood intervention and life-cycle skill development: Evidence from Head Start." *American Economic Journal: Applied Economics* 1 (3): 111-34.

DeWind, N. K., J. Park, M. G. Woldorff, and E. M. Brannon. 2019. "Numerical encoding in early visual cortex." *Cortex* 114: 76-89.

Diamond, A. 2013. "Executive functions." *Annual Review of Psychology* 64: 135-168.

Dillon, M.R., H. Kannan, J.T. Dean, E.S. Spelke, and E. Duflo. 2017. "Cognitive science in the field: A preschool intervention durably enhances intuitive but not formal mathematics." *Science* 357 (6346): 47-55.

Duflo, E., P. Dupas, and M. Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya ." *American Economic Review* 10 (5): 1739-74.

Duncan, G.J., and K. Magnuson. 2013. "Investing in preschool programs." *Journal of Economic Perspectives* 27 (2): 109-32.

Evans, D.K., and A., Popova. 2016. "What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews." *The World Bank Research Observer* 31 (2): 242-270.

Evans, D.K., and F. Yuan. 2021. "What We Learn about Girls' Education from Interventions that Do Not Focus on Girls." *The World Bank Economic Review.*

Evans, D.K., P. Jakiela, and H.A. Knauer. 2021. "The impact of early childhood interventions on mothers." *Science* 372 (6544): 794-796.

Fernald, L.C., A. Weber, E. Galasso, and L. Ratsifandrihamanana. 2011. "Socioeconomic gradients and child development in a very low income population: evidence from Madagascar." *Developmental Science* 14 (4): 832-847.

Fernald, L.C., E. Prado, and Kariger, P. Raikes, A. 2017. *A toolkit for measuring early childhood development in low and middle-income countries.* Washington D.C.: The World Bank.

Galasso, E., and A. Wagstaff. 2019. "The aggregate income losses from childhood stunting and the returns to a nutrition Intervention aimed at reducing stunting." *Economics & Human Biology* (Policy Research Note World Bank) 34: 225-238.

Gallego, F.A., E. Näslund-Hadley, and M. Alfonso. 2021. "Changing Pedagogy to Improve Skills in Preschools: Experimental Evidence from Peru." *The World Bank Economic Review* 35 (1): 261-286.

Ganimian, A.J., K. Muralidharan, and C.R. Walters. 2021. *Augmenting State Capacity for Child Development: Experimental Evidence from India.* (No. w28780): National Bureau of Economic Research.

Garces, E., D. Thomas, and J. Currie. 2002. "Longer-term effects of Head Start." *American Economic Review* 92 (4): 999-1012.

Gelber, A., and A. Isen. 2013. " Children's schooling and parents' behavior: Evidence from the Head Start Impact Study." *Journal of Public Economics* 101: 25-38.

Haacker, M., T.B. Hallett, and R. Atun. 2020. "On discount rates for economic evaluations in global health." *Health Policy and Planning* 35 (1): 107-114.

Hamlin, J., T. Ullman, J. Tenenbaum, and N. Baker, C. Goodman. 2013. "The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model." *Developmental Science* 16 (2): 209-226.

Havnes, T., and M. Mogstad. 2011. "No child left behind: Subsidized child care and children's long-run outcomes." *American Economic Journal: Economic Policy* 3 (2): 97-129.

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

Heckman, J.J. 2007. "The economics, technology, and neuroscience of human capability formation. ." *Proceedings of the National Academy of Sciences* 104(33): 13250-13255.

Heckman, J.J., S.H. Moon, R. Pinto, P.A. Savelyev, and A. and Yavitz. 2010. "The rate of return to the HighScope Perry Preschool Program." *Journal of Public Economics* 94 (1-2): 114-128.

Hedberg, E.C. 2011. *ROBUMETA: Stata module to perform robust variance estimation in meta-regression with dependent effect size estimates.* Boston: Boston College Working Papers in Economics.

Hedges, L.V., E. Tipton, and M.C. Johnson. 2010. "Robust variance estimation in meta-regression with dependent effect size estimates." *Research Synthesis Methods* 1 (1): 39-65.

Hendren, N., and B. Sprung-Keyser. 2020. "A unified welfare analysis of government policies." *The Quarterly Journal of Economics* 135 (3): 1209-1318.

Hidrobo, M., J. Hoddinott, N Kumar, and M. Olivier. 2018. "Social protection, food security, and asset formation." *World Development* (101): 88-103.

Horowitz, J.L., and C.F. Manski. 2000. "Nonparametric analysis of randomized experiments with missing covariate and outcome data." *Journal of the American Statistical Association* 95 (449): 77-84.

Ioannidis, J.P. 2008. "Why most discovered true associations are inflated." *Epidemiology* 640-648.

Ioannidis, J.P., T.D. Stanley, and H. Doucouliagos. 2017. "The power of bias in economics research."

Jara-Ettinger, J., E Gibson, and Kidd, C. Piantadosi, S. 2016. "Native Amazonian children forego egalitarianism in merit-based tasks when they learn to count." *Developmental Science* 19 (6): 1104-1110.

Johnson, R.C., and C.K. Jackson. 2019. "Reducing inequality through dynamic complementarity: Evidence from Head Start and public school spending." *American Economic Journal: Economic Policy* 11 (4): 310-49.

Kline, P., and C.R. Walters. 2016. "Evaluating public programs with close substitutes: The case of Head Start." *The Quarterly Journal of Economics,* 131 (4): 1795-1848.

Knudsen, E.I. 2004. "Sensitive periods in the development of the brain and behavior." *Journal of Cognitive Neuroscience* 16 (8): 1412-1425.

Knudsen, E.I., J.J. Heckman, J.L. Cameron, and J.P. Shonkoff. 2006. "Economic, neurobiological, and behavioral perspectives on building America's future workforce." *Proceedings of the national Academy of Sciences* 103 (27): 10155-10162.

Kremer, M., and A. Holla. 2009. "Improving education in the developing world: what have we learned from randomized evaluations?" *Annu. Rev. Econ.* 1 (1): 513-542.

Krueger, A.B. 1999. "Experimental estimates of education production functions." *The Quarterly Journal of Economics* 114 (2): 497-532.

Lee, D.S. 2009. "Training, wages, and sample selection: Estimating sharp bounds on treatment effects." *The Review of Economic Studies* 76 (3): 1071-1102.

Lipsey, M.W., and D.B., Wilson. 2001. *Practical meta-analysis.* SAGE publications, Inc.

Lipsey, M.W., D.C. Farran, and K. Durkin. 2018. "Effects of the Tennessee Prekindergarten Program on children’s achievement and behavior through third grade." *Early Childhood Research Quarterly* 45: 155-176.

Martinez, S., S. Naudeau, and V. Pereira. 2017. *Preschool and child development under extreme poverty: evidence from a randomized experiment in rural Mozambique.* Washington D.C.: The World Bank.

McCoy, D.C., C. Salhi, H. Yoshikawa, M. Black, P. Britto, and G. Fink. 2018. "Home-and center-based learning opportunities for preschoolers in low-and middle-income countries." *Children and Youth Services Review* 88: 44-56.

McCoy, D.C., H. Yoshikawa, K.M. Ziol-Guest, G.J. Duncan, H.S. Schindler, K. Magnuson, R. Yang, A. Koepp, and J.P., Shonkoff. 2017. "Impacts of early childhood education on medium-and long-term educational outcomes." *Educational Researcher* 46 (8): 474-487.

Meager, R. 2019. "Understanding the average impact of microcredit expansions: A bayesian hierarchical analysis of seven randomized experiments." *American Economic Journal: Applied Economics* 11 (1): 57-91.

Naudeau, S., S. Martinez, P. Premand, and D. Filmer. 2011. "Cognitive development among young children in low-income countries, is No Small Matter." In *The Impact of Poverty Shocks, and Human Capital Investments in Early Childhood Development*, by editor Harold Alderman, 9-50. Washington D.C.: The World Bank.

Nores, M., and W.S. Barnett. 2010. "Benefits of early childhood interventions across the world:(Under) Investing in the very young." *Economics of Education Review* 29 (2): 271-282.

Ozler, B., L. C. H. Fernald, P. Kariger, C. McConnell, M. Neuman, and E. Fraga. 2018. "Combining preschool teacher training with parenting education: A clusterrandomized controlled trial." *Journal of Development Economics* (The World Bank.) 133: 448–467.

Phillips, D., M. Lipsey, K. Dodge, R Haskins, D. Bassok, M. Burchinal, G. Duncan, M. Dynarski, K. Magnuson, and Ch. Weiland. 2017. *The Current State of Scientific Knowledge on Pre-Kindergarten Effects.* Washington D.C.: The Brookings Institute.

Raver, C.C., S.M. Jones, C. Li-Grining, F. Zhai, M.W. Metzger, and B. Solomon. 2009. "Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial." *Journal of Consulting and Clinical Psychology* 77 (2): 302.

Rossin-Slater, M., and M. Wüst. 2020. "What is the Added Value of Preschool for Poor Children? Long-Term and Intergenerational Impacts and Interactions with an Infant Health Intervention." *American Economic Journal: Applied Economics* 12 (3): 255-86.

Schady, N., J. Behrman, M.C. Araujo, R. Azuero, R. Bernal, D. Bravo, F. Lopez-Boo, et al. 2014. *Wealth gradients in early childhood cognitive development in five Latin American countries.* Washington D.C.: The World Bank.

Schultz, T.P. 2010. "Population and Health Policies." In *Handbook of Development Economics 5*, edited by Mark Rosenzweig and Dani Rodrik, 4785-4881.

Shonkoff, J.P., D.A. Phillips, and National Research Council. 2000. *The developing brain.* National Academies Press (US): In From neurons to neighborhoods: The science of early childhood development. .

Spelke, E., and K. Shutts. forthcoming. "Chapter 1: Learning in the Early Years." In *Quality Early Learning: Nurturing Children's Potential*, edited by A. Devercelli and M Bendini.

Spier, E., K. Kamto, A. Molotsky, A. Rahman, N. Hossain, Z. Nahar, and H. and Khondker. 2020. "Bangladesh Early Years Preschool Program Impact Evaluation."

Tamis-LeMonda, C.S, K. E Adolph, S. A. Lobo, L. B. Karasik, S.. Ishak, and K. A. Dimitropoulou. 2008. "When infants take mothers' advice: 18-month-olds integrate perceptual and social information to guide motor action." *Developmental Psychology* 44 (3): 734.

Tanner-Smith, E.E, and E. Tipton. 2014. "Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS." *Research Synthesis Methods* 5 (1): 13-30.

Tanner-Smith, E.E., E. Tipton, and J.R. Polanin. 2016. "Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R." *Journal of Developmental and Life-Course Criminology* 2 (1): 85-112.

Tipton, E. 2015. "Small sample adjustments for robust variance estimation with meta-regression." *Psychological Methods* 20 (3): 375.

Tipton, E., J.E. Pustejovsky, and H. Ahmadi. 2019. "Current practices in meta-regression in psychology, education, and medicine." *Research Synthesis Methods,* 10 (2): 180-194.

Tomasello, M., M. Carpenter, J. Call, and T., Moll, H. Behne. 2005. "Understanding and sharing intentions: The origins of cultural cognition." *Behavioral and brain sciences* 28 (5): 675-691.

UIS, (UNESCO Institute for Statistics). 2018. *UIS Stat.* Accessed March 4, 2021. http://data.uis.unesco.org/.

UNICEF. 2019. *A World Ready to Learn: Prioritizing quality early childhood education. Global Report.* . New York: UNICEF.

van Huizen, T., and J. Plantenga. 2018. "Do children benefit from universal early childhood education and care? A meta-analysis of evidence from natural experiments." *Economics of Education Review* 66: 206-222.

Vivalt, E. 2020. "How much can we generalize from impact evaluations?" *Journal of the European Economic Association* 18 (6): 3045-3089.

Wang, Y., R. Williams, L. Dilley, and D. M. Houston. 2020. "A meta-analysis of the predictability of LENA™ automated measures for child language development." *Developmental Review* 57: 100921.

Wolf, S., J.L. Aber, J.R. Behrman, and E. and Tsinigo. 2019a. "Experimental impacts of the "Quality Preschool for Ghana" interventions on teacher professional well-being, classroom quality, and children's school readiness." *Journal of Research on Educational Effectiveness* 12 (1): 10-37.

Wolf, S., J.L. Aber, J.R. Behrman, and M. Peele. 2019b. "Longitudinal causal impacts of preschool teacher training on Ghanaian children’s school readiness: Evidence for persistence and fade-out." *Developmental Science* 22 (5): e12878.

World Bank. 2013. *What Matters Most for Early Childhood Development: A Framework Paper*. Washington, D.C: World Bank. SABER Working Paper Series.

World Bank. 2018. *World Development Report 2018 : Learning to Realize Education's Promise.* Washington, DC: World Bank.

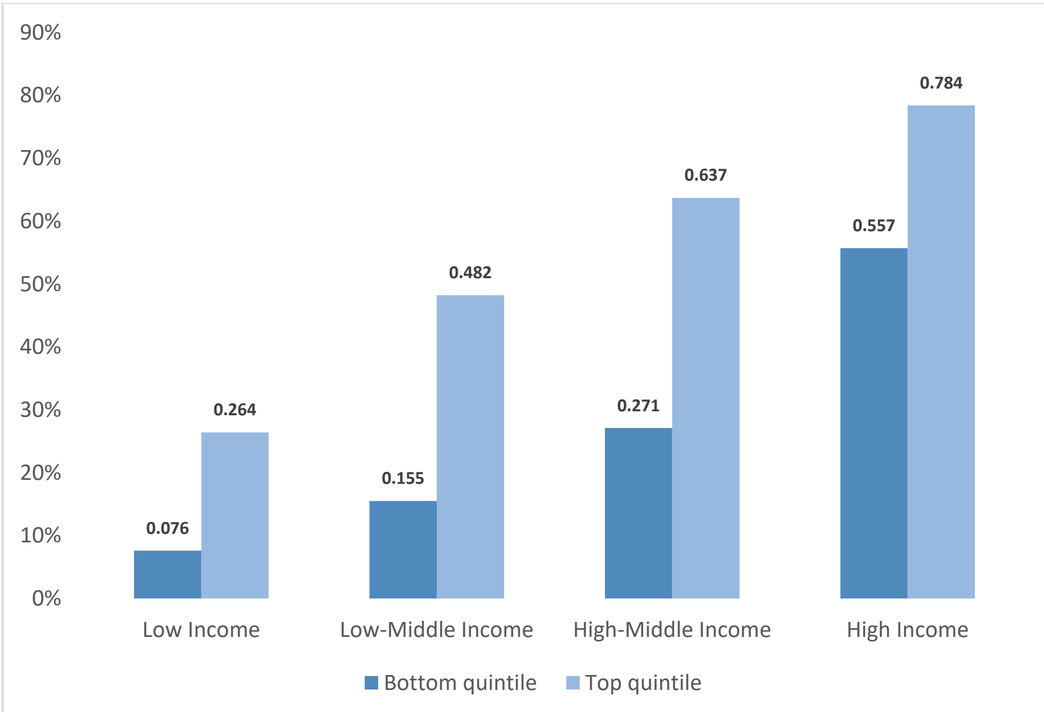
Yoshikawa, H., C. Weiland, J. Brooks-Gunn, M.R. Burchinal, L.M. Espinosa, W.T. Gormley, J. Ludwig, K.A. Magnuson, D. Phillips, and M.J. Zaslow. 2013. *Investing in Our Future: The Evidence Base on Preschool Education. Society for Research in Child.* Society for Research in Child Development.

Yuan, S., and C. Fisher. 2009. "Really? She blicked the baby? Two-year-olds learn combinatorial facts about verbs by listening." *Psychological Science* 619-626.

Zelazo, P.D., U. Müller, D. Frye, S. Marcovitch, G. Argitis, J. Boseovski, J.K. Chiang, et al. 2003. "The development of executive function in early childhood." *Monographs of the society for research in child development* i-151.

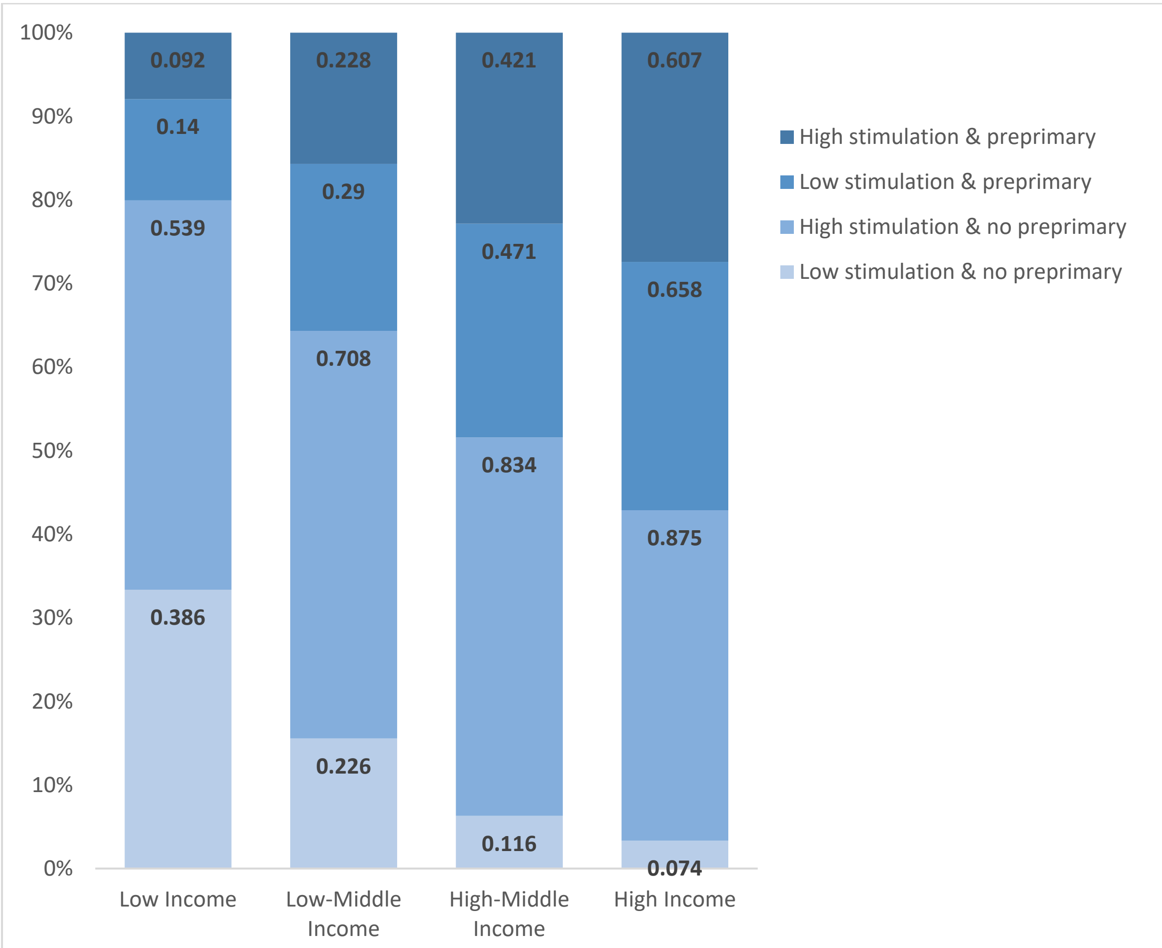
Figures and Tables

Figure 1: Children in early childhood education by country income and household wealth



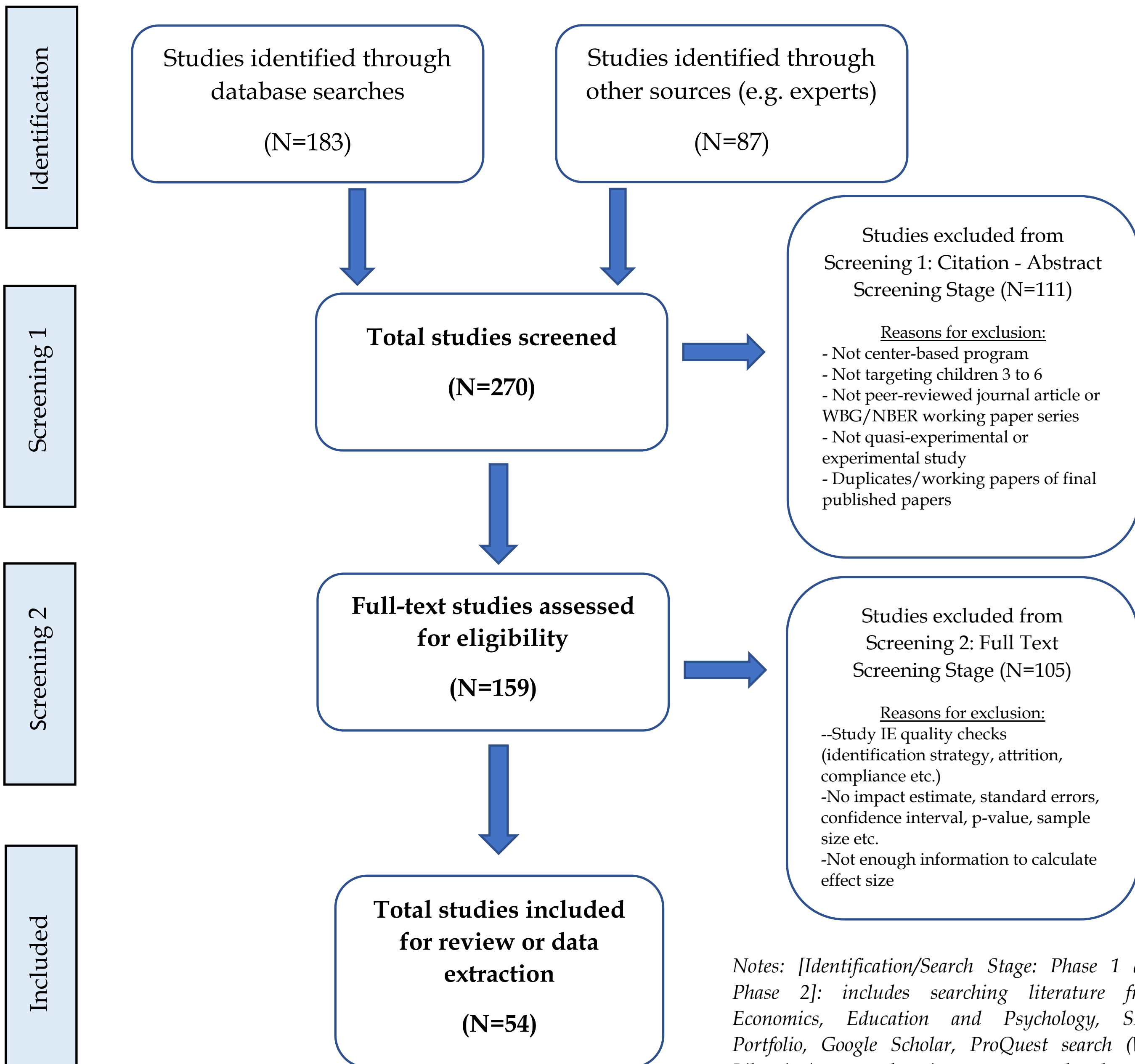
Notes: Data from the Multiple Indicator Cluster Surveys, reported in McCoy et al. (2018)

