Cross-Productivities of Executive Functions

By Maximiliaan W. P. Thijssen[a], Mari Rege[b], Ingeborg F. Solli[c], and Ingunn Størksen[d]

[a] Department of Economics and Finance
University of Stavanger Business School
University of Stavanger
maximiliaan.thijssen@uis.no

[b] Department of Economics and Finance
University of Stavanger Business School
University of Stavanger
mari.rege@uis.no

[c] Department of Economics and Finance
University of Stavanger Business School
University of Stavanger
ingeborg.f.solli@uis.no

[d] Norwegian Centre for Learning Environment and Behavioural Research
Faculty of Educational Sciences and Humanities
University of Stavanger
ingunn.storksen@uis.no

Abstract

Preschool programs can have lasting impacts on educational and labor market outcomes, but we have limited knowledge on what skills these programs should target. Executive function skills at school start are strong predictors of academic success, however, we do not know if preschool program-induced improvements in executive functions promote academic achievement in primary school. We combine experimental data with a skill-building model and find that preschool program-induced improvements in executive functions led to improvements in mathematical and language skills in primary school. This suggests important dynamic complementarities: Preschool investments in executive functions increase the effectiveness of primary school investments promoting mathematical and language skills.

Keywords. Executive functions, school readiness, self-productivity, cross-productivity, dynamic complementarity, randomized controlled trial.

JEL codes. I21, I24, J24, H75.

Maximiliaan Willem Pierre Thijssen https://orcid.org/0000-0001-8622-3106
Mari Rege https://orcid.org/0000-0002-9574-5851
Ingeborg Foldøy Solli https://orcid.org/0000-0002-9016-9920
Ingunn Størksen https://orcid.org/0000-0002-2258-1788

This article is based on the doctoral dissertation of Maximiliaan W. P. Thijssen: "*Human Capital Production in Childhood: Essays on the Economics of Education*." We received funding from Grant 270703 from the Norwegian Research Council. We thank Torberg Falch and participants in the Quantitative Forum at the University of Stavanger for helpful comments. Correspondence concerning this article should be addressed to Maximiliaan W. P. Thijssen, University of Stavanger, Business School, N-4036 Stavanger, Norway. Email: maximiliaan.thijssen@uis.no.

1. Introduction

Preschool programs can have positive and lasting effects on child development and labor market productivity (*e.g.*, Berlinski et al., 2008, 2009; Cornelissen et al., 2018; Felfe et al., 2015; Gray-Lobe et al., 2021; Havnes and Mogstad, 2011; Heckman et al., 2010, 2013). One challenge in designing these programs is knowing which skills to target. Language, mathematics and social-emotional skills are strong predictors of long term development (*e.g.*, Duncan et al., 2007; McClelland et al., 2007; Romano et al., 2010), but this does not necessarily mean that targeting these capacities in preschool will produce lasting effects. Some of these skills will soon be mastered anyway when taught in the early years of primary school (Bailey et al., 2020). A challenge for preschools is to foster skills that are central for children's capacity to navigate the transition to formal schooling (Troller-Renfree et al., 2022).

The transition from preschool to primary school often entails a substantial shift in pedagogical practices from a play-based to a more instruction-based approach. As a result, children face new demands on their ability to regulate behaviors, such as paying attention and following instructions (Blair and Diamond, 2008; DiPrete and Jennings, 2012). In addition, they have to form new friendships, often with less support from teachers scaffolding development of good relationships. Executive functioning skills (EF-skills) may help children handle these challenges and give them a better start academically and socially (Blair and Diamond, 2008). EF skills are the cognitive processes that control behavior and a substantial literature has demonstrated that these skills are strong predictors of academic and social development in school (*e.g.*, Blair, 2002; Blair and Razza, 2007).

While there is evidence that early childhood education programs can improve EF-skills (Diamond and Lee, 2011; Diamond and Ling, 2020), there is only limited evidence on whether program-induced improvements in EF skills lead to improvements in other skills. The child development literature (*e.g.*, Bierman et al., 2008; Raver et al., 2011) has identified EF skills as important mediators of intervention effects on school readiness. However, these studies measured EF and school readiness skills simultaneously, which sheds limited light on the cross-productivity of program-induced EF skills.[1]

We investigate cross-productivities of EF skills by combining high-quality experimental data from the Agder Project (Rege et al., 2021) with a skill-building model (Cunha and Heckman, 2007). The Agder Project is a nine-month-long intervention in Norway targeting 701 five-year-old preschool children. Through a randomized controlled trial, Norway's (relatively) unstructured pedagogical tradition was compared with a play-based, comprehensive and structured curriculum. Rege et al. (2021) report a sizable, positive treatment effect on EF skills post-intervention. We exploit the experimental design to investigate how these preschool program-induced improvements in EF skills led to improvements in mathematical and language skills in first year of primary school. Our investigation is possible because the Agder Project design uses the same skills assessments before and after treatment and at the one-year follow-up at the end of first grade. Our approach follows the decomposition framework in Heckman et al. (2013), which assumes that treatment-induced changes in measured and unmeasured skills are independent, as do Berger et al. (2020), Conti et al. (2016), and Kosse et al. (2020).

We find that preschool program-induced improvements in EF skills led to improvements in mathematical skills and language skills in primary school. This finding suggests a dynamic complementarity between the preschool curriculum (*i.e.*, an investment in preschool process quality) and the investments made in primary school (*e.g.*, teacher quality, instructional practices).

[1] McCoy et al. (2019) is an exception. They investigate the long-term effects of treatment-induced EF skills on self-reported high school performance.

Indeed, primary school seems to be more productive in promoting mathematical and language skills for children who have improved EF skills.

Our article contributes not only to the child development literature discussed above, but also to the literature concerning the economics of human development (see Heckman and Mosso, 2014, for a recent survey). This strand of literature regards preschool education as an investment in skill formation (Blau and Currie, 2006), and considers cross-productivity and dynamic complementarity to be among the salient features of skill formation models (Cunha and Heckman, 2007). Typically, researchers study either the effect of early childhood education programs (*e.g.*, Chor et al., 2016; Currie, 2001; Gormley Jr and Gayer, 2005; Jenkins et al., 2018) or cross-production of skills (*e.g.*, Cunha and Heckman, 2008; Cunha et al., 2010). Few studies investigate those relationships jointly, which is unfortunate because doing so can provide valuable insights into the mechanisms by which differences in skills subsequently emerge (or disappear) as a result of variations in preschool education.

Some notable exceptions include the following[2]: Heckman et al. (2013) investigate the mechanisms by which an early childhood education program, the HighScope Perry Preschool Program, affected a variety of education, labor market, crime, and health outcomes. They find that persistent changes in what they refer to as personality skills played a substantial role in the program's success. Second, Attanasio et al. (2020) investigate the mechanisms of an early childhood education program in Colombia. Their findings show that program-induced gains in cognitive and social-emotional skills are mainly the result of parents changing their investments. Additionally, they find that cognitive skills cross-produce social-emotional skills. While Heckman et al. (2013) and Attanasio et al. (2020) investigate the mechanisms of early childhood interventions, they do not focus on EF skills.

2. Background and Conceptual Framework

This section explains the concepts of self-productivity and dynamic complementarity and how those concepts form the basis of our hypothesis of cross-productivity. Before we explain our research hypothesis, let us briefly review what is meant by the term "executive functions" (or EF skills) and the context within which we test our hypothesis.