Figure 2: Children receiving stimulation at home and attending early childhood education by country income



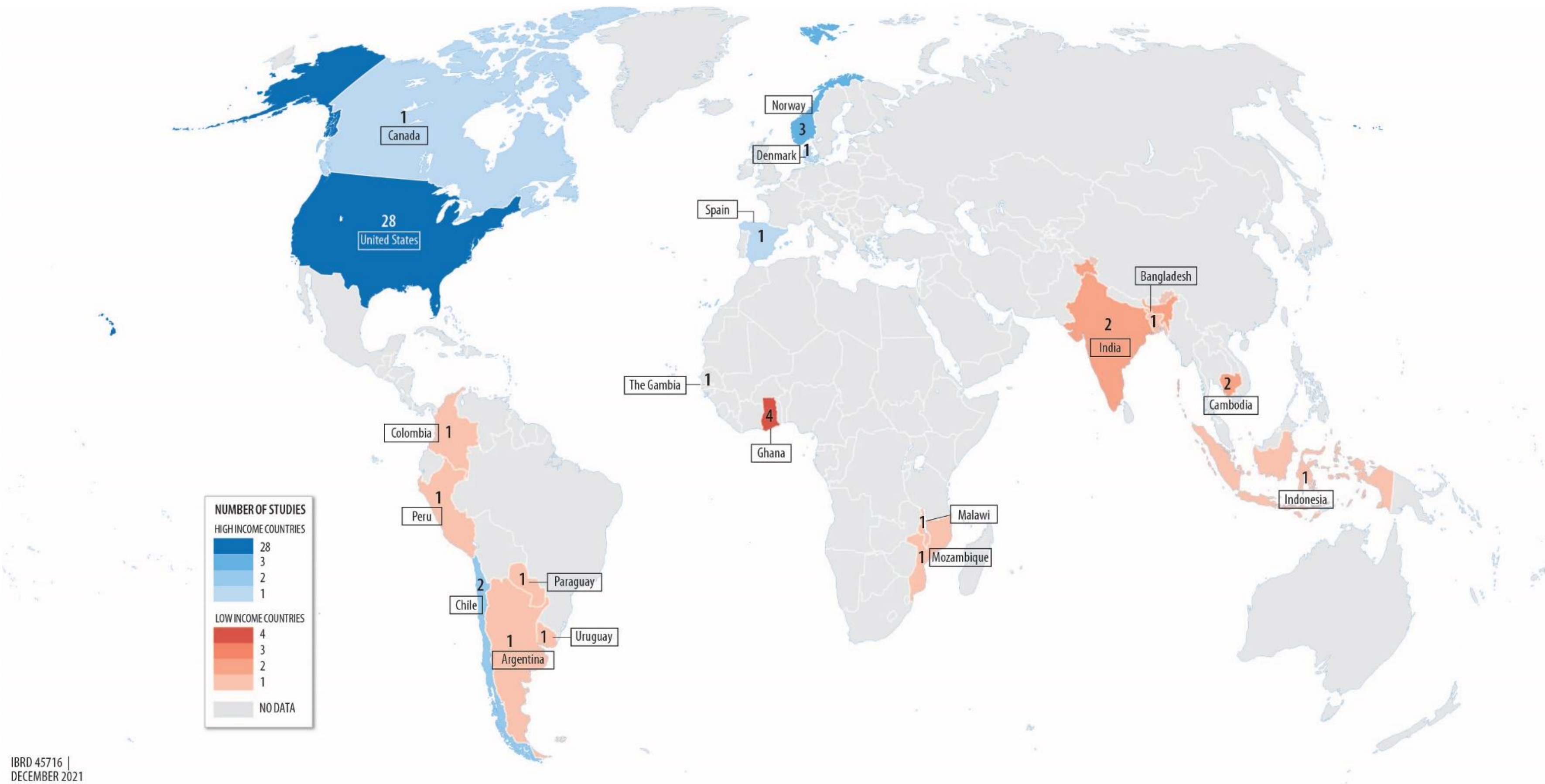
Notes: Data from the Multiple Indicator Cluster Surveys, reported in McCoy et al. (2018)

Figure 3: Study Selection Process



Notes: [Identification/Search Stage: Phase 1 and Phase 2]: includes searching literature from Economics, Education and Psychology, SIEF Portfolio, Google Scholar, ProQuest search (WB Librarian), external reviewers, experts, hand search and bibliographies. [Screening 1: Phase 3]: review of citations and abstracts to eliminate duplicates, not published papers and papers that do not estimate rigorously the impacts of preprimary programs or policies. [Screening 2: Phase 4]: review of full text studies, eliminate papers based on lack of a rigorous quantitative analysis (e.g. identification strategy, attrition, compliance etc.). [Data Extraction: Phase 5]: coding of information at the study, intervention and outcomes (effect size) level.

Figure 4: Geographical Distribution of Included Studies (N=54)



Notes: The sample of 54 studies included in our review covers the following countries: Argentina, Bangladesh, Cambodia, Canada, Chile, Colombia, Denmark, Gambia, Ghana, India, Indonesia, Malawi, Mozambique, Norway, Paraguay, Peru, Spain, U.S., and Uruguay.

Table 1: Search and inclusion criteria for studies

Panel A: Search	
Sources	APA PcycNet, Academic Search Elite/EBSCO host, Econlit, ERIC, IDEAS, ScienceDirect, Social Science, Research Network (SSRN), Google Scholar, ProQuest Database, MEDLINE, NBER, WBG Working Papers, Strategic Impact Evaluation Fund (SIEF) Portfolio, JSTOR, JOLIS library catalogue – International Monetary Fund, World Bank and International Finance Corporation, PsycINFO, PubMed, and Web of Science
Search terms	Preprimary OR preprimary OR Pre-school OR Kindergarten OR early childhood development OR early childhood education OR early child development OR early child education OR early learning OR early cognitive development OR early skills development) AND (impact evaluation OR field experiment OR randomized OR program evaluation OR meta-analysis OR systematic review OR synthesis review OR qualitative study)>2005. Age 3 to 6; Countries: Global
Panel B: Inclusion criteria	
Program beneficiaries	Children aged 3-6 years
Interventions	Programs that provide group-based childcare to children 3 to 6 years old in a center setting with a formal or informal developmental and educational focus. This includes formal preprimary schools, community schools, preschools, kindergarten, pre-kindergarten, or daycare with an educational component. We also include supplementary / co-interventions targeting parents, teachers, or other inputs (i.e., teacher training, targeting curricula, pedagogical approaches, infrastructure) whose primary outcomes of interest are either child outcomes, or intermediary outcomes that affect the quality of a pre-school program.

Study design	Studies implementing experimental and quasi-experimental methods with a credible source of exogenous variation. We included studies employing one of the following designs: randomized control trial (RCT), regression discontinuity design (RDD), differences-in-differences (DID), instrumental variables (IV), and matching methods. At least one treatment arm must be able to isolate the effect of a preprimary program.
Outcomes of interest	Child outcomes: school participation and progression, cognitive skills, social-emotional skills, behavior, long-term educational attainment, adult health, and labor outcomes Teacher/parent (adult) outcomes: classroom practices and engagement in stimulation activities
Publication type	Peer-reviewed journal articles, WB/NBER Working Papers, and other working/discussion papers only if they included a formal institution/citation. We excluded studies that are not part of a working paper series (including PhD Dissertations, job market papers, and conference working papers) and other institutional policy publications.
Publication date	Any
Geography	Global

Table 2: Study characteristics

	Mean	Standard deviation
	(1)	(2)
Focused on high-income country	0.67	0.48
Publication timing		
2007-2012	0.33	0.48
2013-2018	0.41	0.50
2019-2021	0.26	0.44
Published in peer-reviewed journal	0.85	0.36
Discipline		
Economics	0.50	0.50
Education	0.44	0.50
Psychology	0.06	0.23
Randomized control trial	0.70	0.46
Evaluation of expansion in coverage	0.46	0.50

Notes: Observations are studies, N=54.

Table 3: Characteristics of interventions

	High-income countries	Low- & middle-income countries
	(1)	(2)
Program category		
Teacher professional development	0.55	0.69
	(0.07)	(0.08)
Subsidized or free access	0.49	0.03
	(0.07)	(0.03)
Change in curriculum	0.29	0.16
	(0.02)	(0.07)
Change in pedagogy	0.04	0.13
	(0.03)	(0.06)
Provision of materials	0.14	0.28
	(0.05)	(0.08)
Provision of new staff	0.00	0.09
	(0.00)	(0.05)
Provision of health and nutrition services	0.08	0.00
	(0.04)	0.00
School construction	0.02	0.34
	(0.02)	(0.09)
Parental engagement	0.04	0.50
	(0.03)	(0.09)
Community outreach	0.00	0.16
	(0.00)	(0.07)
Extension of school day	0.02	0.00
	(0.02)	(0.00)
Teacher payments	0.00	0.09
	(0.00)	(0.05)
Type of provision		
Formal preprimary	0.90	0.56
	(0.04)	(0.09)
Community-based preprimary	0.00	0.44
	(0.00)	(0.09)
Daycare with educational component	0.10	0.00
	(0.04)	(0.00)
School management		
Public	0.57	0.63
	(0.07)	(0.09)
Public-private	0.43	0.25
	(0.07)	(0.08)
Community	0.00	0.13
	(0.00)	(0.06)
Teacher formally qualified	0.80	0.50
	(0.06)	(0.09)
Program targets disadvantaged population	0.71	0.66
	(0.07)	(0.09)

Notes: Observations are evaluated programs, N=141 interventions or contrasts. Standard errors of presented means are in parentheses.

Table 4: Characteristics of extracted outcomes

	Mean	Standard deviation	Median
	(1)	(2)	(3)
Sample size	502,289	3,299,459	2,449
Fraction that are estimates for separate SES subpopulations	0.24	0.43	
Fraction measured during period			
Preprimary (3-5 years)	0.63	0.48	
Post-preprimary (6-18 years)	0.29	0.45	
Adulthood (after 18 years)	0.08	0.08	
Fraction measuring domain			
School participation and progression	0.19	0.39	
Language	0.11	0.32	
Literacy	0.06	0.24	
Math	0.08	0.27	
General cognitive	0.06	0.23	
Executive function	0.09	0.29	
Social-emotional learning	0.14	0.34	
Behavior	0.09	0.28	
Health	0.11	0.32	
Teacher practices	0.06	0.24	

Notes: Observations are extracted outcomes from studies in the sample, N=798. SES refers to socioeconomic status

Table 5: Average effects of preprimary interventions on participation and skills *during* the preprimary period

	Pooled sample	High-income	Low- & middle-income	Expansion	Quality
	(1)	(2)	(3)	(4)	(5)
Panel A: School participation and progression					
School participation and progression	1.4048**			1.538**	
	(0.4776)			(0.513)	
N	29			26	
Panel B: Cognitive skills					
Pooled skills	0.1474***	0.1787***	0.1223***	0.0950***	0.197***
	(0.0259)	(0.0431)	(0.0330)	(0.0263)	(0.0420)
N	131	76	55	58	73
Language	0.1084***	0.1150**	0.0992	0.0770***	0.157**
	(0.0295)	(0.0367)	(0.0526)	(0.0226)	(0.0591)
N	53	35	18	23	30
Literacy	0.2162***	0.2385**			0.213*
	(0.0634)	(0.0783)			(0.0844)
N	31	28			24
Math	0.2238***	0.3605**	0.1643***		0.265***
	(0.0469)	(0.1108)	(0.0406)		(0.0602)
N	28	10	18		16
General	0.0844*		0.0803	0.0866	
	(0.0367)		(0.0464)	(0.0472)	
N	19		16	16	
Panel C: Executive functions and social-emotional skills					
Pooled skills	0.1209***	0.1354**	0.1138***	0.0488	0.159***
	(0.0279)	(0.0494)	(0.0330)	(0.0412)	(0.0317)
N	114	75	39	32	82
Executive functions	0.0951*	0.1693*			0.176***
	(0.0387)	(0.0778)			(0.0365)
N	42	31			33
Social-emotional skillls	0.1150***	0.0935**	0.1301**		0.117***
	(0.0301)	(0.0318)	(0.0463)		(0.0329)
N	50	22	28		32
Behavior	0.0750	0.0750			
	(0.0500)	(0.0500)			
N	22	22			

Notes: Standard errors estimated using small sample adjustments as in Tipton (2015) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Number of observations (N) refers to number of study-specific average effects used in the specification. Specifications that pool skills use one large category for extracted outcomes - either cognitive skills (which combines language, literacy, math, and general cognitive skills) or what is typically referred to as non-cognitive skills (executive functions, social-emotional skills, and behaviors). See Appendix 1 for details on definitions of school participation and progression; language, literacy, math, and general skills; executive functions; social-emotional skills; and behavior. Cells are blank when the sample size is too small to be able to obtain trustworthy p-values, when the degrees of freedom fall below 4 or when all observations came from a single study.

Table 6: Average effects of preprimary interventions on participation and skills *after* the preprimary period

	Pooled sample	High-income	Low- & middle-income	Expansion	Quality
	(1)	(2)	(3)	(4)	(5)
Panel A: School participation and progression					
School participation and progression	0.1423			0.131	
	(0.0888)			(0.106)	
N	22			20	
Panel B: Cognitive skills					
Pooled skills	0.0718***	0.0604	0.0862**	0.0604	0.0865*
	(0.0236)	(0.0375)	(0.0311)	(0.0336)	(0.0346)
N	62	23	39	31	31
Literacy	0.0626			0.0655	
	(0.0344)			(0.0515)	
	14			9	
Math	0.0571*	0.0648	0.0512	0.0535	0.0622
	(0.0311)	(0.0566)	(0.0361)	(0.0547)	(0.0399)
	29	7	22	9	20
General	0.0603				
	(0.0476)				
	10				
Panel C: Executive functions and social-emotional skills					
Pooled skills	0.0686***		0.0750***		
	(0.0190)		(0.0178)		
N	70		34		
Social-emotional skillls	0.0945**		0.1056**		
	(0.0342)		(0.0361)		
N	34		21		

Notes: Standard errors estimated using small sample adjustments as in Tipton (2015) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Number of observations (N) refers to number of study-specific average effects used in the specification. Specifications that pool skills use one large category for extracted outcomes - either cognitive skills (which combines language, literacy, math, and general cognitive skills) or what is typically referred to as non-cognitive skills (executive functions, social-emotional skills, and behaviors). See Appendix 1 for details on definitions of school participation and progression; language, literacy, math, and general skills; executive functions; social-emotional skills; and behavior. Cells are blank when the sample size is too small to be able to obtain trustworthy p-values, when the degrees of freedom fall below 4 or when all observations came from a single study (Tipton, 2015).

Table 7: Difference in average effects of preprimary interventions by socio-economic status

	Pooled sample (1)	High-income (2)	Low- & middle-income (3)
Panel A: School participation and progression			
Low socio-economic status	0.0959	0.0938	
	(0.0685)	(0.0686)	
N	32	24	
Panel B: Skills (Pooled)			
Low socio-economic status	0.0535*		
	(0.0246)		
N	82		

Notes: Standard errors estimated using small sample adjustments as in Tipton (2015) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Number of observations (N) refers to number of study-specific average effects used in the specification. Appendix Table 2 presents the groups classified as "low socio-economic" sub-populations. All specifications include a constant term (the average effect size), which is omitted here. The coefficient on the low socio-economic status variable represents the average difference between study-specific average effect sizes between children coming from low and high socio-economic backgrounds.

Table 8: Returns to cognitive and social-emotional skills from low- and middle-income studies

	Total costs per child (dollars)	Preprimary-augmentation to lifetime earnings (dollars)						Lower-bound benefit-to-cost ratio	Upper-bound benefit-to-cost ratio
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Discount rate = 3%									
India (Ganimian et al, 2021)								11.1	11.8
India (Dillon et al, 2017)	$53	$913	$2,845	$1,429	$302	$2,234	$817	5.7	54.0
Malawi (Ozler et al, 2018)	$72	$782	$2,010	$1,109	$259	$1,487	$586	3.6	28.1
Ghana (Wolf et al, 2019a)	$16	$563	$1,657	$855	$186	$1,280	$478	11.6	103.5
Peru (Gallego et al, 2021)	$37	$3,264			$1,080			29.2	88.2
Bangladesh (Spier et al, 2020)	$145	$1,542	$4,175	$2,244	$510	$3,143	$1,212	3.5	28.8
Panel B: Discount rate = 5%									
India (Dillon et al, 2017)	$53	$411	$1,281	$643	$136	$1,006	$368	2.6	24.3
Malawi (Ozler et al, 2018)	$72	$370	$950	$524	$122	$703	$277	1.7	13.3
Ghana (Wolf et al, 2019a)	$16	$271	$796	$411	$89	$615	$230	5.6	49.7
Peru (Gallego et al, 2021)	$37	$1,587			$525			14.2	42.9
Bangladesh (Spier et al, 2020)	$145	$738	$1,999	$1,074	$244	$1,505	$580	1.7	13.8
High cognitive wage returns		Yes	Yes	Yes	No	No	No		
Low cognitive wage returns		No	No	No	Yes	Yes	Yes		
High social-emotional wage returns		No	Yes	No	No	Yes	No		
Low social-emotional wage returns		No	No	Yes	No	No	Yes		

Notes: For each country, we first calculate the present discounted value of a child's lifetime earnings, assuming that the child will work from age 22 to 65. To do this, we use the average nominal monthly wages for the latest year available, as well as the labor force participation rate and the average growth in real wages for the 2015-2019 period. See Appendix Table 4 for these parameters and our sources of data for all studies except for Ganimian et al, (2021), for which we directly use their estimates of benefit to cost ratios. In Panel A, when discounting future earnings to the present value, we assume a discount rate of 3 percent per year. In Panel B, we assume 5 percent per year. To calculate gains to lifetime earnings from preprimary interventions, we use the lowest significant treatment effect on cognitive skills reported in standard deviations in the paper and the lowest significant treatment effect on social-emotional skills. We use high and low estimates from the literature on the returns to a standard deviation increase in cognitive or social-emotional skills. Kline and Walters (2016) use an annual return of 13 percent per standard deviation of cognitive skills, while Galasso and Wagstaff (2018) use 4.3 percent. For the returns to social-emotional skills, Belfield et al (2015) use a range of 4 percent to 15 percent. We vary our assumptions on these returns in Columns 2 through 7. Because papers (with the exception of Ganimian et al., (2021)) reported costs per child in US dollars, we discounted lifetime earnings to the year costs were incurred and used nominal exchange rates collected by the International Monetary Fund to convert earnings to dollars in the year costs were incurred. In Columns 8 and 9, we report the ratio of benefits to costs for an individual child.

Appendix—For Online Publication Only

Appendix 1. Mapping of extracted outcomes

This section describes the skill categories we used to aggregate outcomes extracted from each study.

Literacy outcomes include early reading skills (e.g., letter/word recognition, decoding, print and phonological awareness), and writing, as well as reading comprehension and performance in achievement tests in studies that focus on longer-term impact. We classify constructs like expressive and receptive vocabulary, and oral comprehension as *language*.

Math outcomes include early mathematics knowledge and skills such as numerical identification and shape recognition, as well as math test scores beyond preprimary. Under *general cognition* we mapped more general knowledge and learning skills ranging from general intellectual ability, such as problem solving and communication skills, to summary indexes of achievement scores on literacy, language, and math skills, as captured by instruments such as the Early Development Instrument (Janus and Offord, 2007), International Development and Early Learning Assessment (Pisani, Borisova, and Dowd 2015), and the Woodcock–Johnson (Woodcock, McGrew, & Mather, 2001).

We classified outcomes representing constructs such as attention, inhibition, working memory, and cognitive flexibility as *executive functions. Social-emotional* outcomes include social cognition, social competence, emotional regulation, prosociability, and relationships with peers, among others. *Behavioral outcomes* include reported or observed measures of aggression, internalizing and externalizing behavior problems, conduct, and disciplinary actions during preprimary and basic education, as well as indicators such as crime, arrests, convictions, and incarceration during adulthood. There is some overlap across these last three outcome categories, which include a very wide range of measures that are often grouped together in the early childhood development literature.

Health outcomes include anthropometric measures such as weight, height, and their standardized equivalents (WAZ, or z-scores for weight-for-age and HAZ, or z-scores for height-for-age), as well as indicators for general child health. For adults, health included both physical and mental health. *Motor development* encompasses indicators of fine and gross motor development.

Appendix 2. Process for standardizing outcomes across studies

Continuous variables

To standardize continuous variables that were not already standardized, we needed to scale the estimated coefficient (or "raw" effect size, ES_{raw}) by the standard deviation (sd) of the outcome variable for the control group at follow-up to get the standardized effect d.

$$d = \frac{ES_{raw}}{SD} \qquad (1)$$

Very few studies, however, report this standard deviation. When the standard deviation was not available, we use an approximation for standard deviation using the sample sizes of treatment (N_T) and control (N_C) groups:

$$d = \frac{ES_{raw}}{SE \times \sqrt{\frac{N_T \times N_C}{(N_T + N_C)}}} \qquad (2)$$

where SE corresponds to the standard error of ES_{raw}.

When studies did not include even a standard error of the estimated coefficient, we used the p-value (or an approximation of it) to estimate the standardized effect size for outcomes using the following formula:

$$SE = \frac{ES_{raw}}{-qnorm(p_{value}/2)} \qquad (3)$$

where *qnorm* gives us the z-score of the p^{th} quantile of the normal distribution.

Some studies reported baseline or follow-up sample sizes only for the entire sample, not separately for the treatment and control groups. In these cases, we assume that the total sample size was equally distributed across experimental groups, unless information in the text or elsewhere indicated otherwise. For example, in the Baker et al., (2008) differences-in-differences study, we assigned 23 percent of the sample to the treatment group since 23 percent of the Canadian population lives in Quebec. When the follow-up sample size for an estimation is not specified, we assume the sample size reported at baseline and adjust for attrition, if reported.

We use the following formula to calculate the standard errors associated with d, where t corresponds to the t-statistic corresponding to the raw estimated coefficient.

$$SE_d = \frac{\mathrm{d}}{t} \qquad (4)$$

Finally, if a study only reported stars to indicate statistical significance at the 1, 5 or 10 percent levels, we calculate the t-stat, assuming the following p-values: 0.005, 0.04, and 0.09, respectively.[1]

Binary variables

To standardize binary variables, we first have to convert raw effect sizes into odds ratios (OR) using follow-up means of success (enrollment rate, for example) and failure (non-enrollment rate).

$$OR = \frac{Mean_T/(1-Mean_T)}{Mean_C/(1-Mean_C)} \qquad (5)$$

For randomized studies conducted at the individual level that report raw means at follow-up, we would need the outcome variable means at follow-up— $Mean_C$, the outcome mean for the control group at follow-up and and $Mean_T = Mean_C + ES_{raw}$, the mean of the treatment group at follow-up. Because some studies do not report raw means at follow-up and instead report estimates of impact adjusted by baseline characteristics or fixed effects, we can also add the impact estimate of treatment obtained from the regression model to the follow-up mean of the control group to get the covariate-adjusted success rate in the treatment group.

When treatment is assigned at an aggregated level (for example, schools or communities) rather than the individual level, estimated standard errors must also take into account intracluster correlation in the outcome variable. Otherwise, we might overstate the precision of the estimates. We follow Wilson (2011) and convert the logged OR (ln (OR)) into a standardized effect size (d). As the logistic distribution is similar to the normal distribution, we can convert each ln (OR) into a d using the following formula:

$$d = \frac{\ln(OR)}{1.814} \qquad (4)$$

[1] Some studies only reported "Not significant" or p-value = 0 or p-value = 1. For these cases, we assume p-value = 0.55, 0.005, and 0.99, respectively.

Then, the standard error of d can be calculated using the standard error of the coefficient estimate for the treatment indicator from the regression model as follows:

$$SE_d = \frac{\mathrm{d}}{z} \qquad (5)$$

where z is either a z- or t-test associated with the treatment effect from the regression model. As we did for continuous variables, we used an approximation to calculate the t-stat when studies only reported levels of significance.

In sum, to calculate standardized effects sizes and their standard errors within each study for binary variables, we needed the outcome mean in the control group at follow-up and the estimated coefficient. These two elements allowed us to calculate the covariate-adjusted follow-up mean outcome in the treatment group. Then, we calculate ln(OR) and convert it into d using Equation (4). Finally, we calculate the standard error of d using the formula (5).

When the follow-up mean was not reported for the sample or for a subgroup (T or C), we assumed that the follow-up mean was equal to the baseline mean for the control group. We excluded a study from a regression if it only reported an impact estimate without any reference to a control mean at baseline or follow-up. We also excluded studies from a regression when the follow-up mean outcome in the control group plus the covariate-adjusted impact estimate from the regression model was larger than 1, as in these cases, it is not possible to calculate an odds ratio.

Appendix 3. Tables

Appendix Table 1: Characteristics of the included studies for review

	Country (1)	Program Name (2)	Study (2)	Intervention Components (3)	Discipline (4)	Publication Type (5)	IE Design (6)
	HIC - High Income Countries						
1	Canada	Quebec Universal Childcare Program	(Baker et al. 2008)	Subsidized or free access	Economics	Journal article	DID
2	Chile	Chile - Un Buen Comienzo	(Bowne et al. 2016b)	Teacher PD/ Pedagogy change/ Provision of materials	Education	Journal article	RCT + HLM
3	Chile	Chile - Un Buen Comienzo	(Yoshikawa et al. 2015)	Teacher PD/ Pedagogy change/ Provision of materials	Psychology	Journal article	RCT
4	Denmark	Denmark Preschool Expansion (1933-1960)	(Rossin-Slater and Wust 2020)	Preschool construction	Economics	Journal article	DID
5	Norway	Norway Universal Childcare Program I	(Drange et al 2016)	Subsidized or free access/ Curriculum	Economics	Journal article	DID
6	Norway	Norway Universal Childcare Program II	(Havnes and Mogstad 2011)	Subsidized or free access	Economics	Journal article	DID
7	Norway	Norway Universal Childcare Program II	(Havnes and Mogstad 2015)	Subsidized or free access	Economics	Journal article	DID + QR
8	Spain	Spain Universal Childcare Program	(Felfe et al 2014)	Subsidized or free access/ Teacher PD/ Curriculum	Economics	Journal article	DID
9	U.S.	USA High/Scope Perry Preschool Program	(Heckman et al 2010)	Subsidized or free access	Economics	Journal article	RCT
10	U.S.	USA Boston Pre-K - Building Blocks Curriculum	(Clements et al. 2011)	Teacher PD/ Curriculum/ Provision of materials	Education	Journal article	RCT + HLM
11	U.S.	USA Boston Pre-K - Building Blocks Curriculum	(Clements et al. 2013)	Teacher PD/ Curriculum/ Provision of materials	Education	Journal article	RCT + HLM
12	U.S.	USA Boston Pre-K - Subsidized Preschool	(Weiland and Yoshikawa 2013)	Subsidized or free access/ Teacher PD/ Curriculum	Education	Journal article	RDD
13	U.S.	USA Boston Pre-K - Subsidized Preschool	(Gray-Lobe et al 2021)	Subsidized or free access/ Teacher PD/	Economics	Journal article	RCT + IV

				Curriculum			
14	U.S.	USA Great Start Teacher PD Initiative	(Neuman and Cunningham 2009)	Teacher PD	Education	Journal article	RCT + ANOVA
15	U.S.	USA Head Start - CSRP	(Raver et al. 2008)	Teacher PD/ Health and nutrition	Education	Journal article	RCT + HLM
16	U.S.	USA Head Start - CSRP	(Raver et al. 2009)	Teacher PD/ Health and nutrition	Psychology	Journal article	RCT + HLM
17	U.S.	USA Head Start - CSRP	(Raver et al. 2011)	Teacher PD/ Health and nutrition	Education	Journal article	RCT + HLM
18	U.S.	USA Head Start - CSRP	(Watts et al 2018)	Teacher PD/ Health and nutrition	Education	Journal article	RCT
19	U.S.	USA Head Start - Increased Spending	(Johnson and Jackson 2019)	Subsidized or free access	Economics	Journal article	DID + IV
20	U.S.	USA Head Start - REDI Program	(Bierman et al. 2008)	Teacher PD/ Curriculum/ Provision of materials/ Parental engagement	Education	Journal article	RCT + HLM
21	U.S.	USA Head Start - Subsidized Preschool	(Bitler et al 2014)	Subsidized or free access	Economics	Working Paper	RCT + IV
22	U.S.	USA Head Start - Subsidized Preschool	(Carneiro and Ginja 2014)	Subsidized or free access	Economics	Journal article	RDD
23	U.S.	USA Head Start - Subsidized Preschool	(Frisvold and Lumeng 2011)	Subsidized or free access/ Preschool day extension	Economics	Journal article	DID
24	U.S.	USA Head Start - Subsidized Preschool	(Kline and Walters 2016)	Subsidized or free access	Economics	Journal article	RCT
25	U.S.	USA Head Start - Subsidized Preschool	(Bloom and Weiland 2015)	Subsidized or free access	Education	Working Paper	RCT + HLM
26	U.S.	USA Head Start - Subsidized Preschool	(Bailey et al 2020)	Subsidized or free access	Economics	Journal article	DID
27	U.S.	USA Head Start - Teacher PD & PATHS Curriculum	(Hamre et al 2012b)	Teacher PD/ Curriculum	Education	Journal article	RCT + HLM
28	U.S.	USA Head Start - Teacher PD Program II	(Pianta et al. 2017)	Teacher PD	Education	Journal article	RCT + HLM
29	U.S.	USA Head Start - Teacher PD Program I	(Powell et al. 2010)	Teacher PD	Education	Journal article	RCT + HLM
30	U.S.	USA Oklahoma/Tulsa Universal Pre-K	(Gormley et al. 2011)	Subsidized or free access	Education	Journal article	PSM

31	U.S.	USA ParentCorp Project	(Brotman et al. 2016)	Teacher PD/ Curriculum/ Parental engagement	Education	Journal article	RCT + PSM
32	U.S.	USA Pre-K Programs (State level)	(Wong et al. 2008)	Subsidized or free access	Economics	Journal article	RDD
33	U.S.	USA Tennessee Pre-K	(Lipsey et al. 2018)	Subsidized or free access	Education	Journal article	RCT + HLM
34	U.S.	USA Tools of the Mind Curriculum	(Barnett et al. 2008)	Teacher PD/ Curriculum/ Provision of materials	Education	Journal article	RCT
35	U.S.	USA Tools of the Mind Curriculum	(Blair and Raver 2014)	Teacher PD/ Curriculum	Psychology	Journal article	RCT
36	U.S.	USA Tools of the Mind Curriculum	(Diamond et al 2007)	Teacher PD/ Curriculum	Economics	Journal article	RCT
	LMIC - Low- & Middle-Income Countries						
37	Argentina	Argentina Preschool Construction	(Berlinski et al. 2009)	Preschool construction/ Curriculum	Economics	Journal article	DID
38	Bangladesh	Bangladesh Early Years Preschool Program	(Spier at al 2020)	Subsidized or free access	Education	Technical Report	RCT + IV
39	Cambodia	Cambodia Preschool Construction II	(Berkes et al 2019)	Preschool construction/ Community outreach/ Parental engagement	Economics	Working Paper	RCT
40	Cambodia	Cambodia Preschool Construction I	(Bouguen et al. 2018)	Preschool construction/ Teacher PD/ Provision of materials	Economics	Journal article	RCT
41	Colombia	Colombia Preschools - HIM & FE Programs	(Andrew et al. 2019)	Teacher PD/ Provision of materials/ Provision of staff	Economics	Working Paper	RCT
42	Gambia	Gambia Preschool Construction	(Blimpo et al 2019)	Preschool construction/ Curriculum/ Teacher PD	Education	Working Paper	RCT
43	Ghana	Quality Preschool for Ghana (QP4G)	(Wolf 2019)	Teacher PD/ Parental engagement	Education	Journal article	RCT + HLM
44	Ghana	Quality Preschool for Ghana (QP4G)	(Wolf and Peel 2019)	Teacher PD/ Parental engagement	Education	Journal article	RCT + HLM
45	Ghana	Quality Preschool for Ghana (QP4G)	(Wolf et al. 2018)	Teacher PD/ Parental engagement	Education	Journal article	RCT + HLM
46	Ghana	Quality Preschool for Ghana (QP4G)	(Wolf et al. 2019)	Teacher PD/ Parental engagement	Education	Journal article	RCT + HLM

47	India	India Pratham Game-Based Math Curriculum	(Dillon et al. 2017)	Curriculum/ Provision of materials	Economics	Journal article	RCT
48	India	India's Integrated Child Development Services (ICDS)	(Ganimian et al 2021)	Provision of staff	Economics	Journal article	RCT
49	Indonesia	Indonesia Preschool Construction	(Brinkman et al. 2017)	Preschool construction/ Teacher PD/ Community outreach	Education	Journal article	DID + IV
50	Malawi	Malawi Community-based Childcare Centers	(Ozler et al. 2018)	Teacher PD/ Teacher Payments/ Parental engagement	Economics	Journal article	RCT
51	Mozambique	Mozambique Preschool Construction	(Martinez et al. 2017)	Preschool construction/ Teacher PD/ Parental engagement	Economics	Working Paper	RCT
52	Paraguay, Peru	Inquiry and Problem-Based Pedagogy (IPP)	(Bando et al 2019)	Teacher PD/ Pedagogy change/ Provision of materials	Economics	Working Paper	RCT
53	Peru	Inquiry and Problem-Based Pedagogy (IPP)	(Gallego et al 2019)	Teacher PD/ Pedagogy change/ Provision of materials	Economics	Journal article	RCT
54	Uruguay	Uruguay Preschool Construction	(Berlinski et al. 2008)	Preschool construction	Economics	Journal article	IV

Appendix Table 2: Characteristics of the programs evaluated

Country	Program name	Intervention components	Provision type	Target group (years)	Target population	Type of location	Implementer type	Number of studies in review
(1)	(2)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HIC - High Income Countries								
Canada	Quebec Universal Childcare Program	Subsidized or free access	Daycare (w/education component)	3-4	General	Urban & Rural	Government	1
Chile	Chile - Un Buen Comienzo	Teacher PD/ Pedagogy change/ Provision of materials	Formal classroom	4-5	Disadvantaged	Urban	Government	2
Denmark	Denmark Preschool Expansion (1933-1960)	Preschool construction	Preschool - formal classroom	3-7	Disadvantaged	Urban & Rural	Government + Private	1
Norway	Norway Universal Childcare Program I	Subsidized or free access/ Curriculum	Formal classroom	5-6	General	Urban & Rural	Government	1
Norway	Norway Universal Childcare Program II	Subsidized or free access	Daycare (w/education component)	3-6	General	Urban & Rural	Government + Private	2
Spain	Spain Universal Childcare Program	Subsidized or free access/ Teacher PD/ Curriculum	Daycare (w/education component)	3	General	Urban & Rural	Government + Private	1
U.S.	USA High/Scope Perry Preschool Program	Subsidized or free access	Preschool - formal classroom	3-5	Disadvantaged	Urban & Rural	Government	1
U.S.	USA Boston Pre-K - Building Blocks Curriculum	Teacher PD/ Curriculum/ Provision of materials	Formal classroom	3-5	Disadvantaged	Urban	Government	2
U.S.	USA Boston Pre-K - Subsidized Preschool	Subsidized or free access/ Teacher PD/ Curriculum	Formal classroom	4-5	General	Urban	Government	2
U.S.	USA Great Start Teacher PD Initiative	Teacher PD	Formal classroom	3-5	Disadvantaged	Urban	Government + NGO	1
U.S.	USA Head Start - CSRP	Teacher PD/ Health and nutrition	Formal classroom	3-5	Disadvantaged	Urban & Rural	Government + Private	4
U.S.	USA Head Start - Increased Spending	Subsidized or free access	Formal classroom	4	Disadvantaged	Urban & Rural	Government	1
U.S.	USA Head Start - REDI Program	Teacher PD/ Curriculum/ Provision of materials/ Parental engagement	Formal classroom	4	Disadvantaged	Urban & Rural	No info	1
U.S.	USA Head Start - Subsidized Preschool	Subsidized or free access	Formal classroom	4	Disadvantaged	Urban & Rural	Government + Private	6
U.S.	USA Head Start - Teacher PD & PATHS Curriculum	Teacher PD/ Curriculum	Formal classroom	3-4	Disadvantaged	No info	Government	1
U.S.	USA Head Start - Teacher PD Program I	Teacher PD	Formal classroom	4	Disadvantaged	Urban & Rural	No info	1
U.S.	USA Head Start - Teacher PD Program II	Teacher PD	Formal classroom	4	Disadvantaged	Urban	Government + Private	1
U.S.	USA Oklahoma/Tulsa Universal Pre-K	Subsidized or free access	Formal classroom	4	Disadvantaged	Urban	Government	1

U.S.	USA ParentCorp Project	Teacher PD/ Curriculum/ Parental engagement	Daycare (w/education component)	4	Disadvantaged	Urban	Government	1
U.S.	USA Pre-K Programs (State level)	Subsidized or free access	Formal classroom	4	General	Urban & Rural	Government	1
U.S.	USA Tennessee Pre-K	Subsidized or free access	Formal classroom	4-5	Disadvantaged	Urban & Rural	Government	1
U.S.	USA Tools of the Mind Curriculum	Teacher PD/ Curriculum	Formal classroom	4-5	Disadvantaged	Urban	Government	3
LMIC - Low- & Middle-Income Countries								
Argentina	Argentina Preschool Construction	Preschool construction/ Curriculum	Formal classroom	3-5	Disadvantaged	Urban	Government	1
Bangladesh	Bangladesh Early Years Preschool Program	Subsidized or free access	Formal classroom	4	General	Rural	Government	1
Cambodia	Cambodia Preschool Construction I	Preschool construction/ Teacher PD/ Provision of materials	Community-based	3-6	Disadvantaged	Rural	Government	1
Cambodia	Cambodia Preschool Construction II	Preschool construction/ Community outreach/ Parental engagement	Community-based	3-5	General	Urban & Rural	Government	1
Colombia	Colombia Preschools - HIM & FE Programs	Teacher PD/ Provision of materials/ Provision of staff	Community-based	2-5	Disadvantaged	Urban & Rural	Government + NGO	1
Gambia	Gambia Preschool Construction	Preschool construction/ Curriculum/ Teacher PD	Community-based	3-6	Disadvantaged	Rural	Government	1
Ghana	Quality Preschool for Ghana (QP4G)	Teacher PD/ Parental engagement	Formal classroom	4-6	Disadvantaged	Urban & Rural	Government	4
India	India Pratham Game-Based Math Curriculum	Curriculum/ Provision of materials	Formal classroom	5	Disadvantaged	Urban	International Non-Governmental (ING)	1
India	India's Integrated Child Development Services (ICDS)	Provision of staff	Formal classroom	3-6	Disadvantaged	Urban & Rural	Government	1
Indonesia	Indonesia Preschool Construction	Preschool construction/ Teacher PD/ Community outreach	Community-based	3-6	Disadvantaged	Rural	Government + IO	1
Malawi	Malawi Community-based Childcare Centers	Teacher PD/ Teacher Payments/ Parental engagement	Community-based	3-5	Disadvantaged	Rural	Government + ING	1
Mozambique	Mozambique Preschool Construction	Preschool construction/ Teacher PD/ Parental engagement	Community-based	3-5	Disadvantaged	Rural	International Non-Governmental (ING)	1
Paraguay, Peru	Inquiry and Problem-Based Pedagogy (IPP)	Teacher PD/ Pedagogy change/ Provision of materials	Formal classroom	5	General	Urban & Rural	Government	2
Uruguay	Uruguay Preschool Construction	Preschool construction	Formal classroom	3-5	General	Urban & Rural	Government	1

Appendix Table 3: Mapping of subpopulations to high and low socioeconomic status.

Low socioeconomic status	**High socioeconomic status**
Assets below median	Assets above median
Black	Non-Black
Father not at home	Father at home
Free lunch	Not free lunch
Low income	High income
Mother less educated	Mother more educated
Mother without education	Mother with education
Mother without high school	Mother with high school
No parent with high school	One parent with high school
Parents with high school or less	Parents with more than high school
Poor	Non-poor
Single-parent	Two-parent
Stunted child	Unstunted child
Non-White	White
	Mother with college
Hispanic	
High-poverty schools	

Appendix Table 4: Parameter values and data sources for calculating benefit-to-cost ratio for preprimary interventions in low- and middle-income countries

	Labor force participation rate	Average nominal monthly wage (country currency)	Annual real wage growth	Lowest treatment effect on cognitive skills (SD)	Lowest treatment effect on social-emotional skills (SD)	Data sources
	(1)	(2)	(3)	(5)	(6)	(7)
India (Dillon et al, 2017)	46.3	13143	0	0.09	0.165	Periodic Labour Force Survey, 2019; India Ministry of Statistics and Programme Implementation; Bloomberg Newswire
Malawi (Ozler et al, 2018)	37.2	125000	3.4	0.185	0.252	Integrated Household Survey, 2017 ; National Statistical Office of Malawi
Ghana (Wolf et al, 2019a)	57	618	2.16	0.107	0.18	Living Standards Survey, 2017; Ghana Statistical Service; Bloomberg Newswire
Peru (Gallego et al, 2019)	77.4	1570	1.7	0.19	---	Encuesta Nacional de Hogares, 2019; ILO SIALC; Bloomberg Newswire
Indonesia (Brinkman et al, 2017)	70.1	2913897	-0.2	---	0.158	National Labour Force Survey, 2020; Statistics Indonesia of the Republic of Indonesia; Bloomberg Newswire
Bangladesh (Spier et al, 2020)	59.4	12016	3	0.25	0.37	Labour Force Survey, 2017; Bangladesh Bureau of Statistics

Notes: Missing recent data for average real wage growth for Ghana, we averaged values across other countries. Country currencies were the following: India (INR), Malawi (MWK), Mozambique (MZN), Ghana (GHS), Peru (PEN), Indonesia (IDR), and Bangladesh (BDT)

Appendix Table 5: Average effects of preprimary interventions on participation and skills *during* the preprimary period

				Expansion		Quality improvement	
	Pooled sample	High-income	Low- & middle-income	High-income	Low- & middle-income	High-income	Low- & middle-income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: School participation and progression							
School participation and progression	1.4048**						
	(0.4776)						
N	29						
Panel B: Cognitive skills							
Pooled skills	0.1474***	0.1787***	0.1223***	0.1117***	0.0762	0.2492**	0.1576***
	(0.0259)	(0.0431)	(0.0330)	(0.0270)	(0.0504)	(0.0794)	(0.0416)
N	131	76	55	25	33	51	22
Language	0.1084***	0.1150**	0.0992			0.1427*	
	(0.0295)	(0.0367)	(0.0526)			(0.0692)	
N	53	35	18			23	
Literacy	0.2162***	0.2385**				0.2402**	
	(0.0634)	(0.0783)				(0.0937)	
	31	28				22	
Math	0.2238***	0.3605**	0.1643***				
	(0.0469)	(0.1108)	(0.0406)				
	28	10	18				
General	0.0844*		0.0803				
	(0.0367)		(0.0464)				
	19		16				

Panel C: Executive functions and social-emotional skills							
Pooled skills	0.1209***	0.1354**	0.1138***			0.1830***	
	(0.0279)	(0.0494)	(0.0330)			(0.0531)	
N	114	75	39			66	
Executive functions	0.0951*	0.1693*					
	(0.0387)	(0.0778)					
N	42	31					
Social-emotional skills	0.1150***	0.0935**	0.1301**				
	(0.0301)	(0.0318)	(0.0463)				
N	50	22	28				
Behavior	0.0750	0.0750					
	(0.0500)	(0.0500)					
N	22	22					

Notes: Standard errors estimated using small sample adjustments as in Tipton (2015) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Number of observations (N) refers to number of study-specific average effects used in the specification. Results under the columns of "Expansion" or "Quality" are estimated average effects from studies evaluating interventions oriented to increase preschool expansion or to improve preschool quality, respectively. Specifications that pool skills use one large category for extracted outcomes - either cognitive skills (which combines language, literacy, math, and general cognitive skills) or what is typically referred to as non-cognitive skills (executive functions, social-emotional skills, and behaviors). See Appendix 1 for details on definitions of school participation and progression; language, literacy, math, and general skills; executive functions; social-emotional skills; and behavior. Cells are blank when the sample size is too small to be able to obtain trustworthy estimations, when the degrees of freedom fall below 4 or when all observations came from a single study.