2.1. What are Executive Functions?

The term executive function (EF, or cognitive control) refers to a set of interrelated, top-down processes needed for concentration and thinking (Diamond and Lee, 2011). It is generally agreed that there are three core EF skills (Diamond, 2013): working memory, inhibitory control (which overlaps substantially with self-regulation), and cognitive flexibility. Working memory is key to knowing what to inhibit, while inhibition enables us to focus on specific content (Diamond, 2016), suggesting a reciprocal relationship. Cognitive flexibility, which requires working memory and inhibitory control, refers to the ability to view things from different perspectives and to think outside the box. The three core EF skills give rise to the higher-order EF skills of reasoning, problem-solving, and planning. Inhibition, working memory, and cognitive flexibility are central to learning, reasoning, problem-solving, and planning (Blair, 2002; Diamond and Lee, 2011), as well as to the regulation of attention, emotion, and behavior (Rueda et al., 2005). EF skills thus enable a child to "block" habitual behaviors and execute less familiar behaviors (Matsumoto and Tanaka, 2004).

[2] Campbell et al. (2014) and Conti et al. (2016) are also exceptions, but they focus primarily on health behaviors.

These EF skills— inhibitory control, working memory, and cognitive flexibility —are key predictors of social and economic success (see, *e.g.*, the review in Diamond and Ling, 2020). For example, EF skills correlate with academic performance (*e.g.*, Blair, 2002; Blair and Razza, 2007), social-emotional skills (*e.g.*, Broidy et al., 2003), criminal activity, (*e.g.*, Moffitt et al., 2011; Nagin and Tremblay, 1999), (risky) health behaviors (*e.g.*, Miller et al., 2011; Moffitt et al., 2011), and several other socioeconomic outcomes (see, *e.g.*, Diamond and Ling, 2020, pp. 157–160). So too, EF skills are thought to be foundational for school readiness (Blair, 2002; Blair and Razza, 2007). It is therefore important to investigate whether preschool education programs can improve EF skills and whether these program-induced improvements aid children to engage and benefit from the learning environment provided in primary school. This discussion concerning EF skills has been presented in narrative rather than formal form. A formal skill-building model can help clarify the cross-productivity of EF skills.

2.2. A Skill-Building Model

Let $\mathbf{S}_{i,t}$ denote a vector the elements of which represent skills, where t ($t = 0, 1, \dots, T$) indexes age over the T periods of childhood. We assume that these T periods cluster in $K \leq T$ stages of development ($k = 1, \dots, K$). Second, let $\mathbf{I}_{i,t}$ denote a vector representing investments. Child i ($i = 1, \dots, N$) develops skills when environmental influences (as experienced by the child) interact with skills previously acquired,

$$\mathbf{S}_{i,t+1} = \mathbf{f}_k(\mathbf{S}_{i,t}, \mathbf{I}_{i,t}). \tag{1}$$

If these environmental influences are enriched, then it is said that an "investment" is made.

One central assumption in the technology of skill formation is "self-productivity." Self-productivity encompasses the idea that skills are self-reinforcing and cross-producing. Self-reinforcement involves skills that are "alike," whereas cross-production involves skills that are "unlike." The following partial derivative formally defines self-productivity:

$$\frac{\partial \mathbf{S}_{i,t+1}}{\partial \mathbf{S}_{i,t}} > 0. \tag{2}$$

The assumed positive relationship between skills at age t and age $t + 1$ has two important implications. First, if skills are self-reinforcing, then investments will not fully depreciate over a given length of time (all else being equal). Second, if skills are cross-productive, then investments will produce synergistic effects (all else being equal).That is to say, boosting a cross-productive skill will also affect other skills, resulting in an effect greater than the effect that would have resulted from an investment in a skill that is *not* cross-productive. Having defined self-productivity and its implications, we next provide some background information relevant for understanding the context within which we formulate our research hypothesis of cross-productivity.

2.3. Background

We will investigate the cross-productivity of EF skills using high-quality experimental data from the Agder project implemented in Norwegian preschools. The welfare system in Norway includes generous social security and family policies. All one-to-five-year-old children are entitled to receive publicly regulated and subsidized preschool education and care. Preschool uptake amounts to about 98 percent among five-year-olds, the Agder Project's target population. Children start school the year they turn six.

The social pedagogical tradition characterizes preschool education in Norway. This tradition emphasizes free play and natural curiosity. As such, it contrasts with school readiness approaches commonly used in English-speaking countries (OECD, 2006). Research in psychology and education suggests that preschool curricula aimed at school readiness are more effective (Clements and Sarama, 2011; Dillon et al., 2017); consequently, returns on investments in terms of skill formation may be sub-optimal in Norway's preschool centers.

This situation motivated the Agder Project, which aimed to foster school readiness and human potential through playful learning in preschool (Rege et al., 2021). The intervention consisted of a comprehensive curriculum with various age-appropriate activities aimed at stimulating EF skills, social competence, mathematical skills, and language skills (Størksen et al., 2018).

Figure 1 shows the Agder Project's experimental design. In the preschool year 2015/2016, preschool teachers in the treatment group attended the credit-based university class and, as part of this training, provided extensive feedback on preschool activities, resulting in revisions of the curriculum. In August 2016, we assessed children's EF, mathematical, and language skills. This assessment is the baseline. The trained preschool teachers subsequently implemented the structured curriculum with the five-year-olds in the preschool center (in preschool year 2016/2017) and were offered coaching during the implementation phase. Centers in the control group continued per usual, according to the social pedagogical tradition of free play and natural curiosity. Immediately following the intervention in June 2017, we assessed the children for the second time (post-intervention). The follow-up assessment in primary school took place in March 2018, after the children had started formal schooling. We used the the same skills assessments before and after treatment and at the one-year follow-up. The preschool teachers in the control group participated in the credit-based university class and received the intervention materials in the preschool year 2017/2018, after the participating children had left preschool.

The main findings of the Agder Project can be summarized as follows (see Rege et al., 2021, for further details). First, the structured curriculum positively affected children's EF skills post-intervention. Second, at the follow-up assessment in primary school, children in the control group appear to have EF skills on par with children in the treatment group. Third, while the treatment effect on EF skills appears to have faded, children in the treatment group appear to have significantly better mathematical skills than children in the control group at the follow-up assessment in primary school.

2.4. Research Hypotheses

The reported evidence in Rege et al. (2021) suggests that the Agder Project's structured curriculum improved EF skills by the end of preschool. It may be that treated children who started primary school with better EF skills developed more mathematical skills, and that this contributed to the follow-up impact on mathematical skills. We are interested in testing this hypothesis. Specifically, we test:

Hypothesis.
Preschool program-induced improvements in EF skills lead to improvements in mathematical and language skills in primary school.

While our hypothesis concerns the cross-productivity of EF skills, support for it might be indicative of a dynamic complementarity. Assume that EF skills are cross-productive of other skills in primary school. In the context of this study, the higher levels of EF skills observed at the start of primary school can be assumed to have resulted from an investment in preschool. Investments made in primary school may complement EF skills in producing other skills. If so,