Appendix Table 6: Average effects of preprimary interventions on participation and skills *after* the preprimary period

				Expansion		Quality improvement	
	Pooled sample	High-income	Low- & middle-income	High-income	Low- & middle-income	High-income	Low- & middle-income
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: School participation and progression							
School participation and progression	0.1423						
	(0.0888)						
N	22						
Panel B: Cognitive skills							
Pooled skills	0.0718***	0.0604	0.0862**	0.0205			0.0488*
	(0.0236)	(0.0375)	(0.0311)	(0.0304)			(0.0231)
N	62	23	39	18			26
Language							
N							
Literacy	0.0626						
	(0.0344)						
	14						
Math	0.0571*	0.0648	0.0512				0.0231
	(0.0311)	(0.0566)	(0.0361)				(0.0317)
	29	7	22				17
General	0.0603						
	(0.0476)						
	10						

Panel C: Executive functions and social-emotional skills							
Pooled skills	0.0686***		0.0750***				
	(0.0190)		(0.0178)				
N	70		34				
Executive functions							
N							
Social-emotional skills	0.0945**		0.1056**				
	(0.0342)		(0.0361)				
N	34		21				
Behavior							
N							

Notes: Standard errors estimated using small sample adjustments as in Tipton (2015) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Number of observations (N) refers to number of study-specific average effects used in the specification. Results under the columns of "Expansion" or "Quality" are estimated average effects from studies evaluating interventions oriented to increase preschool expansion or to improve preschool quality, respectively. Specifications that pool skills use one large category for extracted outcomes - either cognitive skills (which combines language, literacy, math, and general cognitive skills) or what is typically referred to as non-cognitive skills (executive functions, social-emotional skills, and behaviors). See Appendix 1 for details on definitions of school participation and progression; language, literacy, math, and general skills; executive functions; social-emotional skills; and behavior. Cells are blank when the sample size is too small to be able to obtain trustworthy estimations, when the degrees of freedom fall below 4 or when all observations came from a single study.

Appendix Table 7: Average effects of preprimary interventions on participation and skills during preprimary and post-preprimary

	Preprimary								
	School participation and progression			Cognitive skills			Executive functions and social-emotional skills		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Preprimary program	0.3030**	----	----	0.1474***	0.1222***	0.1915***	0.1209***	0.1150***	0.1576***
	(0.0932)			(0.0259)	(0.0332)	(0.0397)	(0.0279)	(0.0331)	(0.0305)
High-income		----	----		0.0541			0.0138	
					(0.0537)			(0.0586)	
Expansion		----	----			-0.0957**			-0.1066*
						(0.0482)			(0.0514)
N	22	----	----	131	131	131	114	114	114
	Post-preprimary								
	School participation and progression			Cognitive skills			Executive functions and social-emotional skills		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Preprimary program	0.0736	----	----	0.0718***	0.0863**	0.0856**	0.0686***	----	----
	(0.0539)			(0.0236)	(0.0311)	(0.0342)	(0.0190)		
High-income		----	----		-0.0260			----	----
					(0.0487)				
Expansion		----	----			-0.0239		----	----
						(0.0481)			
N	12	----	----	62	62	62	70	----	----

Notes: Standard errors estimated using small sample adjustments as in Tipton (2015) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Number of observations (N) refers to number of study-specific average effects used in the specification. Specifications that pool skills use one large category for extracted outcomes - either cognitive skills (which combines language, literacy, math, and general cognitive skills) or what is typically referred to as non-cognitive skills (executive functions, social-emotional skills, and behaviors). "High income" indicates whether the study reports results from high-income countries, while "Expansion" indicates whether the preprimary intervention increased preschool coverage as opposed to improving preschool quality. Cells are blank when the sample size is too small to be able to obtain trustworthy estimations, when the degrees of freedom fall below 4 or when all observations came from a single study.

Appendix Table 8: Average effects of preprimary on teacher practices and health

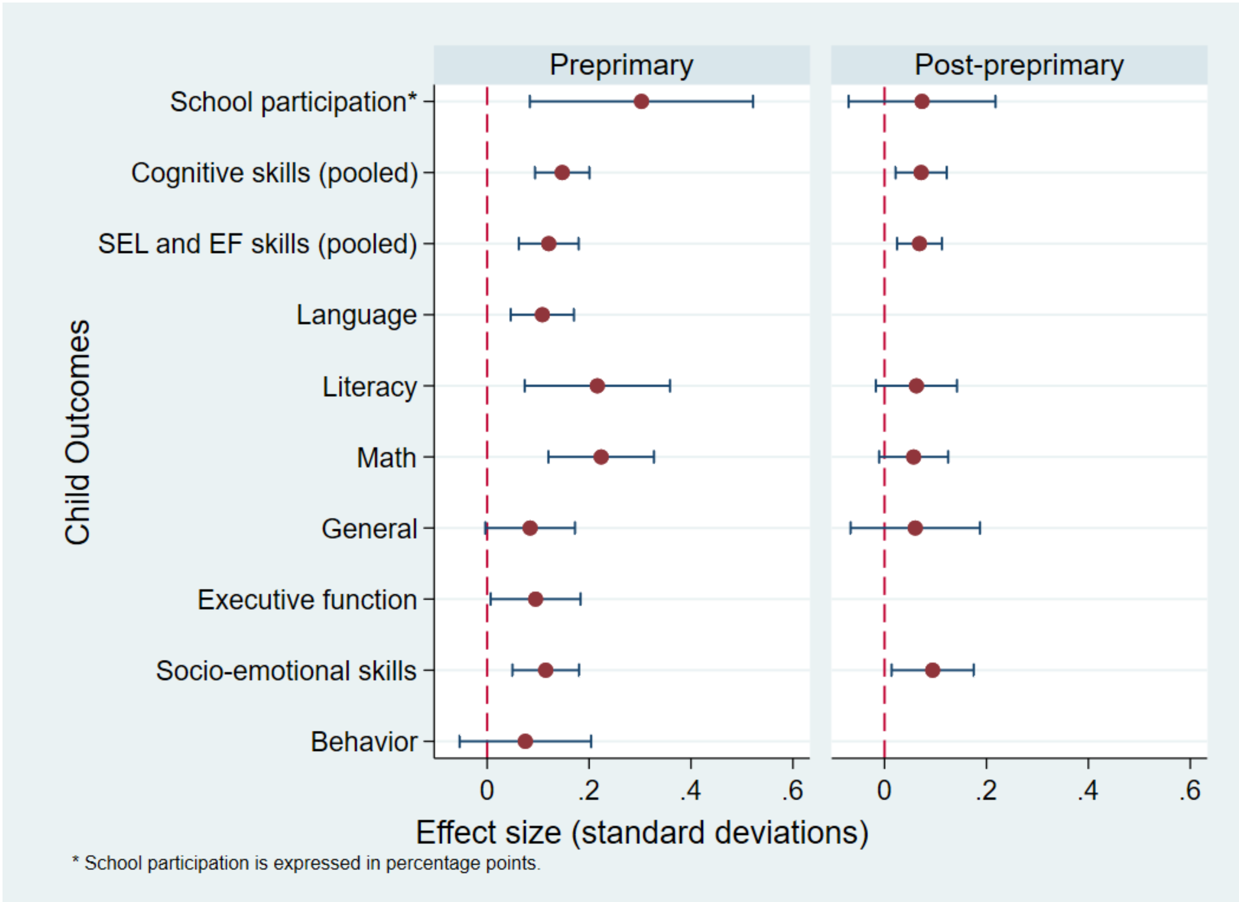
	Pooled sample	High-income	Low- & middle-income
	(1)	(2)	(3)
Teacher practice in the classroom	0.4729***		
	(0.0994)		
N	47		
Health during preprimary	0.0324		0.0280
	(0.0290)		(0.0312)
N	27		23

Notes: Standard errors estimated using small sample adjustments as in Tipton (2015) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1 Number of observations (N) refers to number of study-specific average effects used in the specification. Cells are blank when the sample size is too small to be able to obtain trustworthy estimations, when the degrees of freedom fall below 4 or when all observations came from a single study.

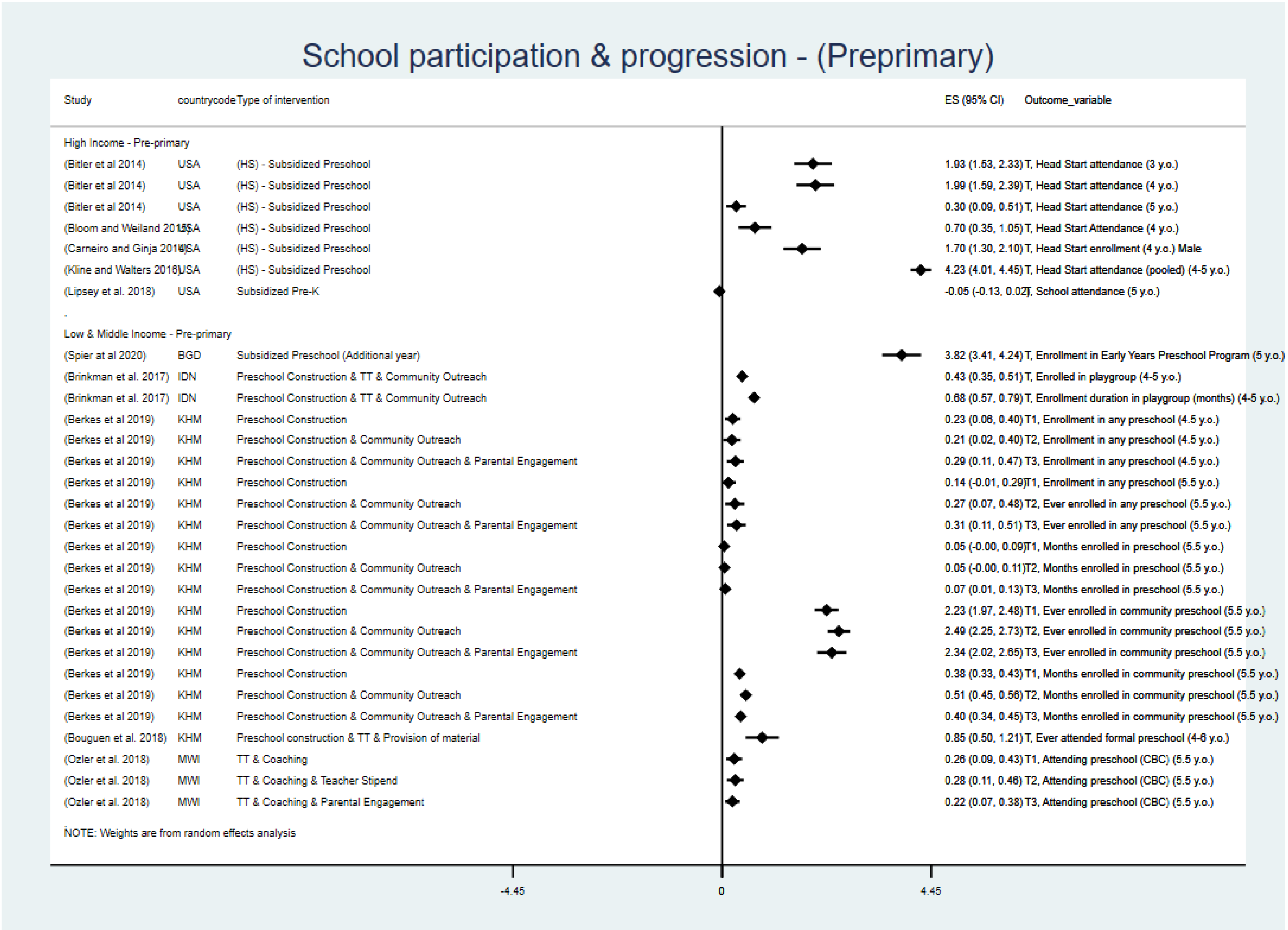
Appendix 4. Figures

Note: These Appendix Figures present study-specific average treatment effects of preprimary programs on outcomes expressed in terms of standard deviations. Results are drawn from 54 experimental and quasi experimental studies that report impacts on a wide variety of outcome categories. Effect sizes to the right of zero represent an improvement of the outcome, while effect sizes to the left of zero represent a worsening of the outcome. (*) indicates that the connotation of the outcome is negative, and the representation of the effect size has been adjusted to show either an improvement or worsening according to each specific case.

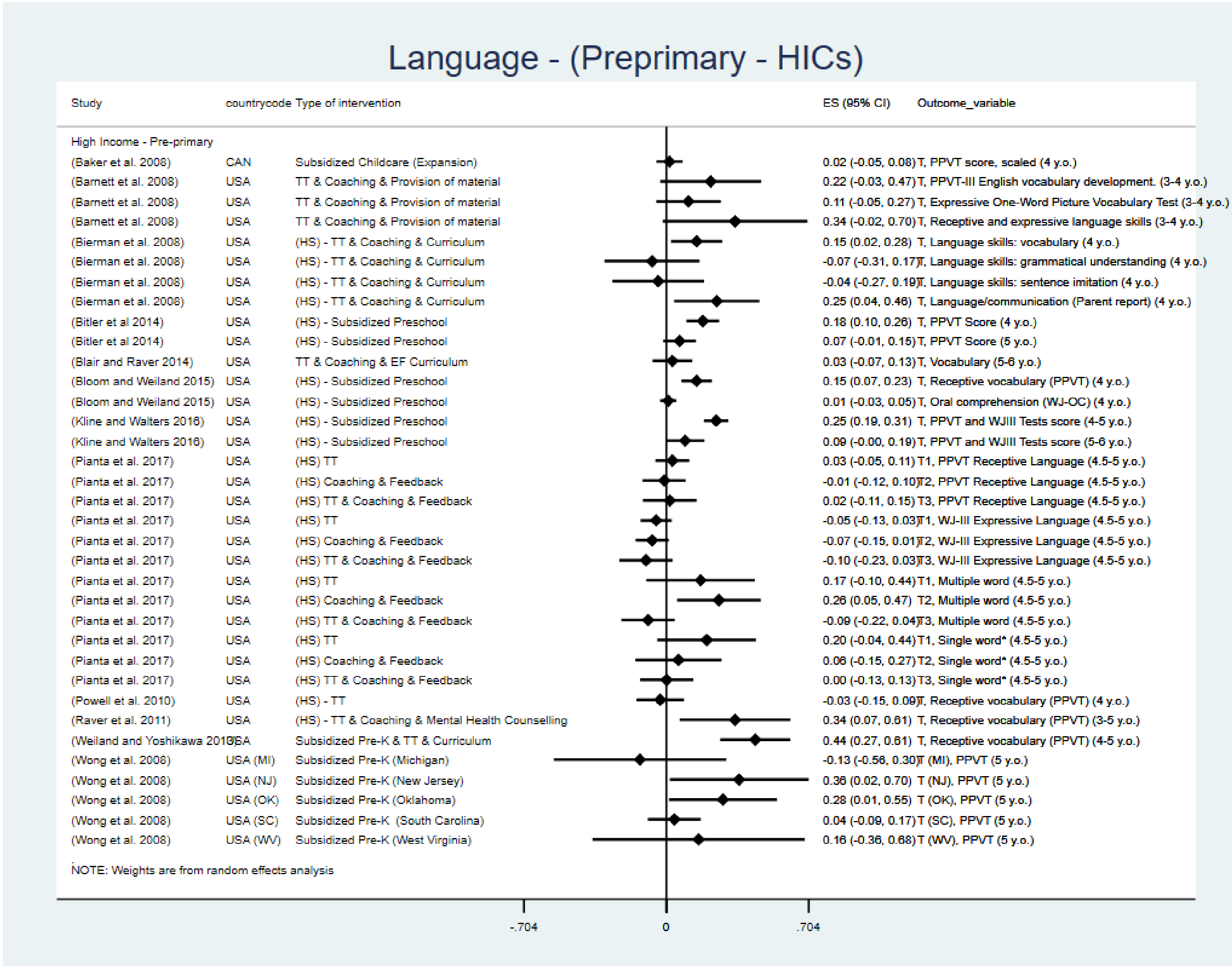
Appendix Figure 1: Average effects of preprimary interventions on participation and skills *during and after* the preprimary period



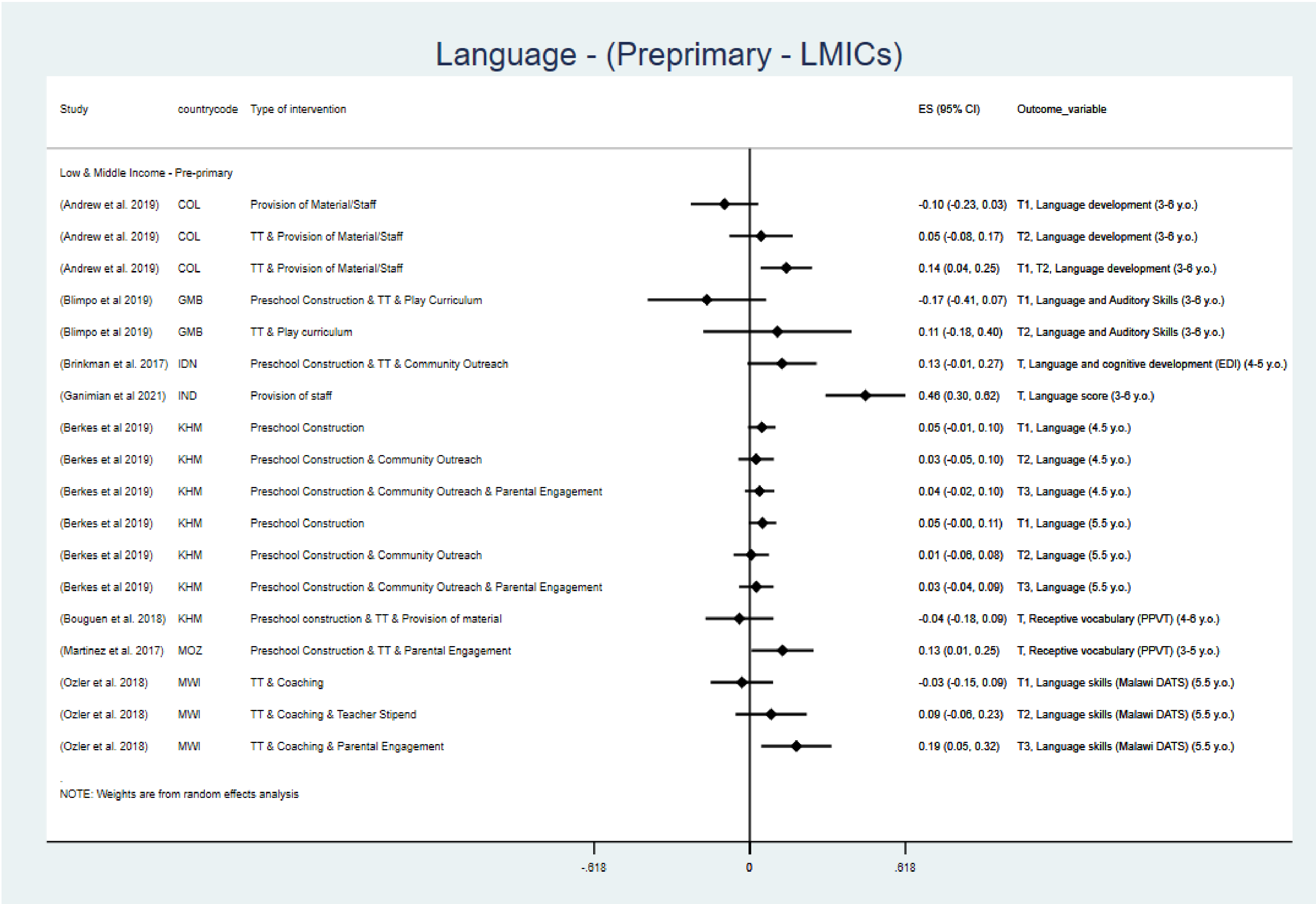
Appendix Figure 2: Effect sizes on school participation and progression (preprimary)