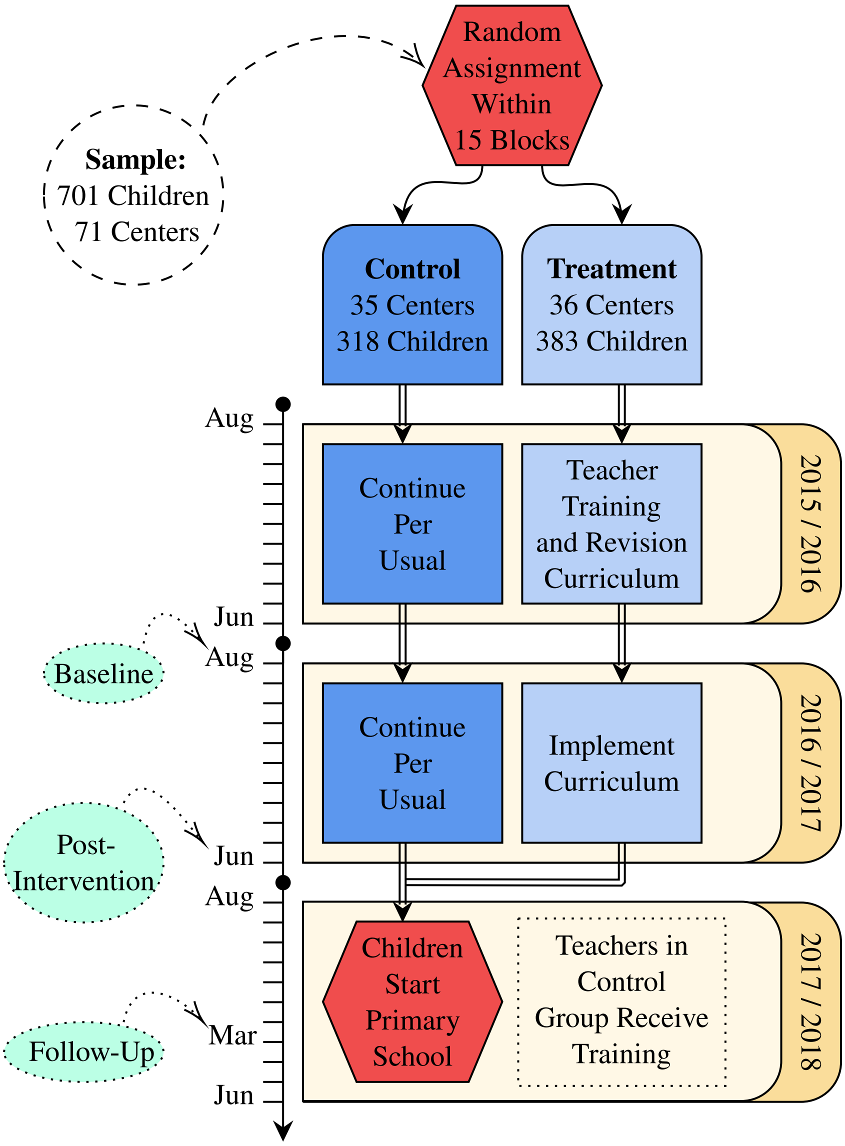

Figure 1. Experimental Design of the Agder Project

then a dynamic complementarity exists because current investments (*i.e.*, those made in primary school) are becoming more effective at producing other future skills, thanks to investments made in the past (*i.e.*, those made in preschool). The mechanism underlying such a dynamic complementarity would then be the more advanced EF skills at the start of primary school. Evidence on such dynamic complementarities can inform public policy and the design of preschool education programs to ensure that all children are ready to learn at school entry.

3. Data

During each assessment wave (*i.e.*, baseline 2016, post-intervention 2017, and the follow-up 2018), testers, who were trained, certified, and blind to treatment status, administered six tests: (1) the Ani Banani Math Test; (2) the Norwegian Vocabulary Test; (3) a Blending Test measuring phonological awareness; (4) the Hearts and Flowers Test; (5) the Head-Toes-Knees-Shoulders Test; and (6) the Forward and Backward Digit Span Test. The Ani Banani Math Test is designed to measure mathematical skills. Performance on the Norwegian Vocabulary Test and the Blending Test reflect language skills. Finally, performance on the Forward and Backward Digit Span Test, the Hearts and Flowers Test, and the Head-Toes-Knees-Shoulders Test provides an indication of the EF skills. We provide further details about each test below. We matched the Agder Project's assessment data to Statistics Norway's registry data, providing data on the child's sex, birth month, the parent's education, income, and whether at least one of the parents is a non-Western immigrant. In Online Appendix A.1, we show descriptives and use the baseline and predetermined variables to conduct balance tests. There has been relatively little attrition, and very few observations are missing (see Online Appendix A.1).

3.1. Mathematics

We used the Ani Banani Math Test to assess mathematical skills (Størksen and Mosvold, 2013), selecting 11 out of the 18 items. We dropped two items because almost all children answered them correctly. During the third assessment wave, technical problems with the tablet computer application caused another five items to become unusable. We omit these five items in each assessment wave to maintain consistency with Rege et al. (2021). During the Ani Banani Math Test, testers asked children to help the monkey Ani Banani with such tasks as counting bananas or setting a table with the correct number of plates. The test takes about ten minutes to complete. ten Braak and Størksen (2021) confirm (i) good concurrent validity, (ii) good discriminant validity when contrasted with measures of EF skills and language, and (iii) predictive validity for mathematics achievement five years later.

3.2. Language

We used the Norwegian Vocabulary Test (Størksen et al., 2013) and a Blending Test to assess early language skills. During the Norwegian Vocabulary Test, pictures appeared on the tablet computer, and the tester would then ask the child to identify the picture. Children received a point for each correct answer, with a total of 20 possible points.

The Blending Test measures phonological awareness. Testers presented a target word with its phonemes and asked the children to select the corresponding picture from four appearing on the tablet computer. Each correct response earned children one point, with a total of 12 possible points. While the Blending Test theoretically reflects language skills, it may be a weak test in practice, since it was originally designed for pedagogical rather than research purposes. The Blending Test simply assesses whether or not children can read, so the distribution of the sum of items is not normal. By contrast, measurement instruments such as the Head-Toes-Knees-Shoulders Test are designed and validated specifically for research purposes (McClelland et al., 2007, 2014).

3.3. Executive Functions

The first measure of EF skills is the Hearts and Flowers Test (Davidson et al., 2006), which assesses inhibitory control and cognitive flexibility using tablet computers. The children were instructed to press a key on the same side as the stimulus when they saw a heart and on the opposite side when the stimulus was a flower. The child received a point for each correctly pressed key. The test consisted of 60 stimuli.

The second measure is the Head-Toes-Knees-Shoulders Test (McClelland et al., 2014), which integrates inhibitory control, cognitive flexibility, and working memory demands into a self-regulation task. The test consists of three blocks with ten items per block. For each item, children received two points when they did the task correctly, one point when they carried out an incorrect movement but ended with a correct response, and zero points for incorrect responses. McClelland et al. (2014) report the psychometric properties of this test. The test relates to inhibition, working memory, and cognitive flexibility when it comes to construct validity. Furthermore, the test predicts academic achievement, particularly from kindergarten to first grade (Lenes et al., 2020).

The last measure is the Forward and Backward Digit Span Test, a component of the Wechsler Intelligence Scale for Children (Wechsler, 1991). The children were asked to listen to a sequence of digits voiced by the tester, then to repeat back the sequence of digits. The forward digit span test simply assesses short-term (auditory) memory, as children are not required to manipulate the information. By contrast, the backward digit span test measures the child's ability to manipulate verbal information in temporary storage (*i.e.*, working memory). The total score is the sum of the combined forward and backward digit span tests and reflects the total number of correctly repeated digit sequences.

4. Empirical Strategy

We first explain how differences between treated and non-treated children in primary school can be decomposed into preschool program-induced improvements. This decomposition closely follows Heckman et al. (2013). Since the six tests are measured with error, we next specify measurement models. We present further details concerning these measurement models (and their identification) in Online Appendix A.2. The last section describes our multi-step estimation procedure.

4.1. A Linear Framework for Decomposing Treatment Effects

We consider a linear-in-parameters production function. We assume that the first stage of development extends from August 2016 (age t) to June 2017 (age $t+1$) and that the second developmental stage extends from June 2017 (age $t+1$) to March 2018 (age $t+2$). It follows from Equation (1) that the parameters are invariant within these stages with respect to time.