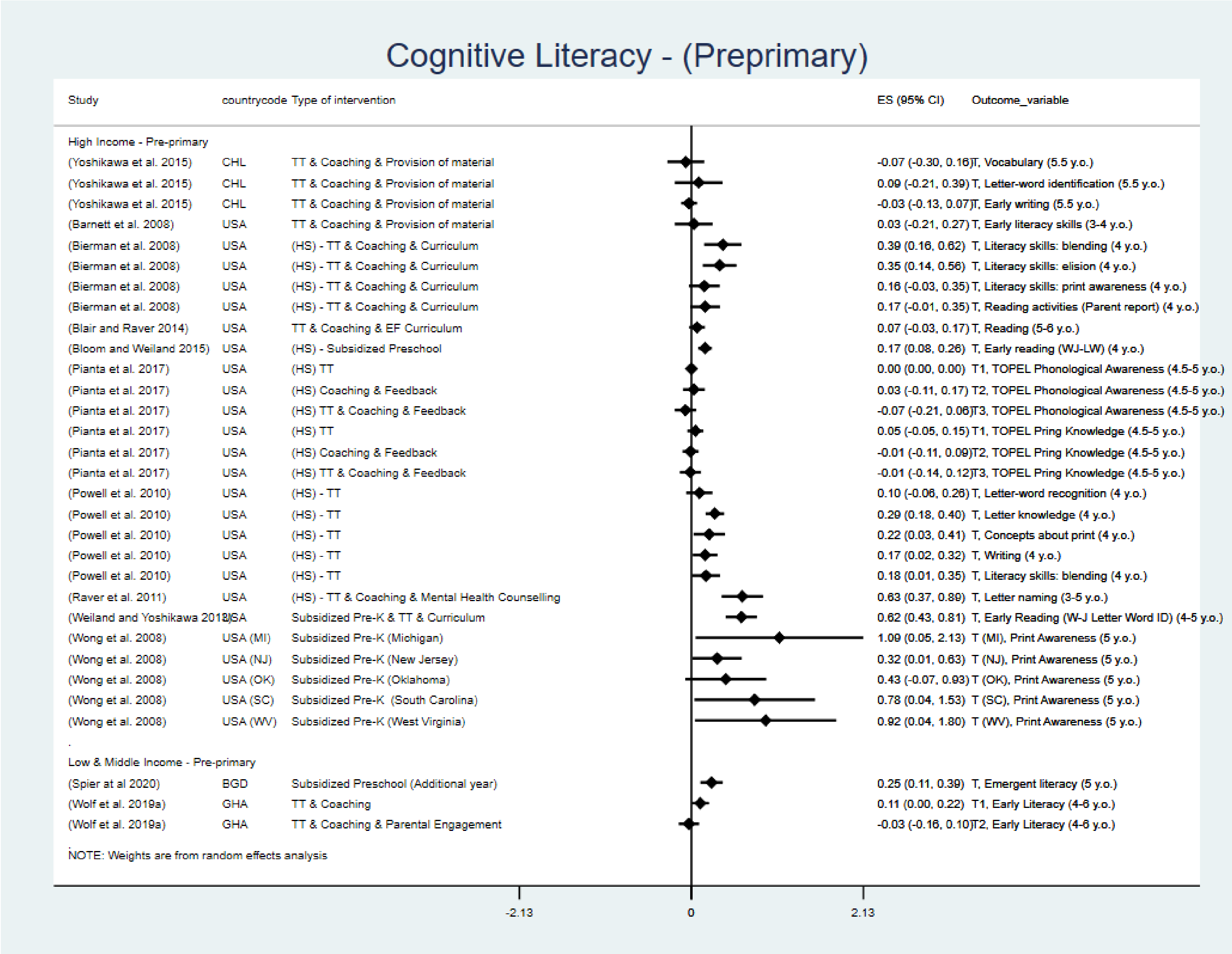
Appendix Figure 3: Effect sizes on language skills in HICs (preprimary)



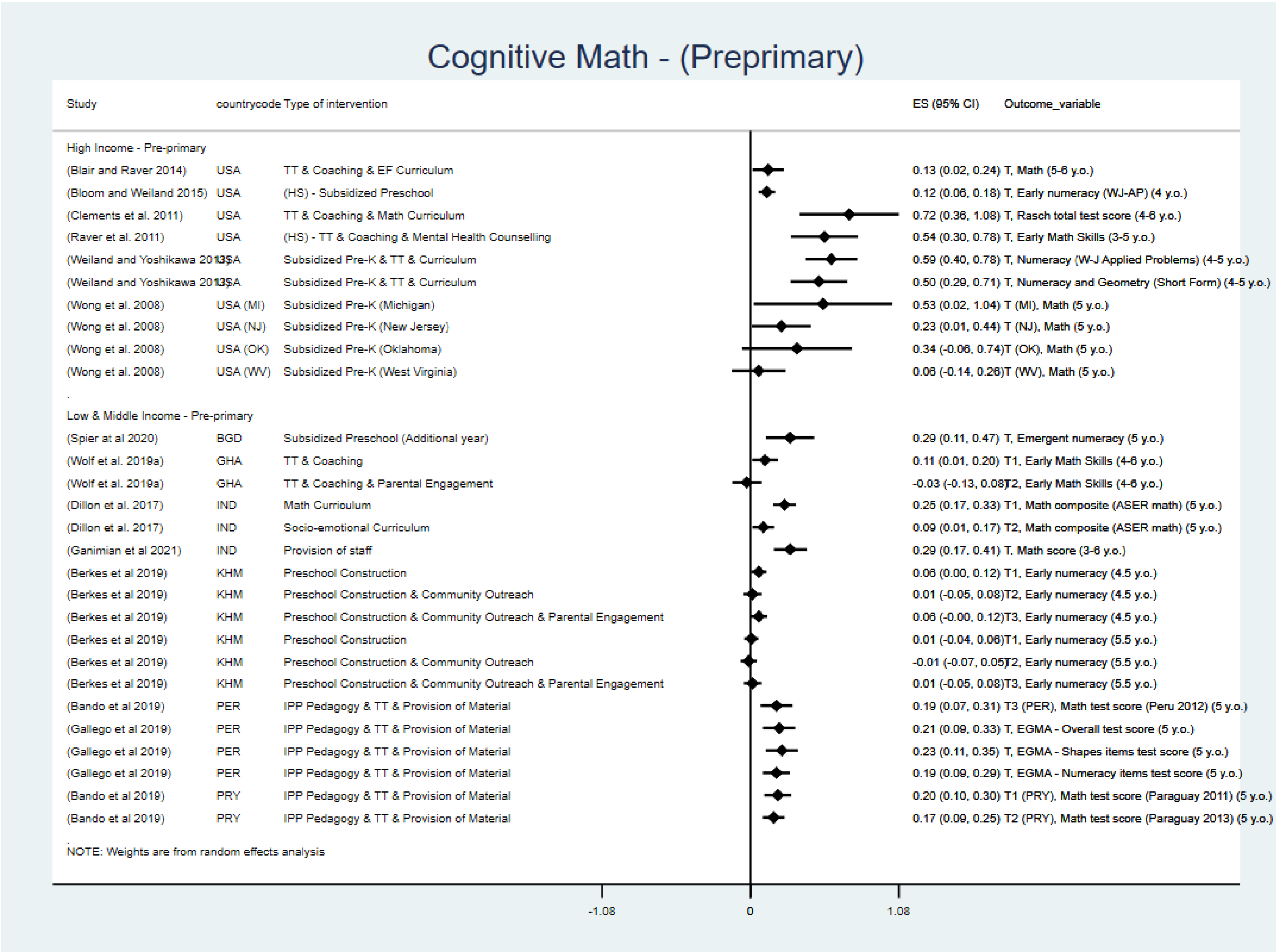
Appendix Figure 4: Effect sizes on language skills in LMICs (preprimary)



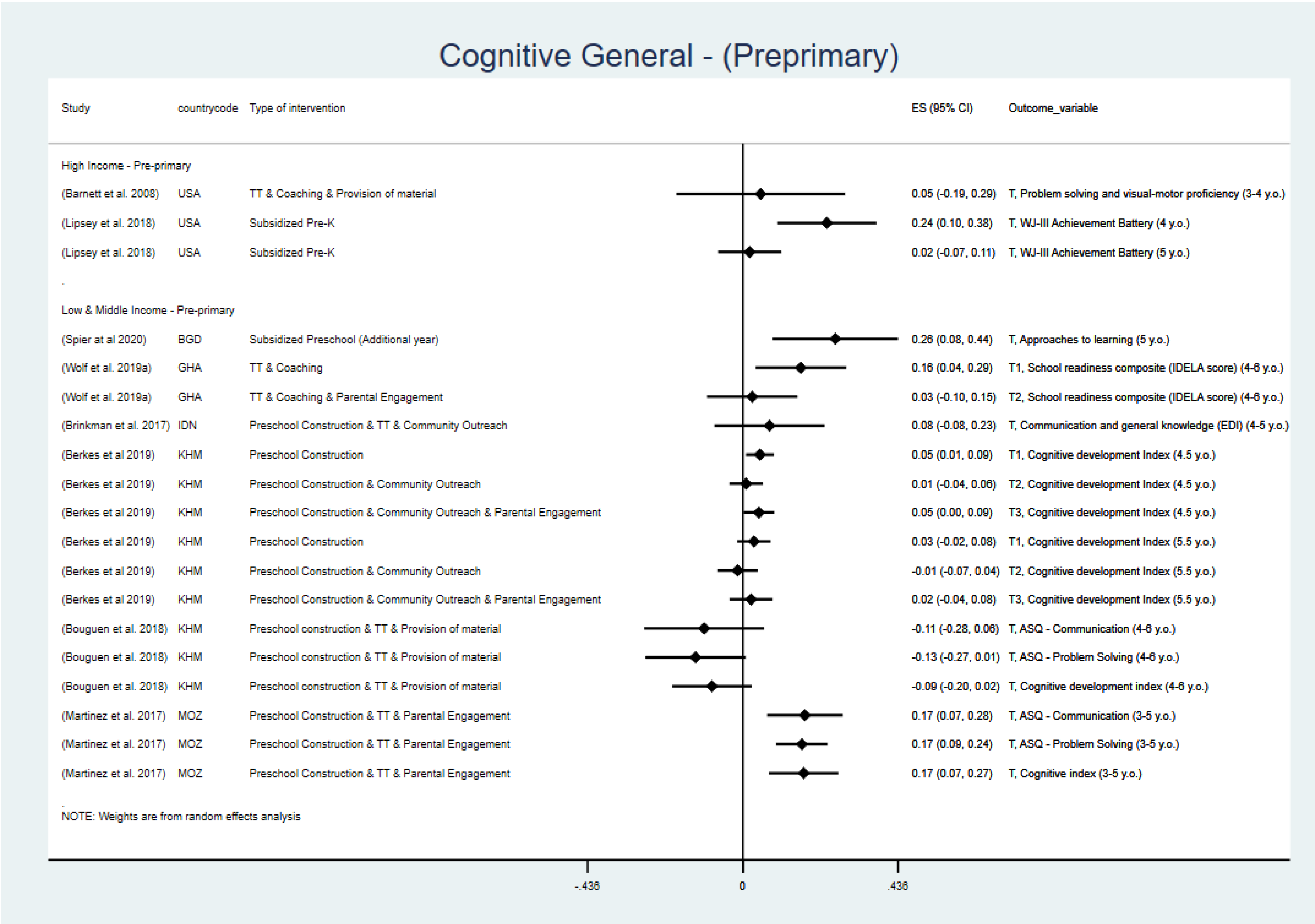
Appendix Figure 5: Effect sizes on cognitive literacy skills (preprimary)



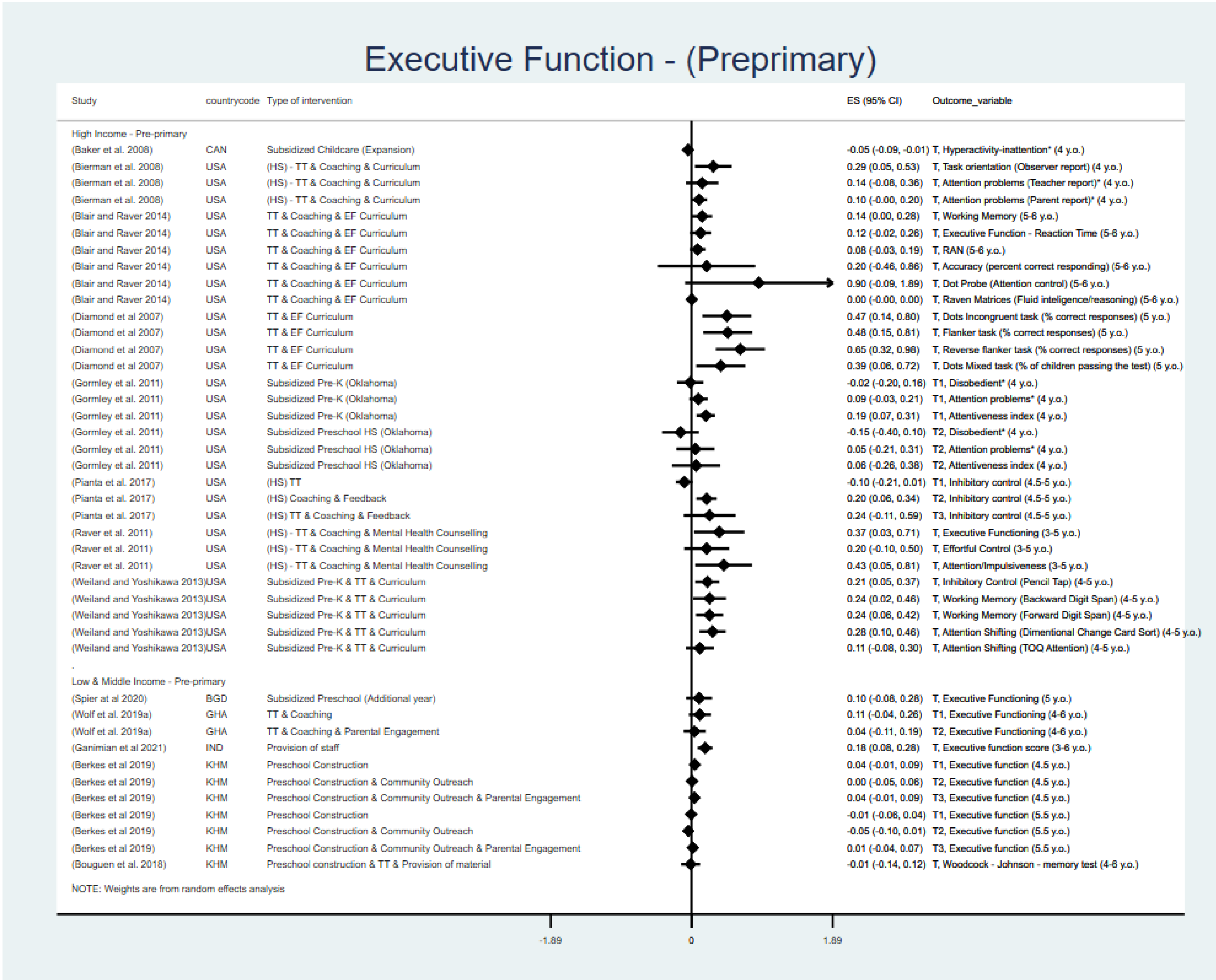
Appendix Figure 6: Effect sizes on cognitive math skills (preprimary)



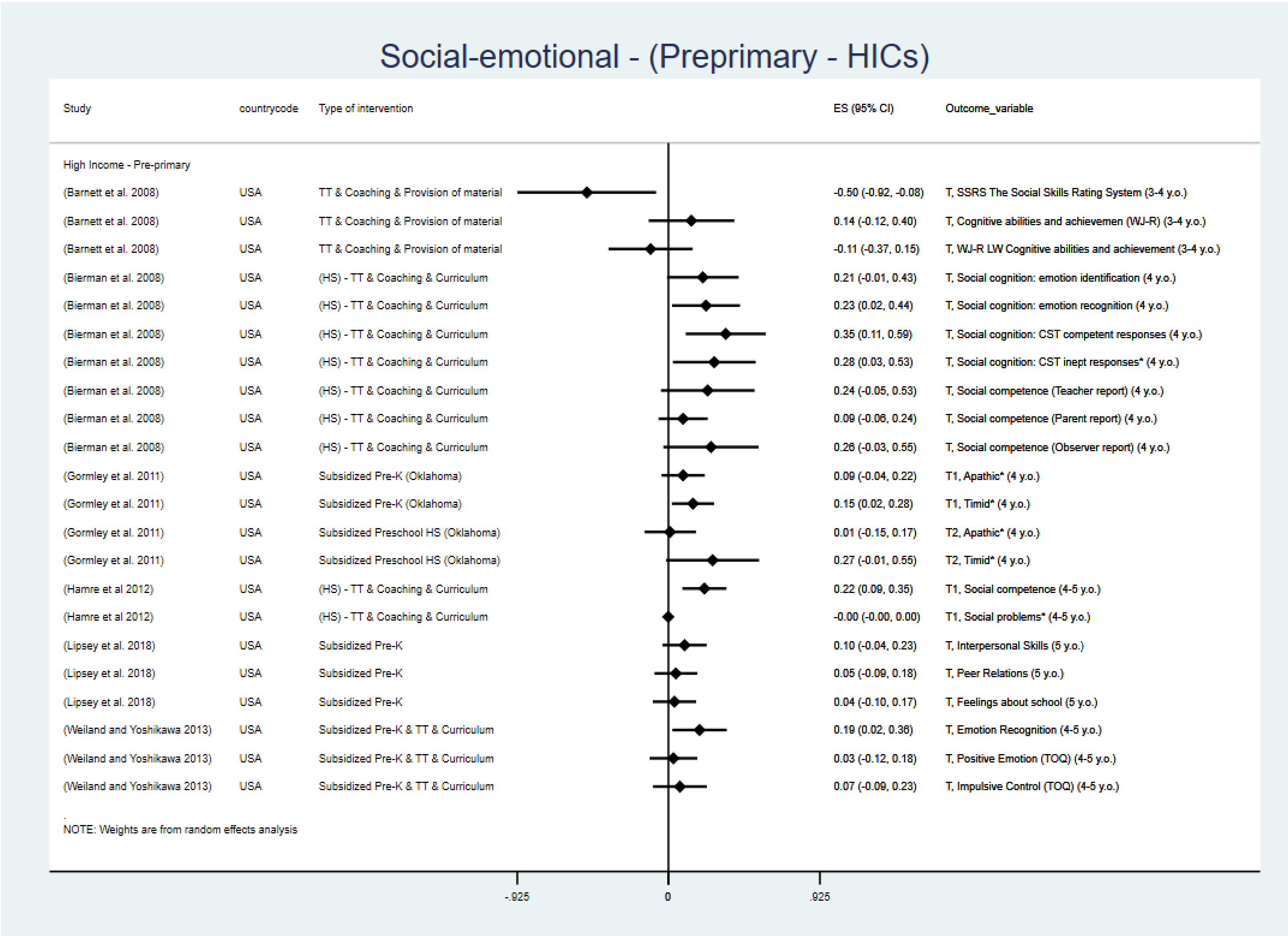
Appendix Figure 7: Effect sizes on cognitive general skills (preprimary)



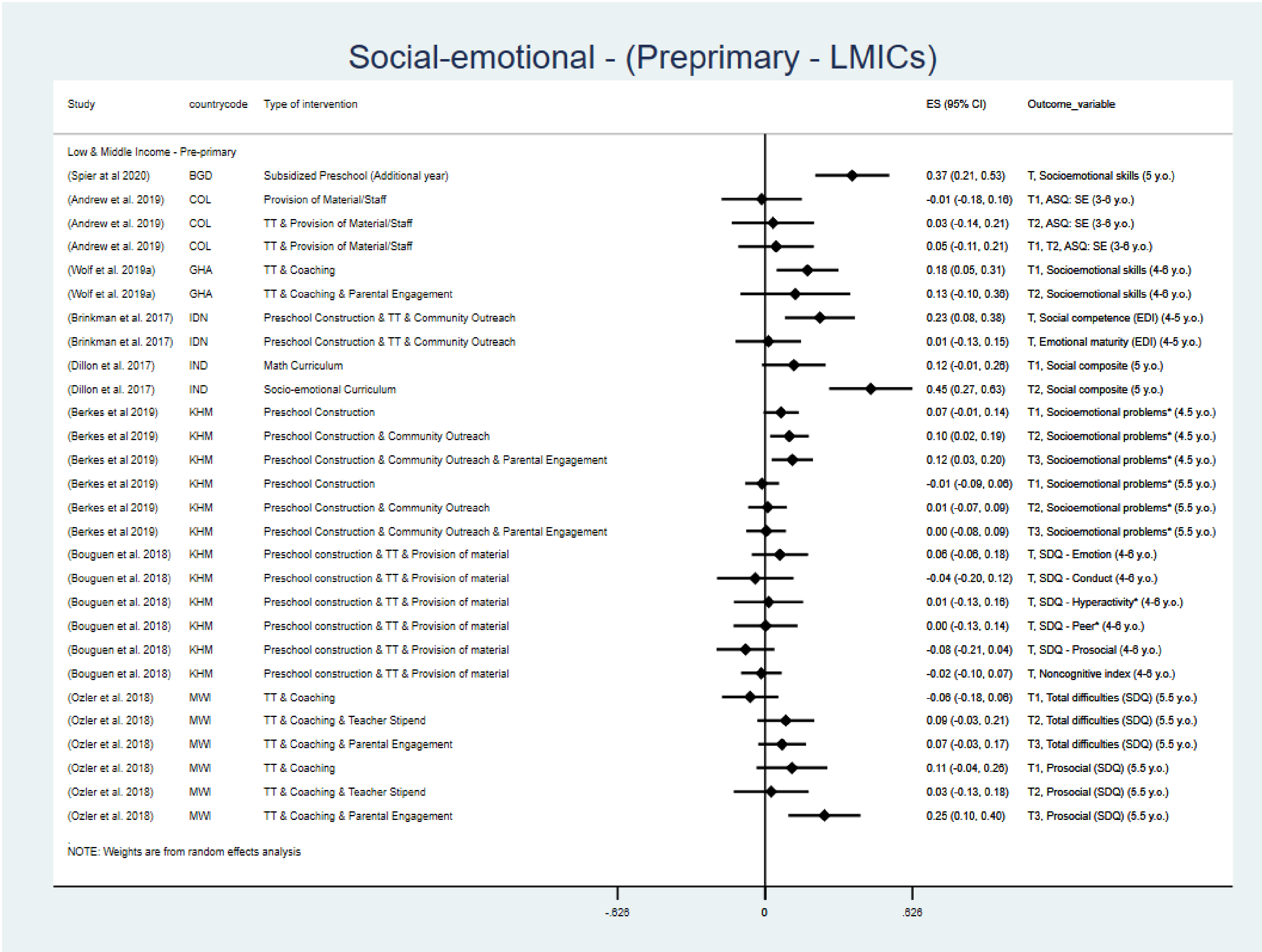
Appendix Figure 8: Effect sizes on executive function skills (preprimary)



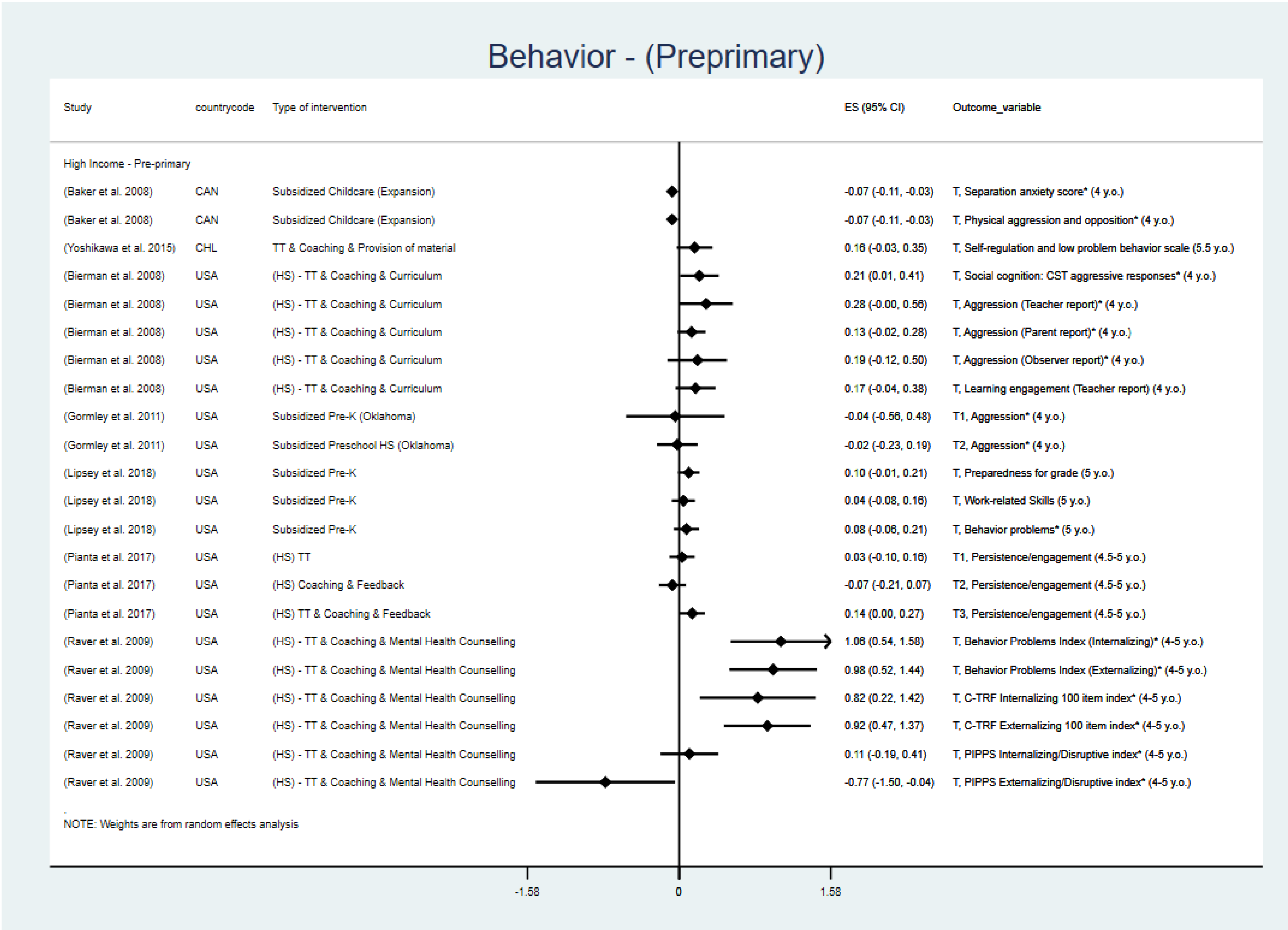
Appendix Figure 9: Effect sizes on social-emotional skills in HICs (preprimary)



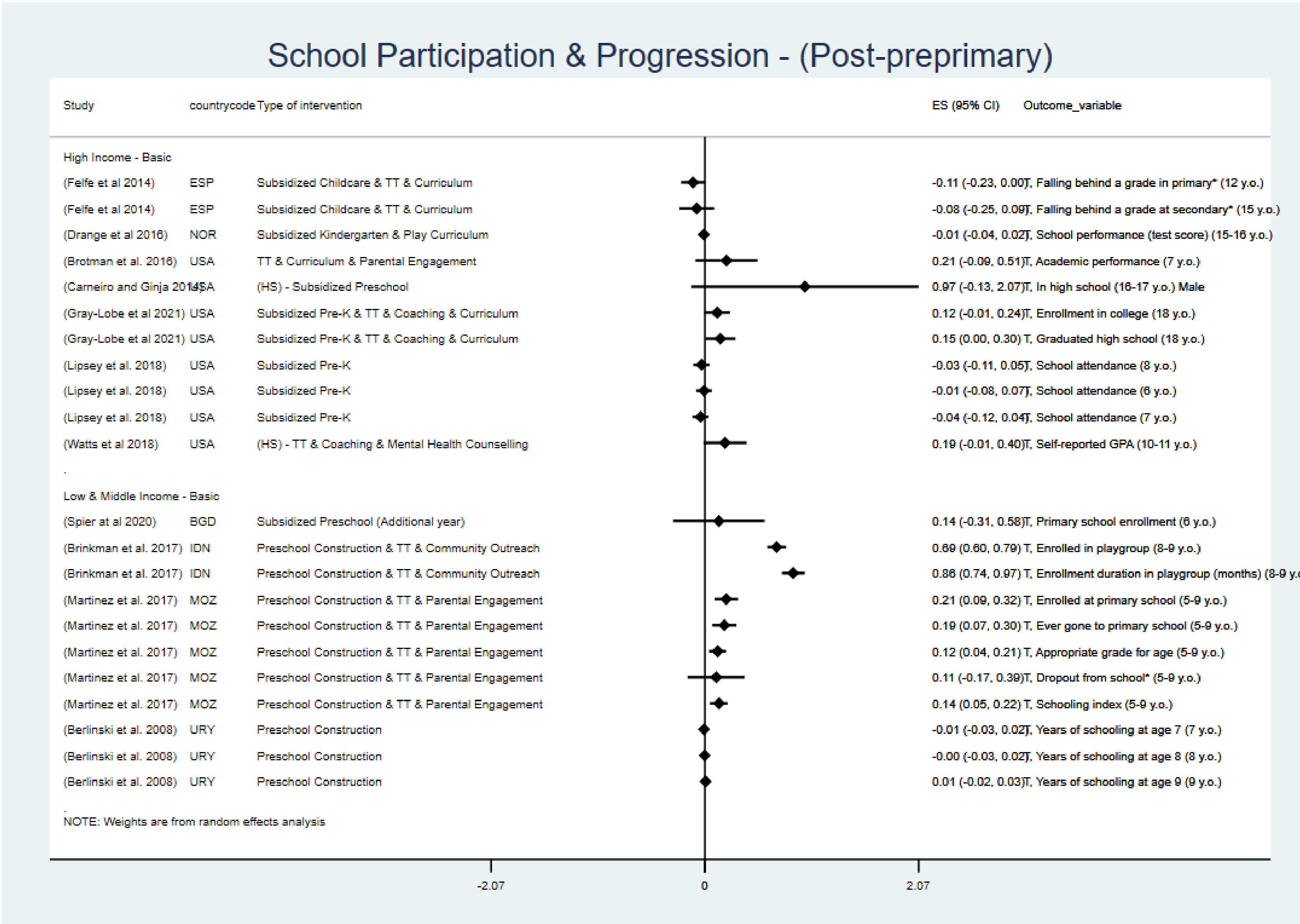
Appendix Figure 10: Effect sizes on social-emotional skills in LMICs (preprimary)



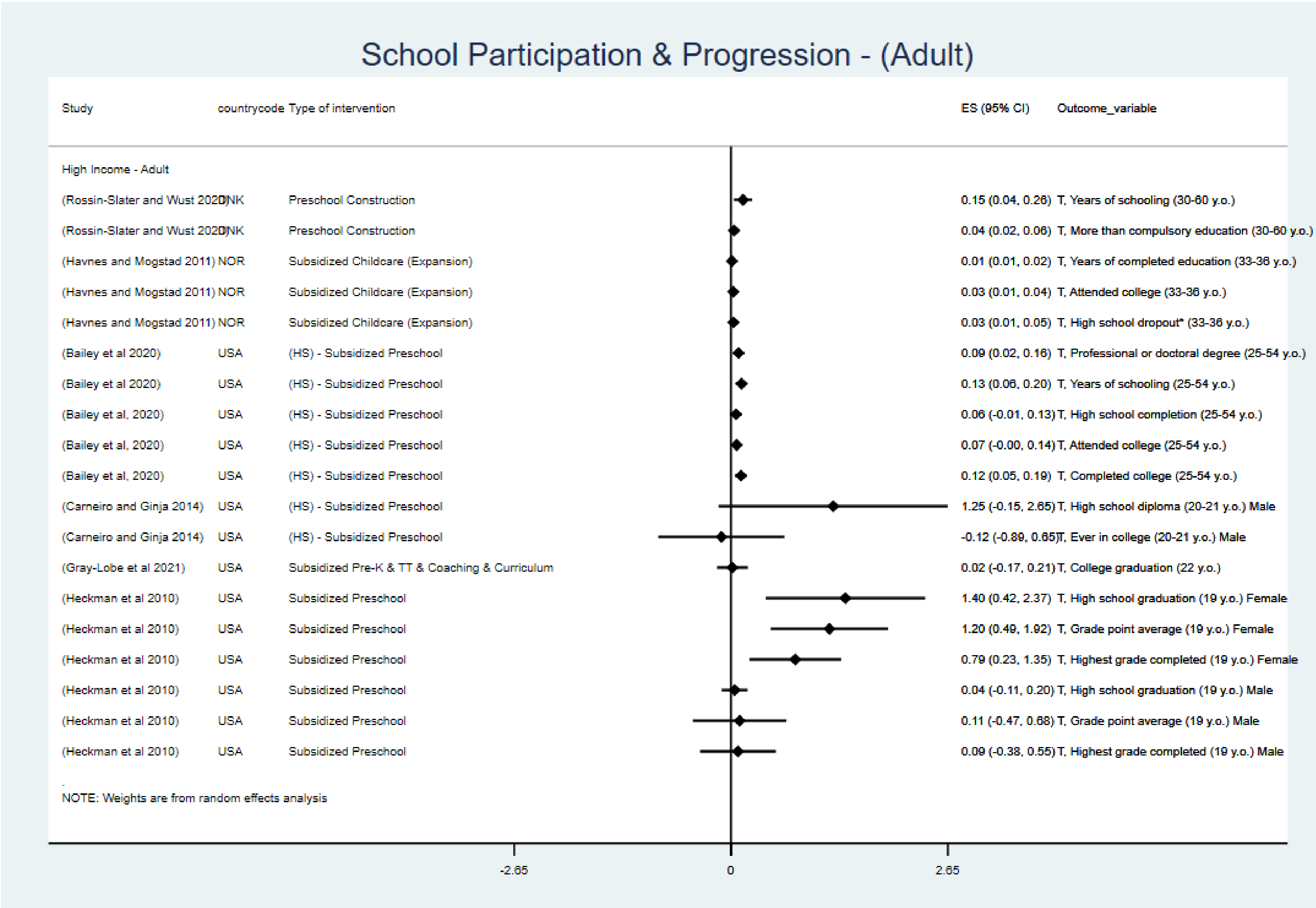
Appendix Figure 11: Effect sizes on behavioral skills (preprimary)



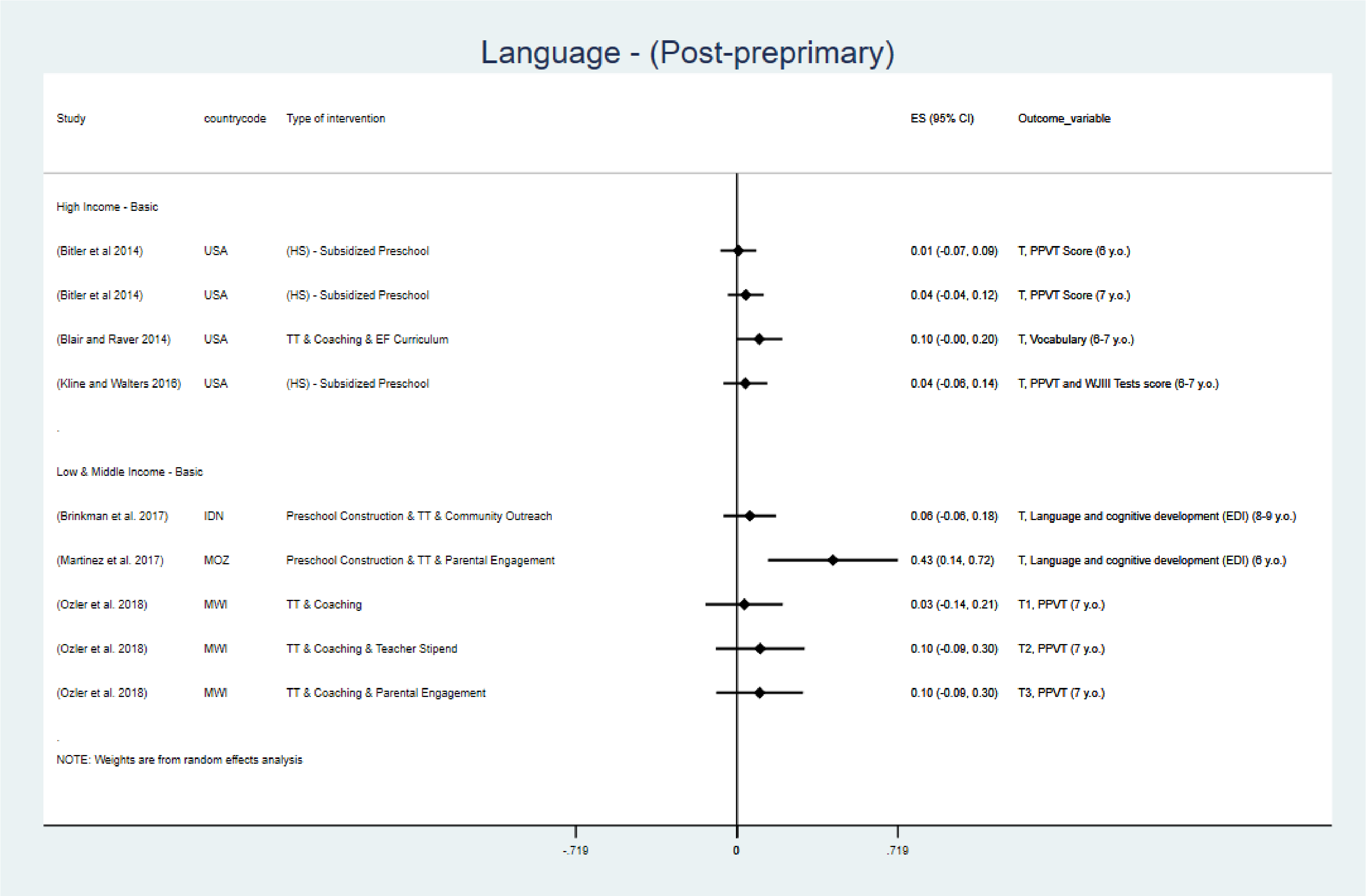
Appendix Figure 12: Effect sizes on school participation and progression (post-preprimary)



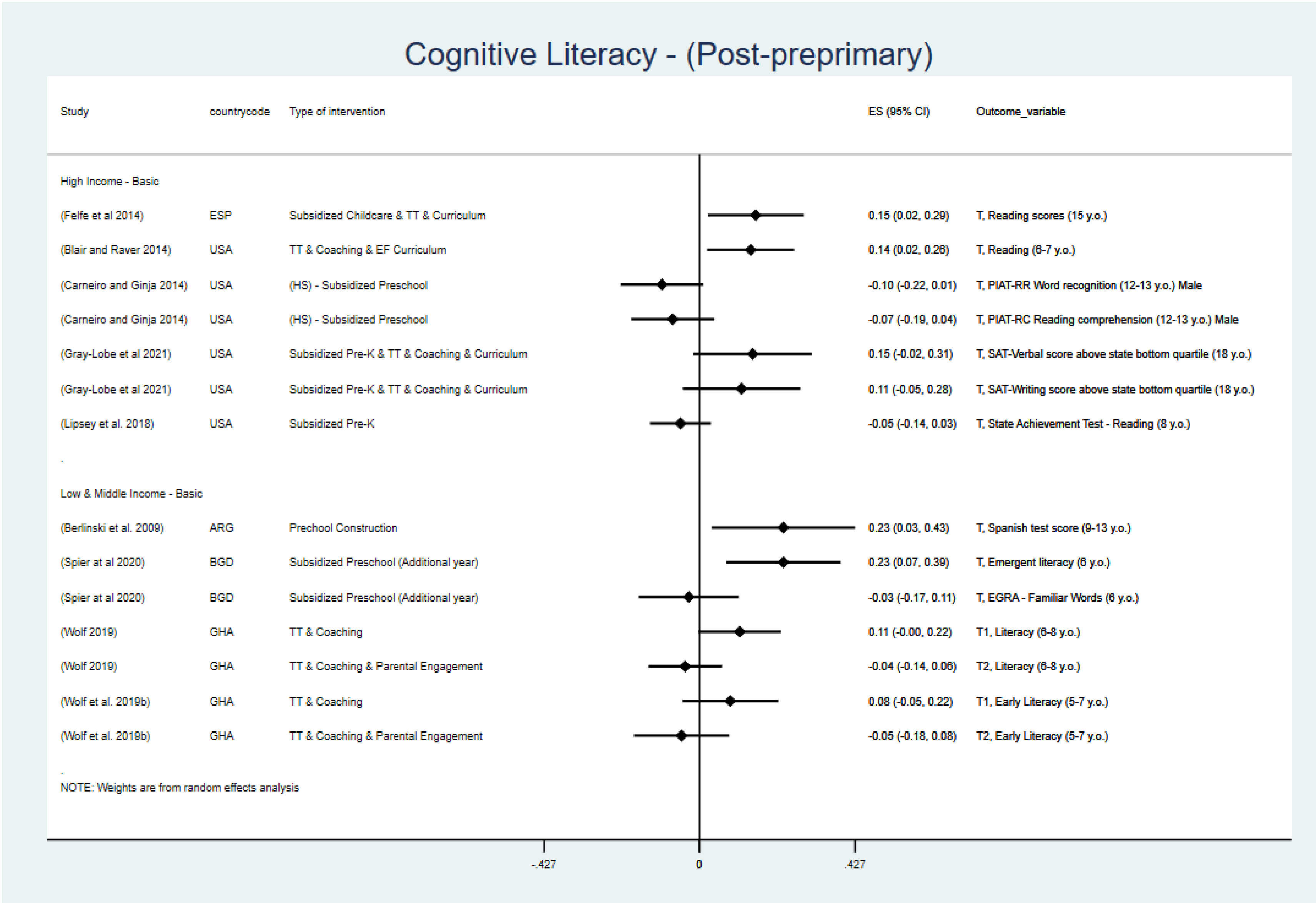
Appendix Figure 13: Effect sizes on school participation and progression (adulthood)



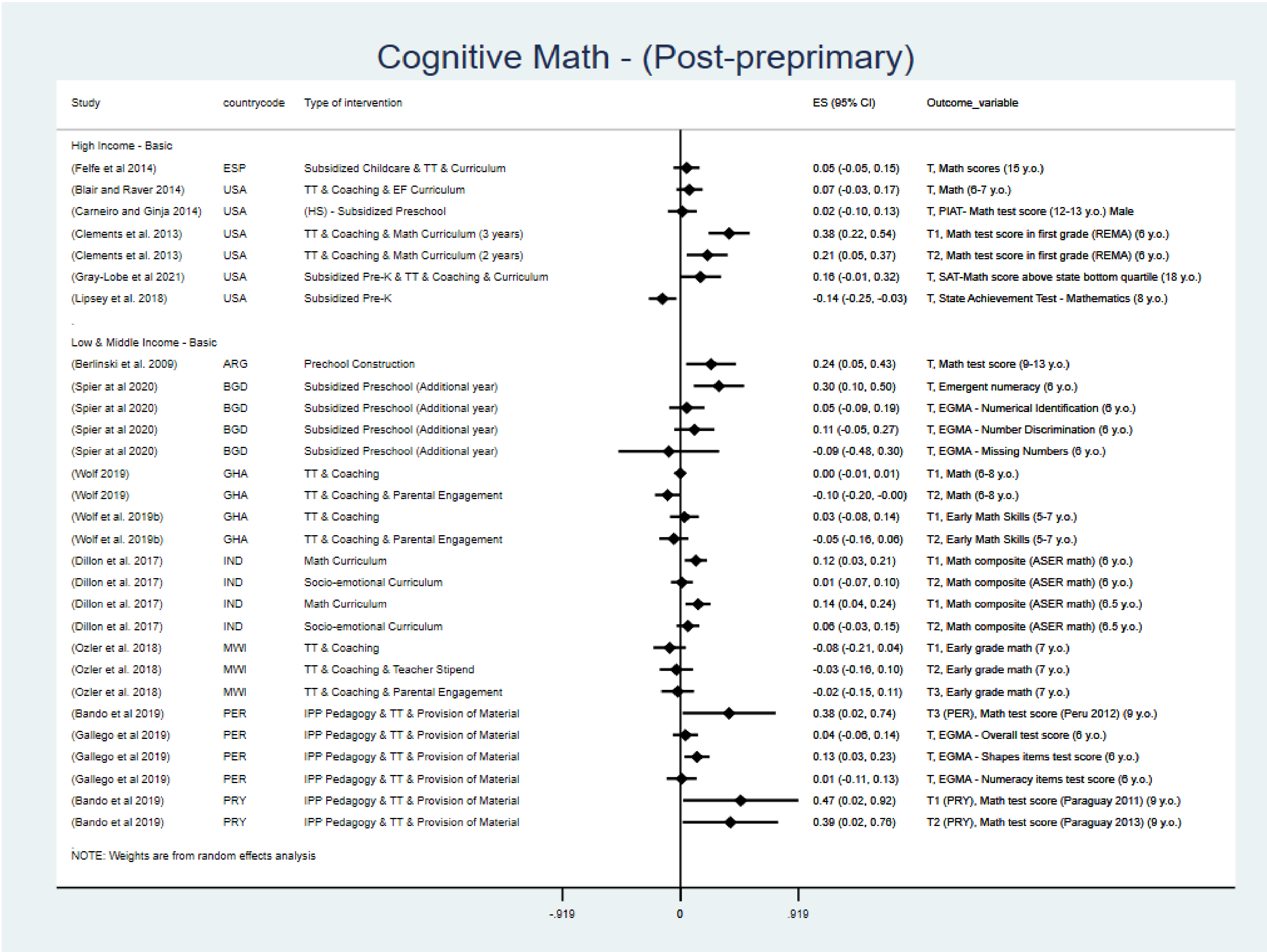
Appendix Figure 14: Effect sizes on language skills (post-preprimary)