Let d index treatment assignment so that $d = 1$ if a child attends a treated preschool center and $d = 0$ otherwise, $d \in \{0, 1\}$. Because of our linear-in-parameters assumption, we can write the first developmental stage,

$$\mathbf{S}_{d,i,t+1} = \mathbf{a}_d + \mathbf{B}\mathbf{S}_{d,i} + \mathbf{C}\mathbf{X}_{d,i} + \mathbf{w}_{d,i,t+1} \tag{3}$$

In Equation (3), $\mathbf{S}_{d,i,t+1}$ denotes a H-dimensional vector representing the counterfactual skill set at age $t+1$ (*i.e.*, post-intervention, June 2017). Second, $\mathbf{a}_d$ is a H-dimensional vector with scalar intercept parameters. Third, $\mathbf{B}$ is (H, H)-dimensional matrix of scalar parameters that characterize the extent to which skills acquired at age t (*i.e.*, baseline, August 2016) are self-productive. The parameters on the diagonal measure how skills reinforce themselves, and the off-diagonal parameters measure how skills cross-produce each other over the period from August 2016 to June 2017. Fourth, $\mathbf{C}$ is a (H, P)-dimensional matrix with scalar parameters measuring how child and parental characteristics plus randomization-block indicators affect skill formation. The last term, $\mathbf{w}_{d,i,t+1}$ is a H-dimensional vector representing idiosyncratic, zero mean shocks.

Since treatment assignment is random (Rege et al., 2021), we know, by definition, that observed (*i.e.*, $\mathbf{S}_{d,i,t}, \mathbf{X}_{d,i}$) and unobserved (*i.e.*, $\mathbf{w}_{d,i,t+1}$) variables balance in expectation. It follows, then, that the mean difference, $\mathbb{E}(\mathbf{S}_{1,i,t+1} - \mathbf{S}_{0,i,t+1}) = \mathbf{a}_1 - \mathbf{a}_0$, identifies program-induced improvements in skills.[3]

The second developmental stage is different. First, the program may have improved variables we did not measure. Second, the program may have changed the extent to which skills acquired at age t are self-productive (*i.e.*, $\mathbf{B} = \mathbf{B}_d$). Third, the program may have changed the extent to which child and parental characteristics affect skill formation (*i.e.*, $\mathbf{C} = \mathbf{C}_d$). Let $\mathbf{U}_{d,i,t+1}$ denote a R-dimensional vector with unmeasured variables affected by the structured curriculum and let $\tilde{\mathbf{B}}_d$ denote the (H, R)-dimensional matrix with scalar parameters that measure the effect of unmeasured variables on measured skills. For the second developmental stage, we can then write:

$$\mathbf{S}_{d,i,t+2} = \mathbf{a}_d + \mathbf{B}_d\mathbf{S}_{d,i,t+1} + \mathbf{C}_d\mathbf{X}_{d,i} + \tilde{\mathbf{B}}_d\mathbf{U}_{d,i,t+1} + \mathbf{w}_{d,i,t+2}. \tag{4}$$

To simplify the decomposition, we assume that $\mathbf{B}_1 = \mathbf{B}_0$ and $\mathbf{C}_1 = \mathbf{C}_0$. The treatment affected skills in primary school, but not self-productivity or the effect of child and parental characteristics. Parenthetically, we test and fail to reject this hypothesis. For details on the intuition of this test, see Heckman et al. (2013). Thus, we proceed by writing $\mathbf{B}_d = \mathbf{B}$ and $\mathbf{C}_d = \mathbf{C}$. We rewrite Equation (4) as follows:

$$\mathbf{S}_{d,i,t+2} = \tilde{\mathbf{a}}_d + \mathbf{B}\mathbf{S}_{d,i,t+1} + \mathbf{C}\mathbf{X}_{d,i} + \tilde{\mathbf{w}}_{d,i,t+2}, \tag{5}$$

where we defined $\tilde{\mathbf{a}}_d = \mathbf{a}_d + \tilde{\mathbf{B}}_d\mathbb{E}(\mathbf{U}_{d,i,t+1})$ as the new H-dimensional vector with intercepts and $\tilde{\mathbf{w}}_{d,i,t+2} = \mathbf{w}_{d,i,t+2} + \tilde{\mathbf{B}}_d[\mathbf{U}_{d,i,t+1} - \mathbb{E}(\mathbf{U}_{d,i,t+1})]$ as the new H-dimensional vector with random zero mean shocks.

If we further assume that program-induced increments in measured and unmeasured skills are (statistically) independent conditional on the child and parental characteristics for the no-treatment equations, then we can identify treatment effects owing to measured skills from the difference in expectations (for further details, see Heckman et al., 2013, pp. 2060–2063),

$$\overbrace{\mathbb{E}(\mathbf{S}_{1,i,t+2} - \mathbf{S}_{0,i,t+2})}^{\substack{\text{Follow-Up} \\ \text{Treatment Effect}}} = \overbrace{(\tilde{\mathbf{a}}_1 - \tilde{\mathbf{a}}_0)}^{\substack{\text{Contribution of} \\ \text{Unmeasured} \\ \text{Variables}}} + \overbrace{\mathbf{B}\mathbb{E}(\mathbf{S}_{1,i,t+1} - \mathbf{S}_{0,i,t+1})}^{\substack{\text{Contribution of} \\ \text{Measured Variables}}} \tag{6}$$

Thus, for the present decomposition, we maintain the same identifying assumption as Heckman et al. (2013), Berger et al. (2020), Conti et al. (2016), and Kosse et al. (2020) concerning the independence of measured and unmeasured skills.

4.2. Specifying a Linear Measurement Model

We specify a measurement model that addresses (classical) measurement error and weighs each test score based on its level of informativeness regarding the skill it manifests. Online Appendix A.2 presents results based on a simple arithmetic average.

Our measurement model defines observed test scores (*i.e.*, manifest variables) as a function of unobserved skills (*i.e.*, common factors) and other latent influences (*i.e.*, unique factors). Formally, let $\mathbf{M}_{h,d,i,t}$ denote a L_h-dimensional vector with (observed) manifest variables in which skill h manifests at age t. Since we observe the same manifest variables in each period,

[3] In Online Appendix A.1, we report the results of a balance test.

we omit a time subscript for L_h. We assume that the manifest variables are additively separable in the common factors they represent. It follows, then, that we can write the following linear system of measurement models:

$$\mathbf{M}_{h,d,i,t} = \boldsymbol{\mu}_{h,t} + \boldsymbol{\lambda}_{h,t} \mathrm{S}_{h,d,i,t} + \boldsymbol{\zeta}_{h,d,i,t}. \tag{7}$$

In Equation (7), $\boldsymbol{\mu}_{h,t}$ denotes a L_h-dimensional vector of intercepts. Furthermore, $\boldsymbol{\lambda}_{h,t}$ denotes a L_h-dimensional vector of factor loadings. These factor loadings weigh each manifest variable based on their informativeness concerning the common factor, $\mathrm{S}_{h,d,i,t}$. Lastly, $\boldsymbol{\zeta}_{h,d,i,t}$ is a L_h-dimensional vector with unique factors.

Since none of the right-hand-side variables in Equation (7) are observable, there is an inherent identification problem. First, we require some normalization to set a scale and location for the factors. We normalize one of the factor loadings (say the first) to one to set a scale. To set the location, we normalize the mean of the common factor to zero. Second, we assume (1) independence between the common and unique factors and (2) independence between the unique factors, conditional on the common factor. We also assume that the unique factors are independent across children. If the measures are continuous, then these assumptions and normalizations are sufficient for identifying the measurement model in Equation (7) (Anderson and Rubin, 1956).

We can identify the factor loadings from the ratio of covariances. With the factor loadings identified, we can (nonparametrically) identify the distribution of the factors by applying Kotlarski's lemma (see Lemma 1, Remark 4, and Remark 5 Kotlarski, 1967, pp. 70–73). However, these identification results no longer hold when the manifest variables are categorical (*e.g.*, ordinal, dichotomous). In those cases, we assume a known distribution for the unique factors. We also require further normalizations since the variances in polychoric (or tetrachoric) correlation matrices are redundant. We normalize the unique factor variances to unity to achieve (local) identification. Online Appendix A.2 provides further notes on identification.