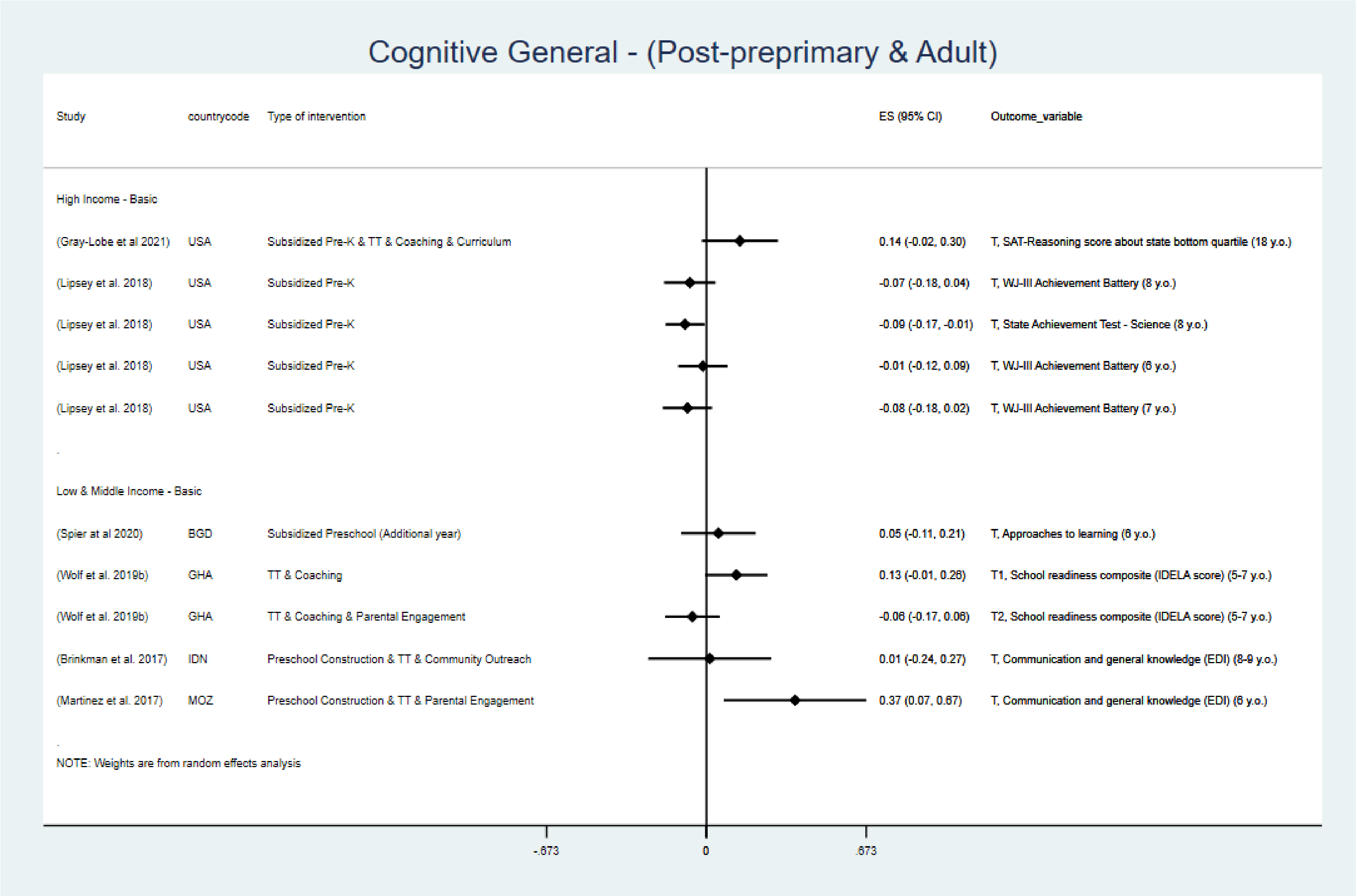
Appendix Figure 15: Effect sizes on cognitive literacy skills (post-preprimary)



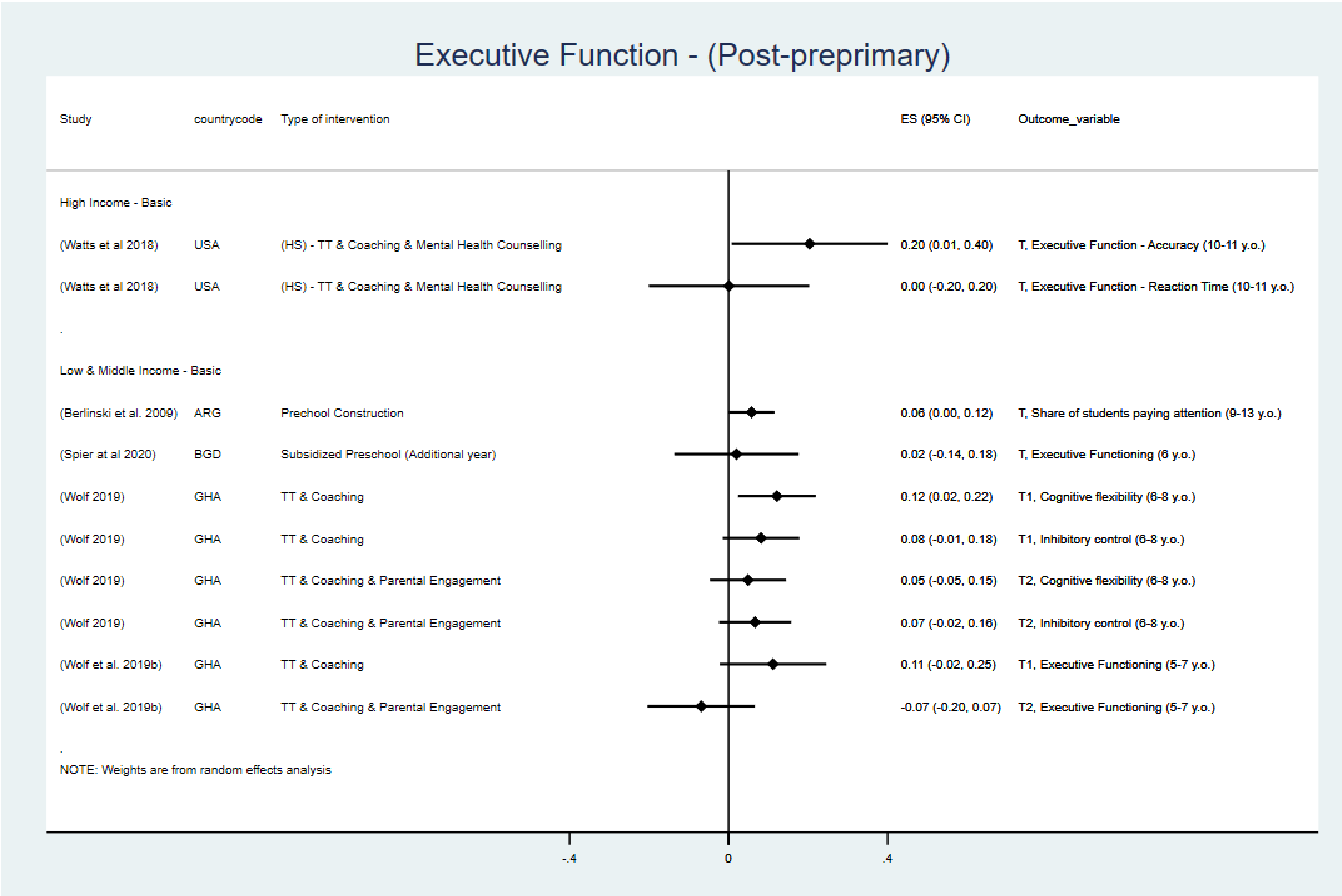
Appendix Figure 16: Effect sizes on cognitive math skills (post-preprimary)



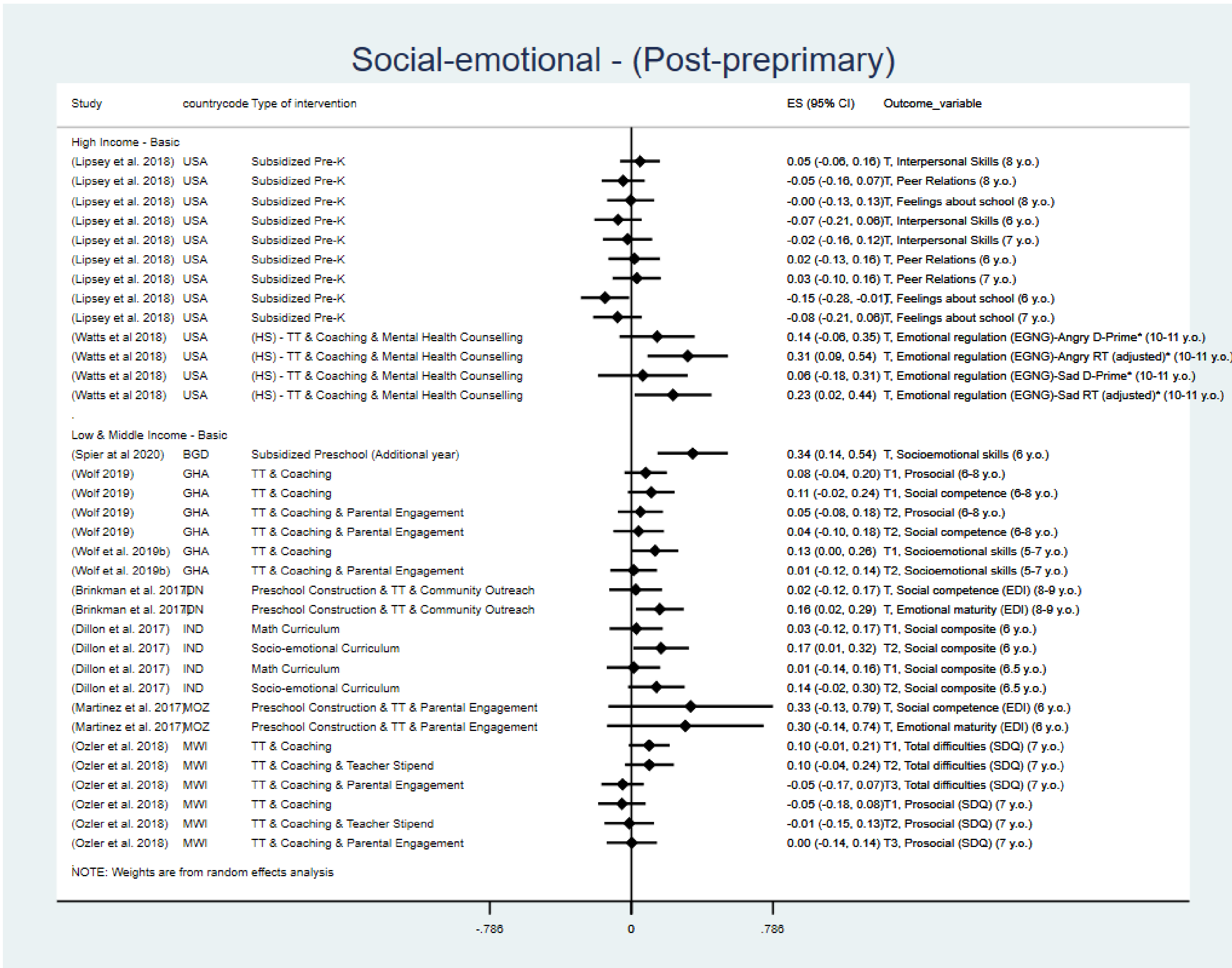
Appendix Figure 17: Effect sizes on cognitive general skills (post-preprimary)



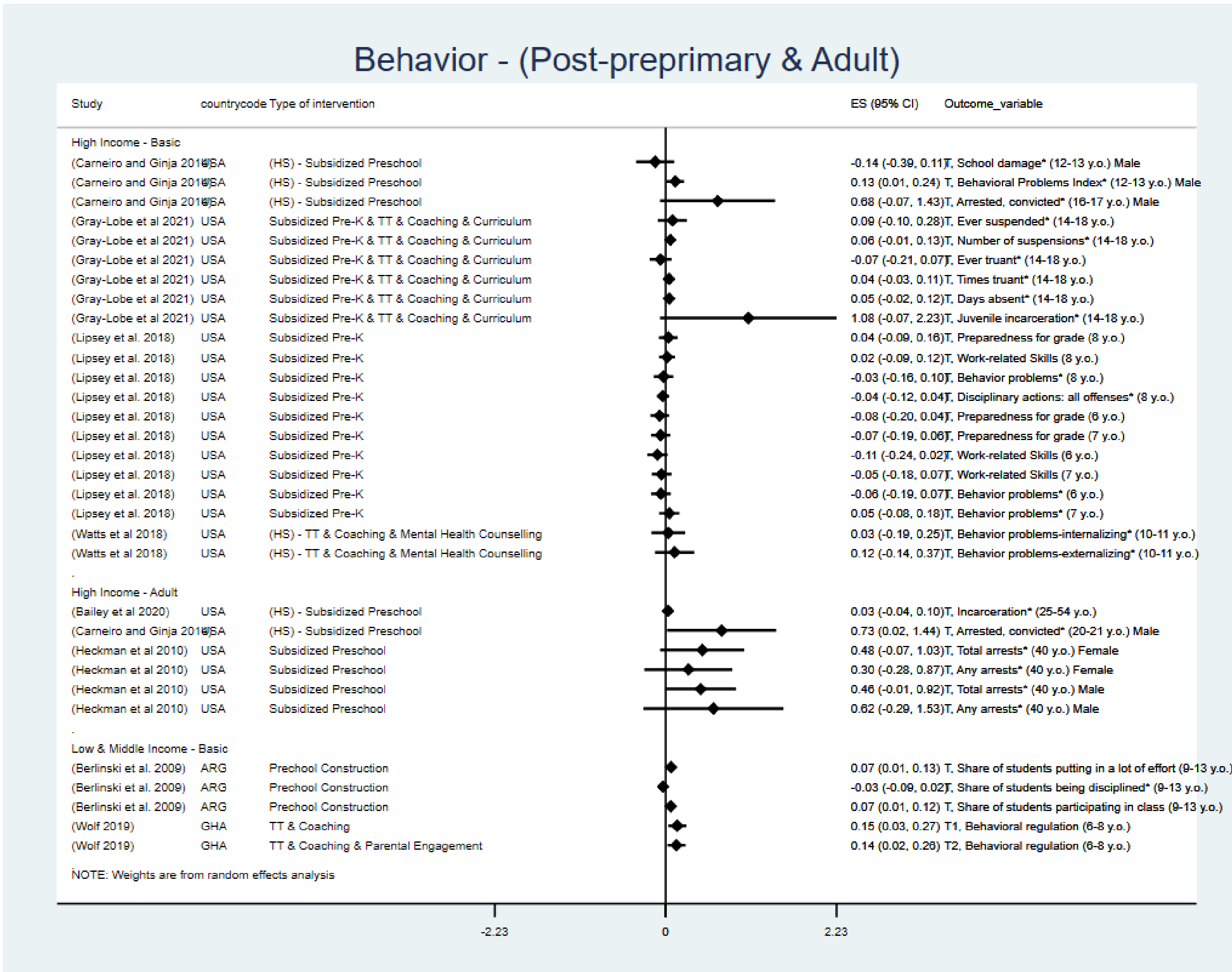
Appendix Figure 18: Effect sizes on executive function skills (post-preprimary)



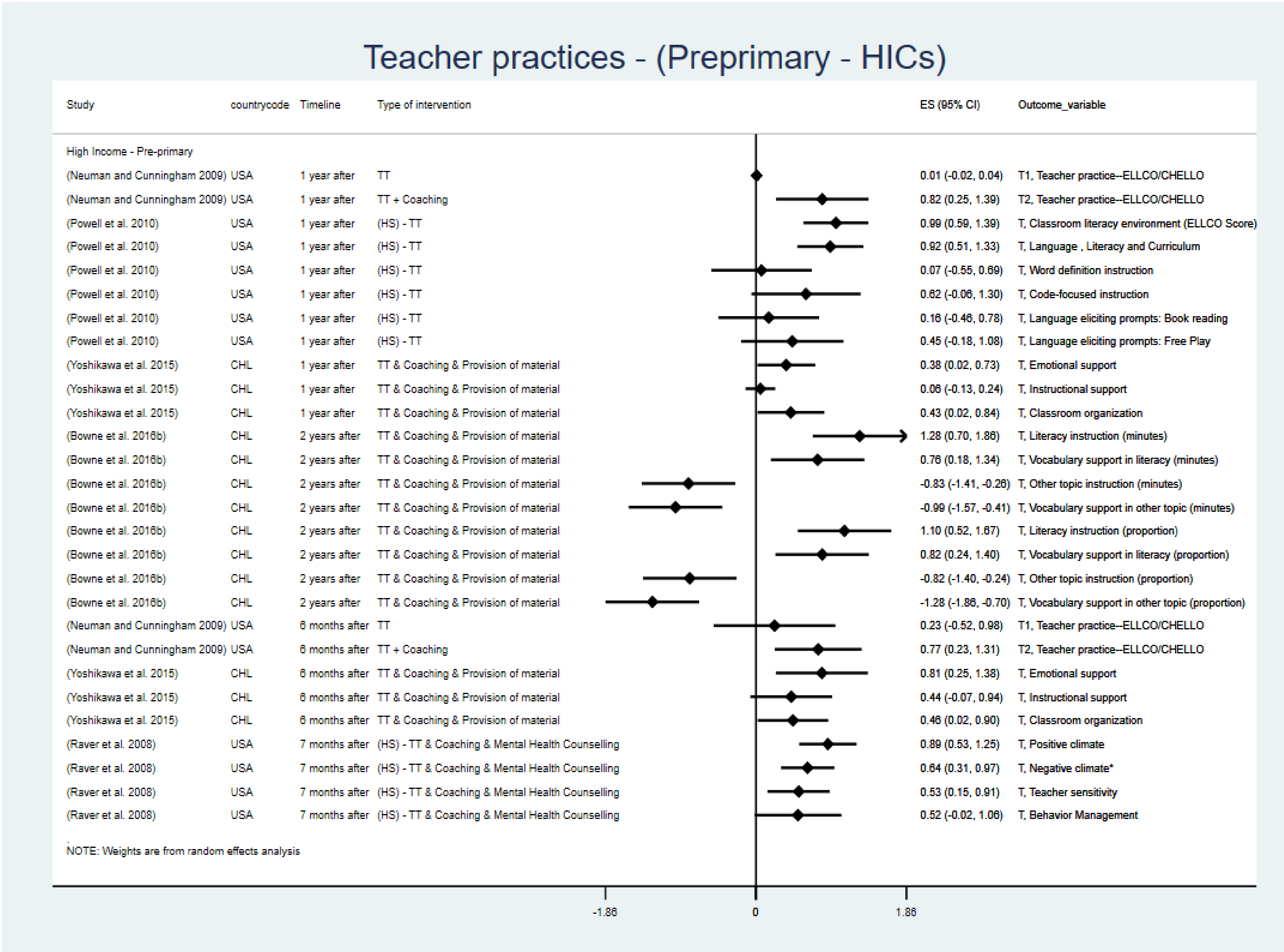
Appendix Figure 19: Effect sizes on social-emotional skills (post-preprimary)