A final consideration is the scale of the common factor. We anchor the scale of the common factor in the scale of one of the tests. We anchor mathematical skills in item 15 of the Ani Banani Math Test. Item 15 asks each child to copy a pattern appearing on the tablet computer. For EF skills, we use the Forward and Backward Digit Span Test. The score on this test represents the total number of correctly repeated number sequences. Lastly, we anchor language in the Blending Test. The score on this test is the total number of correctly chosen alternatives from four pictures.

4.3. Estimation Procedure

We use a multi-step estimation algorithm, following Heckman et al. (2013, p. 2066). In the first step, we estimate the measurement models for EF, mathematical, and language skills (see Online Appendix A.3). In the second step, we predict (Bartlett) factor scores (Bartlett, 1937; Thomson, 1938). The third step estimates the models outlined in Section 4.1 using the predicted factor scores. We apply Croon's correction method in the last step (Croon, 2002). The intuition behind Croon's correction method is to use our knowledge of the common factor variance and unique factor variance (from the first step) to adjust the estimates for prediction error. See Online Appendix A.3 for details. We assume data are missing at random, to address the issue of missing data, and estimate the models using full information parametric maximum likelihood (Anderson, 1957). See Online Appendix A.1 for details on missingness. We assume a normal distribution for the error terms. Note that we do not require this assumption for identification.

After applying this correction, the standard errors are incorrect. We apply a clustered wild residual bootstrap procedure. We draw 1,000 bootstrap samples from the original data and apply the multi-step estimation algorithm to each pseudo-sample. We cluster and re-sample at the (randomization) block level to ensure that each bootstrap sample includes both treated and non-treated children. Lastly, we use the "true" variance, not the variance of the predicted factor score, which is biased, to standardize the interpretation of the parameters.

5. Results

We first present the parameter estimates of self-reinforcement and cross-production for both stages of development. Then, we discuss the post-intervention and follow-up treatment effects. Lastly, we discuss how the follow-up treatment effect results from program-induced improvements in EF, mathematical, language, and unmeasured skills post-intervention.

Table 1 reports the self-reinforcement and cross-production parameter estimates and standard errors. The results suggest the following: First, EF and mathematical skills show strong persistence in both stages of development (the diagonal cells). In our preferred model specification (*i.e.*, the model with control variables), we estimate an auto-regressive parameter of 0.774 for EF skills and 0.663 for mathematical skills (both statistically significant at the one percent level) in developmental stage 1. In developmental stage 2, we find a similar level of persistence for mathematical skills, namely 0.635, but the persistence of EF skills increases to 0.943 (both statistically significant at the one percent level). High persistence implies that (effective) investments depreciate more slowly. Second, we observe that skills are cross-productive in the first development stage, particularly EF and mathematical skills. In the second developmental stage, we observe that EF skills remain cross-productive. By contrast, mathematical skills promote only EF skills, and language skills do not seem to boost either EF or mathematical skills. These findings align with the studies that have determined that EF skills predict success in school (Blair, 2002, 2006; Blair and Raver, 2015; Blair and Razza, 2007).

Table 2, Panel A, reports the parameter estimates and standard errors associated with the post-intervention and follow-up treatment effects. The results in Panel A suggest the following: Consistent with Rege et al. (2021), children develop more EF skills because they have experienced the structured curriculum. Treated children have about a 0.176 standard-deviation-higher level of EF skills than non-treated children (statistically significant at the one percent level). Furthermore, there is suggestive evidence that the structured curriculum boosted the mathematical skills of treated children by the end of preschool. However, we cannot rule out the possibility that this estimate occurred by chance at the conventional cut-off of five percent. Panel A further shows that language skills and EF skills differ in favor of treated children in primary school, but these differences are imprecisely estimated, so we cannot assume a difference at the conventional cut-off of five percent. We do observe positive and statistically significant differences in favor of treated children for mathematical skills in primary school, however. The follow-up treatment effect on mathematical skills implies that treated children have about a 0.198 standard deviation higher level of mathematical skill because they experienced the structured curriculum in preschool (statistically significant at the five percent level). Online Appendix A.3 presents results that do not account for measurement error (except through simple averaging). We find point estimates (and levels of precision) of the post-intervention and follow-up treatment effects even closer to those reported in Rege et al. (2021).

Table 2, Panel B, decomposes the follow-up treatment effect into measured and unmeasured variables. In brackets, we report the relative contributions of these variables. We calculate these relative contributions using absolute values. It may appear counter-intuitive to decompose the follow-up treatment effect on EF skills and language skills, since there is no significant treatment impact on these measures. Note, however, that the follow-up treatment effect captures total differences between treated and nontreated children. There could be unmeasured variables that

Table 1. Self-Reinforcement and Cross-Production of Skills

	Mathematics		Language		EFs	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Developmental Stage 1 (2016 - 2017)						
EFs	0.393	0.363	0.148	0.167	0.785	0.774
	(0.061)	(0.052)	(0.061)	(0.065)	(0.056)	(0.045)
Mathematics	0.691	0.663	0.292	0.293	0.348	0.339
	(0.034)	(0.039)	(0.059)	(0.065)	(0.056)	(0.060)
Language	0.074	0.090	0.279	0.264	0.115	0.113
	(0.040)	(0.042)	(0.059)	(0.061)	(0.044)	(0.046)
Panel B: Developmental Stage 2 (2017 - 2018)						
EFs	0.390	0.424	0.355	0.304	0.966	0.943
	(0.054)	(0.052)	(0.059)	(0.057)	(0.066)	(0.065)
Mathematics	0.668	0.635	0.049	0.081	0.247	0.261
	(0.052)	(0.057)	(0.070)	(0.065)	(0.052)	(0.051)
Language	0.021	0.043	0.243	0.215	−0.018	−0.029
	(0.040)	(0.043)	(0.055)	(0.050)	(0.045)	(0.041)
Control Variables	No	Yes	No	Yes	No	Yes

Notes. This table reports the self-productivity parameter estimates for the first developmental stage (Panel A) and the second (Panel B). The columns denote the dependent variables, and the rows denote the independent variables. Consider the first row (i.e., EF skills) in Panel A; the reported results measure the effect of EF skills measured at baseline (2016) on mathematical skills (Column 1), language skills (Column 3), and EFs (Column 5) measured post-intervention (2017). Columns (2), (4), and (6) differ from Columns (1), (3), and (5). Consequently, the table reports estimates from twelve models. We include the following control variables: child sex, birth month, whether or not at least one of the parents is a non-Western immigrant, parental education, family income, an indicator for late parental consent, and randomization-block indicators. All of the reported parameter estimates are given in standard deviations. We standardize the parameter estimates using the "true" variance, not the predicted variance. We present randomization-block clustered standard errors in parentheses. Standard errors are computed using a wild residual bootstrap procedure (1,000 bootstrap samples). All models are estimated with full-information maximum likelihood (total observations 701). Online Appendix A.1 provides an overview of missingness.

favor non-treated children, thereby reducing the estimate of the total difference. Alternatively, the total difference might be noisier (i.e., less precisely estimated) because its standard error is based on a linear combination of parameter estimates, which is generally more imprecise than of any single parameter estimate. Consequently, while the follow-up treatment effect on EF and language skills are not statistically significant, it is still worthwhile to estimate the extent to which preschool program-induced changes in EF skills contribute to the follow-up differences in language and EF skills.

The findings in Columns (1) through (4) are consistent with our hypothesis. We find that preschool program-induced improvements in EF skills led to improvements in mathematical and language skills in primary school. Specifically, because treated children started primary school with higher EF skills, treated children developed about 0.050 of a standard deviation more mathematical skills and about 0.037 of a standard deviation more language skills (both statistically significant at the five percent level). While the total follow-up treatment effect in language skills was not statistically significant, the extent to which preschool program-induced changes in EF skills contributed to these follow-up differences is. Taken together, these findings illustrate the fundamental role of EF skills in learning, which is consistent with the claim that these skills are beneficial for school success (Diamond and Lee, 2011). In other words, children seem to develop more mathematical and language skills because they started primary school with improved EF skills.