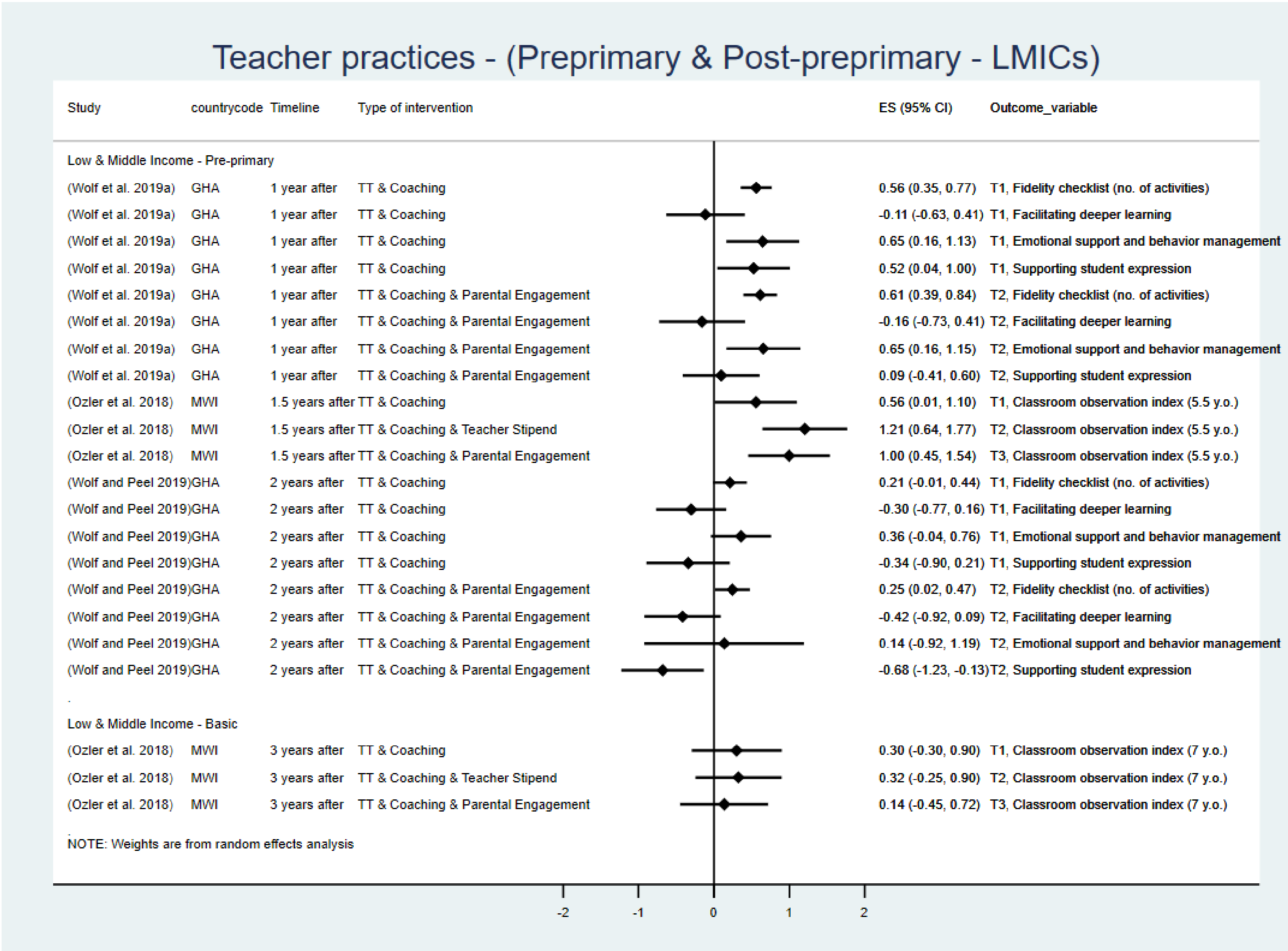
Appendix Figure 20: Effect sizes on behavioral skills (post-preprimary and adulthood)



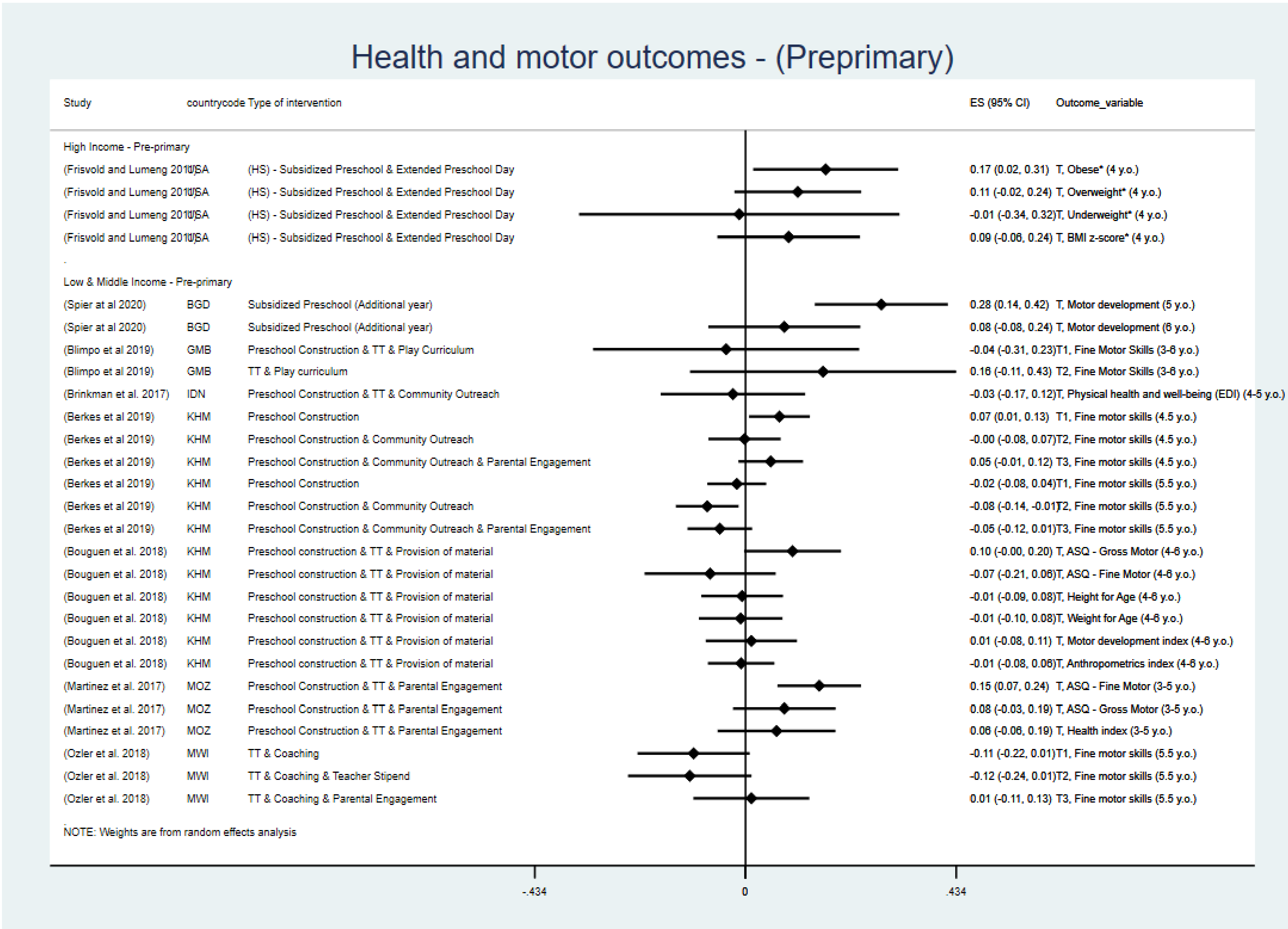
Appendix Figure 21: Effect sizes on teaching practices in HIC (preprimary)



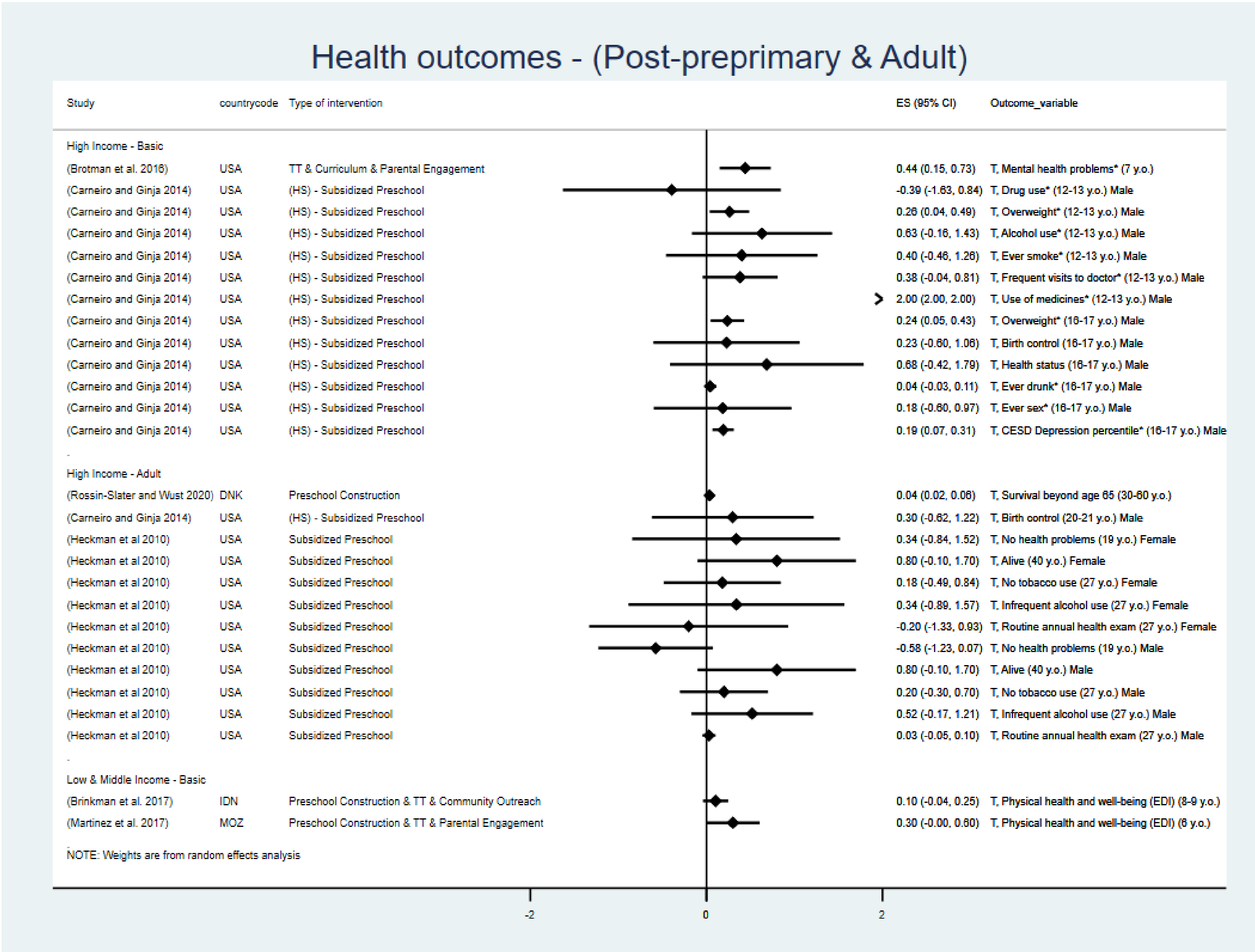
Appendix Figure 22: Effect sizes on teaching practices in LMICs (preprimary & post-preprimary)



Appendix Figure 23: Effect sizes on health and motor outcomes (preprimary)



Appendix Figure 24: Effect sizes on health and motor outcomes (post-preprimary and adulthood)



Appendix 5. Included studies

Andrew, A., O. Attanasio, R. Bernal, L.C. Sosa, S. Krutikova, and M. Rubio-Codina. 2019. *Preschool quality and child development* . (No. w26191): National Bureau of Economic Research.

Bailey, M.J., B.D. Timpe, and S. Sun. 2020. *Prep School for poor kids: The long-run impacts of Head Start on Human capital and economic self-sufficiency* . (No. w28268): National Bureau of Economic Research.

Baker, M., J. Gruber, and K. Milligan. 2008. "Universal child care, maternal labor supply, and family well-being." *Journal of Political Economy,* 116 (4): 709-745.

Bando, R., E. Näslund-Hadley, and P. Gertler. 2019. *Effect of inquiry and problem based pedagogy on learning: Evidence from 10 field experiments in four countries.* (No. w26280): National Bureau of Economic Research.

Barnett, W.S., K. Jung, D.J. Yarosz, J. Thomas, A. Hornbeck, R. Stechuk, and S. Burns. 2008. "Educational effects of the Tools of the Mind curriculum: A randomized trial." *Early Childhood Research Quarterly* 23 (3): 299-313.

Berkes, J., A. Bouguen, D. Filmer, and T. Fukao. 2019. *Improving Preschool Provision and Encouraging Demand: Heterogeneous Impacts of a Large-Scale Program.* Washington D.C.: The World Bank Working Paper Series.

Berlinski, S., S. Galiani, and M. Manacorda. 2008. "Giving children a better start: Preschool attendance and school-age profiles." *Journal of Public Economics* 92 (5-6): 1416-1440.

Berlinski, S., S. Galiani, and P. Gertler. 2009. "The effect of pre-primary education on primary school performance." *Journal of Public Economics* 93 (1-2): 219-234.

Bierman, K.L., C.E. Domitrovich, R.L. Nix, S.D. Gest, J.A. Welsh, M.T. Greenberg, C. Blair, K.E. Nelson, and S. Gill. 2008. "Promoting academic and social-emotional school readiness: The Head Start REDI program." *Child Development* 79 (6): 1802-181.

Bitler, M.P., H.W. Hoynes, and T. Domina. 2014. *Experimental evidence on distributional effects of Head Start.* (No. w20434): National Bureau of Economic Research.

Blair, C., and C.C. Raver. 2014. "Closing the achievement gap through modification of neurocognitive and neuroendocrine function: Results from a cluster randomized controlled trial of an innovative approach to the education of children in kindergarten." *PloS One* 9 (11): e112393.

Blimpo, M., P.M. Carneiro, P. Jervis, and T. Pugatch. 2019. *Improving access and quality in early childhood development programs: Experimental evidence from the Gambia.* Washington D.C.: World Bank Policy Research Working Paper, (8737).

Bloom, H.S., and C. Weiland. 2015. *Quantifying variation in Head Start effects on young children's cognitive and socio-emotional skills using data from the National Head Start Impact Study.* Available at SSRN 2594430.

Bouguen, A., D. Filmer, K. Macours, and S. Naudeau. 2018. "Preschool and parental response in a second best world evidence from a school construction experiment." *Journal of Human Resources* (The World Bank Working Paper Series) 53 (2): 474-512.

Bowne, J.B., H. Yoshikawa, and C.E. Snow. 2016. "Experimental impacts of a teacher professional development program in early childhood on explicit vocabulary instruction across the curriculum." *Early Childhood Research Quarterly* 34: 27-39.

Brinkman, S.A., A. Hasan, H. Jung, A. Kinnell, and M. Pradhan. 2017. "The impact of expanding access to early childhood education services in rural Indonesia." *Journal of Labor Economics* 35 (S1): S305-S335.

Brotman, L.M., S. Dawson-McClure, D. Kamboukos, K.Y. Huang, E.J. Calzada, K. Goldfeld, and E. Petkova. 2016. "Effects of ParentCorps in prekindergarten on child mental health and academic performance: Follow-up of a randomized clinical trial through 8 years of age." *JAMA Pediatrics* 170 (12): 1149-1155.

Carneiro, P., and R. Ginja. 2014. "Long-term impacts of compensatory preschool on health and behavior: Evidence from Head Start." *American Economic Journal: Economic Policy* 6 (4): 135-73.

Clements, D.H., J. Sarama, M.E. Spitler, A.A. Lange, and C.B. Wolfe. 2011. "Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial." *Journal for Research in Mathematics Education* 42 (2): 127-166.

Clements, D.H., J., Sarama, C.B. Wolfe, and M.E. Spitler. 2013. "Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies: Persistence of effects in the third year." *American Educational Research Journal* 50 (4): 812-850.

Diamond, A., W.S. Barnett, J. Thomas, and S. Munro. 2007. "Preschool program improves cognitive control." *Science* 318 (5855): 1387.

Dillon, M.R., H. Kannan, J.T. Dean, E.S. Spelke, and E. Duflo. 2017. "Cognitive science in the field: A preschool intervention durably enhances intuitive but not formal mathematics." *Science* 357 (6346): 47-55.

Drange, N., T. Havnes, and A.M. Sandsør. 2016. "Kindergarten for all: Long run effects of a universal intervention." *Economics of Education Review* 53: 164-181.

Felfe, C., N. Nollenberger, and N. Rodríguez-Planas. 2015. "Can’t buy mommy’s love? Universal childcare and children’s long-term cognitive development." *Journal of Population Economics* 28 (2): 393-422.

Frisvold, D.E., and J.C. Lumeng. 2011. "Expanding exposure can increasing the daily duration of head start reduce childhood obesity?" *Journal of Human Resources* 46 (2): 373-402.

Gallego, F.A., E. Näslund-Hadley, and M. Alfonso. 2021. "Changing Pedagogy to Improve Skills in Preschools: Experimental Evidence from Peru." *The World Bank Economic Review* 35 (1): 261-286.

Ganimian, A.J., K. Muralidharan, and C.R. Walters. 2021. *Augmenting State Capacity for Child Development: Experimental Evidence from India.* (No. w28780): National Bureau of Economic Research.

Gormley, W.T., D.A. Phillips, K. Newmark, K. Welti, and S. Adelstein. 2011. "Social-emotional effects of early childhood education programs in Tulsa." *Child Development* 82 (6): 2095-2109.

Gray-Lobe, G., P.A. Pathak, and C.R. Walters. 2021. *The Long-Term Effects of Universal Preschool in Boston.* (No. w28756): National Bureau of Economic Research.

Hamre, B. K., R. C. Pianta, A. J. Mashburn, and J. T. Downer. 2012. "Promoting young children's social competence through the preschool PATHS curriculum and MyTeachingPartner professional development resources." *Early Education & Development* 23 (6): 809-832.

Havnes, T., and M. Mogstad. 2015. "Is universal child care leveling the playing field?" *Journal of Public Economics* 127: 100-114.

Havnes, T., and M. Mogstad. 2011. "No child left behind: Subsidized child care and children's long-run outcomes." *American Economic Journal: Economic Policy* 3 (2): 97-129.

Heckman, J.J., S.H. Moon, R. Pinto, P.A. Savelyev, and A. and Yavitz. 2010. "The rate of return to the HighScope Perry Preschool Program." *Journal of Public Economics* 94 (1-2): 114-128.

Johnson, R.C., and C.K. Jackson. 2019. "Reducing inequality through dynamic complementarity: Evidence from Head Start and public school spending." *American Economic Journal: Economic Policy* 11 (4): 310-49.

Kline, P., and C.R. Walters. 2016. "Evaluating public programs with close substitutes: The case of Head Start." *The Quarterly Journal of Economics,* 131 (4): 1795-1848.

Lipsey, M.W., D.C. Farran, and K. Durkin. 2018. "Effects of the Tennessee Prekindergarten Program on children’s achievement and behavior through third grade." *Early Childhood Research Quarterly* 45: 155-176.

Martinez, S., S. Naudeau, and V. Pereira. 2017. *Preschool and child development under extreme poverty: evidence from a randomized experiment in rural Mozambique.* Washington D.C.: The World Bank.

Neuman, S.B., and L. Cunningham. 2009. "The impact of professional development and coaching on early language and literacy instructional practices." *American Educational Research Journal* 46 (2): 532-566.

Ozler, B., L. C. H. Fernald, P. Kariger, C. McConnell, M. Neuman, and E. Fraga. 2018. "Combining preschool teacher training with parenting education: A clusterrandomized controlled trial." *Journal of Development Economics* (The World Bank.) 133: 448–467.

Pianta, R., B. Hamre, J. Downer, M. Burchinal, A. Williford, J. Locasale-Crouch, C. Howes, K. La Paro, and C. Scott-Little. 2017. "Early childhood professional development: Coaching and coursework effects on indicators of children's school readiness." *Early Education and Development* 28 (8): 956-975.

Powell, D.R., K.E. Diamond, M.R. Burchinal, and M.J. Koehler. 2010. "Effects of an early literacy professional development intervention on head start teachers and children." *Journal of Educational Psychology* 102 (2): 299.

Raver, C.C., S.M. Jones, C. Li-Grining, F. Zhai, K. Bub, and E. Pressler. 2011. "CSRP's impact on low-income preschoolers' preacademic skills: self-regulation as a mediating mechanism." *Child Development* 82 (1): 362-378.

Raver, C.C., S.M. Jones, C. Li-Grining, F. Zhai, M.W. Metzger, and B. Solomon. 2009. "Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial." *Journal of Consulting and Clinical Psychology* 77 (2): 302.

Raver, C.C., S.M. Jones, C.P. Li-Grining, M. Metzger, K.M. Champion, and L. Sardin. 2008. "Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings." *Early Childhood Research Quarterly* 23 (1): 10-26.

Rossin-Slater, M., and M. Wüst. 2020. "What is the Added Value of Preschool for Poor Children? Long-Term and Intergenerational Impacts and Interactions with an Infant Health Intervention." *American Economic Journal: Applied Economics* 12 (3): 255-86.

Spier, E., K. Kamto, A. Molotsky, A. Rahman, N. Hossain, Z. Nahar, and H. and Khondker. 2020. "Bangladesh Early Years Preschool Program Impact Evaluation."

Watts, T.W., J. Gandhi, D.A. Ibrahim, M.D. Masucci, and C.C. Raver. 2018. "The Chicago School Readiness Project: Examining the long-term impacts of an early childhood intervention." *PloS One* 13 (7): e0200144.

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

Wolf, S. 2019. "Year 3 follow-up of the 'Quality Preschool for Ghana'interventions on child development." *Developmental Psychology* 55 (12): 2587.

Wolf, S., and M.E. Peele. 2019. "Examining sustained impacts of two teacher professional development programs on professional well-being and classroom practices." *Teaching and Teacher Education* 86: 102873.

Wolf, S., J.L. Aber, J.R. Behrman, and E. and Tsinigo. 2019a. "Experimental impacts of the "Quality Preschool for Ghana" interventions on teacher professional well-being, classroom quality, and children's school readiness." *Journal of Research on Educational Effectiveness* 12 (1): 10-37.

Wolf, S., J.L. Aber, J.R. Behrman, and M. Peele. 2019b. "Longitudinal causal impacts of preschool teacher training on Ghanaian children's school readiness: Evidence for persistence and fade-out." *Developmental Science* 22 (5): e12878.

Wong, V.C., T.D. Cook, W.S. Barnett, and K. Jung. 2008. "An effectiveness-based evaluation of five state pre-kindergarten programs." *The Journal of the Association for Public Policy Analysis and Management* 27 (1): 122-154.

Yoshikawa, H., D. Leyva, C.E. Snow, E. Treviño, M. Barata, C. Weiland, C.J. Gomez, et al. 2015. " Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and and child outcomes." *Developmental Psychology* 51 (3): 309.