Based on the follow-up treatment effect of EF skills (Columns 5 and 6), it appears that the impact of the structured curriculum on these skills fades out, as differences between treated and non-treated children are (statistically) indistinguishable. However, Panel B reveals that such a conclusion would be inaccurate. Children who started primary school with superior EF skills – because they experienced the structured curriculum – showed improvement in their EF skills. Treated children developed 0.105 of a standard deviation more EF skills in primary school because they started school with better EF skills (statistically significant at the five percent level). This finding is consistent with the high auto-regressive parameter estimates reported in

Table 2. Treatment Effects and Decomposition

	Mathematics		Language		EFs	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Post-Intervention and Follow-Up Treatment Effects						
Post-Intervention	0.183	0.173	0.026	0.042	0.168	0.176
	(0.099)	(0.098)	(0.098)	(0.104)	(0.063)	(0.063)
Follow-Up	0.205	0.198	0.006	0.042	0.050	0.069
	(0.100)	(0.093)	(0.111)	(0.112)	(0.088)	(0.087)
Panel B: Decomposition of Follow-Up Treatment Effect						
EFs	0.043	0.050	0.041	0.037	0.099	0.105
	(0.020)	(0.021)	(0.018)	(0.015)	(0.049)	(0.047)
	[21.1%]	[25.3%]	[48.2%]	[78.7%]	[57.9%]	[64.4%]
Mathematics	0.031	0.028	0.002	0.004	0.012	0.011
	(0.018)	(0.017)	(0.002)	(0.002)	(0.001)	(0.007)
	[15.2%]	[14.1%]	[2.4%]	[8.5%]	[7.0%]	[6.7%]
Language	0.000	0.001	0.002	0.004	−0.000	−0.000
	(0.001)	(0.002)	(0.006)	(0.009)	(0.001)	(0.002)
	[0.0%]	[0.5%]	[2.4%]	[8.5%]	[0.0%]	[0.0%]
Unmeasured Variables	0.130	0.119	−0.040	−0.002	−0.060	−0.047
	(0.096)	(0.089)	(0.109)	(0.110)	(0.077)	(0.076)
	[63.7%]	[60.1%]	[47.1%]	[4.3%]	[35.1%]	[28.8%]
Control Variables	No	Yes	No	Yes	No	Yes

Notes. This table reports the parameter estimates for the treatment effect decomposition. Panel A reports the post-intervention treatment effect (*i.e.*, differences between treated and non-treated children post-intervention) and the follow-up treatment effect (*i.e.*, differences between treated and non-treated children at the follow-up). Panel B decomposes the follow-up treatment effect in measured and unmeasured variables. The columns denote the dependent variables, and the rows denote the independent variables. Consider the first row (*i.e.*, post-intervention treatment effect) in Panel A; the reported results measure the effect of the structured curriculum on mathematical skills (Column 1), language skills (Column 3), and EF skills (Column 5) measured post-intervention (2017). The table reports estimates from twelve models. We include the following control variables: child sex, birth month, whether or not at least one of the parents is a non-Western immigrant, parental education, family income, an indicator for late parental consent, and randomization-block indicators. All of the reported parameter estimates are in standard deviations. We standardize the parameter estimates using the "true" variance, not the predicted variance. We present randomization-block clustered standard errors in parentheses. Standard errors are computed using a wild residual bootstrap procedure (1,000 bootstrap samples). In brackets, we report the relative contributions. We compute these relative contributions by taking the absolute value of the estimate divided by the sum of absolute values of all estimates multiplied by 100. All models are estimated with full-information maximum likelihood (total observations 701). Online Appendix A.1 provides an overview of missingness.

Table 1 and the stable rank order of children this implies. Also, this finding is consistent with an implication of self-productivity; if skills are self-productive, then investments will not fully depreciate over a given length of time (all else being equal).

The observation that the follow-up treatment effect in Table 2 is smaller than the effect of program-induced changes in EF skills on those skills indicates that non-treated children catch up. The unmeasured variables capture this catch-up mechanism. We do not know precisely what these unmeasured variables are, however. These unmeasured variables may include unmeasured skills, but they might also capture interactions between measured skills (which would vary by treatment assignment) and investments made in primary school (which would not vary by treatment assignment). These interactions may provide a possible explanation for the catch-up. In particular, it could be that investments in primary school (*e.g.*, the structure provided by the teacher through rules and expectations) are compensatory to low levels of EF skills, as is commonly hypothesized (see, *e.g.*, Bierman et al., 2008; Raver et al., 2011; Riggs et al., 2006). These investments may have boosted EF skills for non-treated children, who had lower levels of these skills at the start of primary school, but not the EF skills of treated children, who had already mastered them by the start of primary school. While plausible, such catch-up does not imply that the intervention was not effective. The primary school environment may substitute for lower levels of EF skills, but Table 2 showed that treated children who started school with higher levels of EF skills developed more mathematical and language skills.

6. Conclusion

We investigated the cross-productivity of EF skills. We combined the experimental variation with an econometric model of skill formation to estimate the extent to which program-induced improvements in EF skills caused children to acquire more mathematical and language skills. We found that children did in fact acquire more mathematical and language skills in primary school because they started primary school with higher levels of EF skills. These findings hint at a dynamic complementarity between preschool investments in EF skills and learning in primary school.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika 52*(3), 317–332.

Anderson, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association 52*(278), 200–203.

Anderson, T. W. and H. Rubin (1956). *Statistical Inference in Factor Analysis*, Volume 5 of *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Contributions to Econometrics, Industrial Research, and Psychometry*. Berkeley, California: University of California Press.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020). Estimating the production function for human capital: Results from a randomized controlled trial in Colombia. *American Economic Review 110*(1), 48–85.

Bailey, D. H., G. J. Duncan, F. Cunha, B. R. Foorman, and D. S. Yeager (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest 21*(2), 55–97.

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology 28*(1), 97–104.

Berger, E. M., E. Fehr, H. Hermes, D. Schunk, and K. Winkel (2020). The impact of working memory training on children's cognitive and noncognitive skills. Working Paper 13338, Institute of Labor Economics, Bonn, Germany.

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.

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.

Bierman, K. L., R. L. Nix, M. T. Greenberg, C. Blair, and C. E. Domitrovich (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the head start redi program. *Development and Psychopathology 20*(3), 821–843.

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist 57*(2), 111–127.

Blair, C. (2006). How similar are fluid cognition and general intelligence? a developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences 29*(2), 109–160.

Blair, C. and A. Diamond (2008). Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. *Development and Psychopathology 20*(3), 899—911.

Blair, C. and C. C. Raver (2015). School readiness and self-regulation: A developmental psychobiological approach. *Annual Review of Psychology 66*, 711–731.

Blair, C. and R. P. Razza (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development 78*(2), 647–663.

Blau, D. and J. Currie (2006). Pre-school, day care, and after-school care: Who's minding the kids? In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Volume 2, Chapter 20, pp. 1163–1278. Amsterdam, Netherlands: Elsevier B.V.

Broidy, L. M., D. S. Nagin, R. E. Tremblay, J. E. Bates, B. Brame, K. A. Dodge, D. Fergusson, J. L. Horwood, R. Loeber, R. Laird, D. R. Lynam, T. E. Moffitt, G. S. Pettit, and F. Vitaro (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology 39*(2), 222–245.

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.

Chor, E., M. E. Andresen, and A. Kalil (2016). The impact of universal prekindergarten on family behavior and child outcomes. *Economics of Education Review 55*, 168–181.

Clarke, D., J. P. Romano, and M. Wolf (2020). The romano-wolf multiple-hypothesis correction in stata. *Stata Journal 20*(4), 812–843.

Clements, D. H. and J. Sarama (2011). Early childhood mathematics intervention. *Science 333*(6045), 968–970.

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

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2018). Who benefits from universal child care? estimating marginal returns to early child care attendance. *Journal of Political Economy 126*(6), 2356–2409.

Croon, M. (2002). Using predicted latent scores in general latent structure models. In G. A. Marcoulides and I. Moustaki (Eds.), *Latent Variable and Latent Structure Models*, Book section 10, pp. 195–223. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

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

Cunha, F. and J. J. Heckman (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Currie, J. (2001). Early childhood education programs. *Journal of Economic Perspectives 15*(2), 213–238.

Davidson, M. C., D. Amso, L. C. Anderson, and A. Diamond (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia 44*(11), 2037–2078.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology 64*, 135–168.

Diamond, A. (2016). *Why Improving and Assessing Executive Functions Early in Life is Critical*, pp. 11–43. Washington, D.C.: American Psychological Association.

Diamond, A. and K. Lee (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science 333*(6045), 959–964.

Diamond, A. and D. S. Ling (2020). Review of the evidence on, and fundamental questions about, efforts to improve executive functions, including working memory. In J. M. Novick, M. F. Bunting, M. R. Dougherty, and R. W. Engle (Eds.), *Cognitive and Working Memory Training, Perspectives from Psychology, Neuroscience, and Human Development*, Book section 8, pp. 143–431. New York City, New York: Oxford University Press.

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.

DiPrete, T. A. and J. L. Jennings (2012). Social and behavioral skills and the gender gap in early educational achievement. *Social Science Research 41*(1), 1–15.

Duncan, G. J., C. J. Dowsett, A. Claessens, K. Magnuson, A. C. Huston, P. Klebanov, L. S. Pagani, L. Feinstein, M. Engel, J. Brooks-Gunn, H. Sexton, K. Duckworth, and C. Japel (2007). School readiness and later achievement. *Developmental Psychology 43*(6), 1428–1446.

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

Gormley Jr, W. T. and T. Gayer (2005). Promoting school readiness in oklahoma: An evaluation of tulsa's pre-k program. *Journal of Human Resources 40*(3), 533–558.

Gray-Lobe, G., P. A. Pathak, and C. R. Walters (2021). The long-term effects of universal preschool in Boston. Working Paper 28756, National Bureau of Economic Research, Cambridge, Massachusetts.

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

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 2052–2086.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Yavitz (2010). The rate of return to the highscope perry preschool program. *Journal of Public Economics 94*(1), 114–128.

Heckman, J. J. and S. Mosso (2014). The economics of human development and social mobility. *Annual Review of Economics 6*, 689–733.

Jenkins, J. M., G. J. Duncan, A. Auger, M. Bitler, T. Domina, and M. Burchinal (2018). Boosting school readiness: Should preschool teachers target skills or the whole child? *Economics of Education Review 65*, 107–125.

Kosse, F., T. Deckers, P. Pinger, H. Schildberg-Hörisch, and A. Falk (2020). The formation of prosociality: Causal evidence on the role of social environment. *Journal of Political Economy 128*(2), 434–467.

Kotlarski, I. (1967). On characterizing the gamma and the normal distribution. *Pacific Journal of Mathematics 20*(1), 69–76.

Lenes, R., M. M. McClelland, D. ten Braak, T. Idsøe, and I. Størksen (2020). Direct and indirect pathways from children's early self-regulation to academic achievement in fifth grade in norway. *Early Childhood Research Quarterly 53*, 612–624.

Matsumoto, K. and K. Tanaka (2004). Conflict and cognitive control. *Science 303*(5660), 969–970.

McClelland, M. M., C. E. Cameron, C. M. Connor, C. L. Farris, A. M. Jewkes, and F. J. Morrison (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology 43*(4), 947–959.

McClelland, M. M., C. E. Cameron, R. Duncan, R. P. Bowles, A. C. Acock, A. Miao, and M. E. Pratt (2014). Predictors of early growth in academic achievement: The head-toes-knees-shoulders task. *Frontiers in Psychology 5*(599), 1–14.

McCoy, D. C., K. Gonzalez, and S. Jones (2019). Preschool self-regulation and preacademic skills as mediators of the long-term impacts of an early intervention. *Child Development 90*(5), 1544–1558.

Miller, H. V., J. C. Barnes, and K. M. Beaver (2011). Self-control and health outcomes in a nationally representative sample. *American Journal of Health Behavior 35*(1), 15–27.

Moffitt, T. E., L. Arseneault, D. Belsky, N. Dickson, R. J. Hancox, H. Harrington, R. Houts, R. Poulton, B. W. Roberts, S. Ross, M. R. Sears, W. M. Thomson, and A. Caspi (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America 108*(7), 2693–2698.

Nagin, D. and R. E. Tremblay (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development 70*(5), 1181–1196.

OECD (2006). *Starting Strong II: Early Childhood Education and Care*. Paris, France: OECD Publishing.

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.

Rege, M., I. Størksen, I. F. Solli, A. Kalil, M. M. McClelland, D. ten Braak, R. Lenes, S. Lunde, S. Breive, M. Carlsen, I. Erfjord, and P. S. Hundeland (2021). The effects of a structured curriculum on preschool effectiveness: A field experiment. *Journal of Human Resources*.

Riggs, N. R., M. T. Greenberg, C. A. Kusché, and M. A. Pentz (2006). The mediational role of neurocognition in the behavioral outcomes of a social-emotional prevention program in elementary school students: Effects of the paths curriculum. *Prevention Science 7*(1), 91–102.

Romano, E., L. Babchishin, L. S. Pagani, and D. Kohen (2010). School readiness and later achievement: Replication and extension using a nationwide Canadian survey. *Developmental Psychology 46*(5), 995–1007.

Romano, J. P. and M. Wolf (2005). Stepwise multiple testing as formalized data snooping. *Econometrica 73*(4), 1237–1282.

Rueda, M. R., M. I. Posner, and M. K. Rothbart (2005). The development of executive attention: Contributions to the emergence of self-regulation. *Developmental Neuropsychology 28*(2), 573–594.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*(2), 461–464.

Skrondal, A. and S. Rabe-Hesketh (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models* (1 ed.). Interdisciplinary Statistics. Boca Raton, Florida: Chapman & Hall/CRC.

Størksen, I., I. T. Ellingsen, M. S. Tvedt, and E. M. C. Idsøe (2013). Norsk vokabulartest (nvt) for barn i overgangen mellom barnehage og skole: Psykometrisk vurdering av en nettbrettbasert test [norwegian vocabulary test (nvt) for children in the transition between kindergarten and school: Psychometric assessment of a tablet-based test]. *Spesialpedagogikks Forskingsdel 4*(13), 40–54.

Størksen, I. and R. Mosvold (2013). Assessing early math skills with tablet computers: Development of the ani banani math test (ABMT) for young children [Utdanning2020].

Størksen, I., D. ten Braak, S. Breive, R. Lenes, S. Lunde, M. Carlsen, and M. Rege (2018). *Lekbasert Læring: et Forskningsbasert Førskoleopplegg fra Agderprosjektet [Playful Learning: a Research based Preschool Curriculum from the Agder Project].* Oslo, Norway: GAN Aschehoug.

ten Braak, D. and I. Størksen (2021). Psychometric properties of the Ani Banani Math Test. *European Journal of Developmental Psychology 18*(4), 610–628.

Thomson, G. H. (1938). Methods of estimating mental factors. *Nature 141*(3562), 246.

Troller-Renfree, S. V., M. A. Costanzo, G. J. Duncan, K. Magnuson, L. A. Gennetian, H. Yoshikawa, S. Halpern-Meekin, N. A. Fox, and K. G. Noble (2022). The impact of a poverty reduction intervention on infant brain activity. *Proceedings of the National Academy of Sciences 119*(5), 1–8.

Wald, A. (1950). Note on the identification of economic relations. In T. C. Koopmans and J. Marschak (Eds.), *Statistical Inference in Dynamic Economic Models*, Book section 3, pp. 238–244. New York City, New York: John Wiley & Sons, Inc.

Wechsler, F. (1991). *Wechsler Intelligence Scale for Children (3rd ed.) (WISC-III) Manual.* San Antonio, Texas: The Psychological Corporation.

Supplementary Material

Supplement to “Cross-Productivities of Executive Functions”: Online Appendices

By

Maximiliaan W. P. Thijssen
Mari Rege
Ingeborg F. Solli
Ingunn Størksen

A.1. Further Notes on the Data

Table A.1.1 and Table A.1.2 present an overview of missingness. Table A.1.1 reports the missing values in the Agder Project data, and Table A.1.2 shows the missingness in Statistics Norway's data. To compute the missing values for the assessment data, we first estimate the measurement models for each test. Then, we predict the factor scores. We changed the predicted factor scores to missing if any of the individual items were missing. We then calculate, based on these predicted factor scores, the number of observed and missing values. This procedure implies that a single missing item will result in the deletion of the whole row. However, rarely are only a few items missing. In most cases, either all individual items are missing or none. We use full information (parametric) maximum likelihood estimation (Anderson, 1957). Even when data are only missing at random, full-information maximum likelihood uses all available data and provides valid point estimates.

We regress baseline and predetermined variables on a treatment indicator and randomization-block fixed effects to test differences in baseline characteristics. We cluster the standard errors at the randomization-block level. Table A.1.3 reports the results. All baseline characteristics are balanced, including baseline skills of children. We also allow for multiple hypothesis testing using the Romano and Wolf (2005) procedure implemented in the Stata package "`rwolf`" (Clarke et al., 2020). None of the differences are significant.

Figure A.1.1 shows descriptives for the six assessments calculated over all non-missing observations. We plot the mean and standard deviation across all three assessment waves (baseline 2016, post-intervention 2017, and follow-up 2018). The length of the whiskers signifies one standard deviation above (and below) the mean. Figure A.1.1 shows a strong mean development across all six tests from the start of the last year in preschool (*i.e.*, August 2016) to midway through primary school (*i.e.*, March 2018).

We matched the assessment data shown in Figure A.1.1 to Statistics Norway's registry data. As part of the Agder Project, we collected data on sex, birth month, randomization-block indicators, and whether we received late parental consent. In Statistics Norway's registry data, we observe education levels for the mother and father, income for both parents, and the parent's country of origin. Using the information on country of origin, we construct an indicator variable that takes on the value of one if at least one of the parents is a non-Western immigrant and zero otherwise. Following Rege et al. (2021), we do not include categories for the birth month or the parents' education in our model. Instead, we estimate a single parameter for each of these variables.

Table A.1.4 reports descriptive information for the child and parental characteristics. In Table A.1.4, we observe that about half of the children are female. The median educational attainment for mothers is the first stage of higher education (undergraduate level), whereas for fathers it is the upper secondary (final) year. On average, about 15 percent of the sampled children have at least one parent who is a non-Western immigrant. Mean family income is about NOK 888,416, 63 percent earned by the father (NOK 555,051) and 37 percent by the mother (NOK 330,587).

Table A.1.5 and Table A.1.6 are similar to Table 1 and Table 2, respectively. The difference is the handling of some of the missing values. As explained in the beginning of this section, we changed the predicted factor scores to missing if *any* of the individual items were missing. Consequently, if we observe two out of three tests for a particular child, the child's score is changed to missing. In Table A.1.5 and Table A.1.6, we instead replace missing values based on the available information. If we observe two out of three tests, then we calculate the predicted factor scores using the available tests. The results reported in Table A.1.5 and Table A.1.6 do not qualitatively change our conclusions.

Table A.1.1. Overview of Missing Values: Agder Project Data

		Observed		Missing	
	(1) Period	(2) Obs.	(3) Pct.	(4) Obs.	(5) Pct.
Head-Toes-Knees-Shoulders Test	2016	516	73.6	185	26.4
Hearts and Flowers Test	2017	635	90.6	66	9.4
Digit Span Test	2017	641	91.4	60	8.6
Hearts and Flowers Test	2016	642	91.6	59	8.4
Head-Toes-Knees-Shoulders Test	2017	645	92.0	56	8.0
Blending Test	2017	645	92.0	55	8.0
Ani Banani Math Test	2016	646	92.3	55	7.8
Norwegian Vocabulary Test	2016	647	92.3	54	7.7
Blending Test	2016	648	92.4	53	7.6
Digit Span Test	2016	648	92.4	53	7.6
Digit Span Test	2018	653	93.2	48	6.8
Blending Test	2018	658	93.9	43	6.1
Head-Toes-Knees-Shoulders Test	2016	659	94.0	42	6.0
Norwegian Vocabulary Test	2018	659	94.0	42	6.0
Hearts and Flowers Test	2018	660	94.2	41	5.8
Ani Banani Math Test	2018	661	94.3	40	5.7
Ani Banani Math Test	2016	663	94.6	38	5.4

Notes. This table reports the descriptive frequencies related to observed and missing observations for the assessment data from the Agder Project. For the Head-Toes-Knees-Shoulders Test in 2016, some children did not complete the last ten items (10 out of a total of 30). The reason for the occurrence of these missing values is that the test stops after a child misses a certain number of items. At this point, it is unlikely that the child will complete the later items. For this reason, we replace these ten items with zeros. There is some minor variation in missingness across test items.

Table A.1.2. Overview of Missing Values: Family Characteristics

		Observed		Missing	
	(1) Period	(2) Obs.	(3) Pct.	(4) Obs.	(5) Pct.
Education Father	2016	666	95.0	35	5.0
Education Mother	2016	676	96.4	25	3.6
Non-Western Immigrant	2016	682	97.3	19	2.7
Income Father	2016	683	97.4	18	2.6
Income Mother	2016	698	99.6	3	0.4
Birth Month	2016	701	100.0	0	0.0
The Child is Female	2016	701	100.0	0	0.0
Late Parental Consent	2016	701	100.0	0	0.0

Notes. This table reports the descriptive frequencies of observed and missing observations for the data from Statistics Norway and the child and parental characteristics collected as part of the Agder Project's data collection.

Table A.1.3. Balance Test

	Treatment Group		Control Group		Differences		
	(1) Obs.	(2) Mean	(3) Obs.	(4) Mean	(5) Mean	(6) *p*-val.	(7) RW-*p*
Panel A: Predetermined Variables							
The Child is Female	383	0.48	318	0.52	−0.03	0.44	0.85
Birth Month	383	6.15	318	6.19	−0.02	0.93	0.99
Education Mother	365	4.71	311	4.86	−0.12	0.38	0.85
Education Father	362	4.46	304	4.50	−0.05	0.76	0.99
Non-Western Immigrant	371	0.18	311	0.12	0.05	0.21	0.63
Income Mother (in NOKs)	381	321,391	317	341,640	−20,495	0.38	0.85
Income Father (in NOKs)	372	560,080	311	549,035	11,048	0.58	0.92
Panel B: Premeasured Variables							
Blending Test	383	2.37	318	2.63	−0.29	0.36	0.85
Head-Toes-Knees-Shoulders	383	21.52	318	19.98	1.47	0.40	0.85
Vocabulary Test	383	9.93	318	9.62	0.30	0.48	0.85
Hearts and Flowers Test	383	26.34	318	25.28	1.08	0.35	0.85
Ani Banani Math Test	368	3.14	318	3.12	−0.00	0.98	0.99
Digit Span Test	359	5.39	289	5.46	−0.06	0.76	0.99

Notes. This table reports the results of a balance test. Column (6) reports p-values (*p*-val.) for without correction for multiple hypothesis testing. Column (7) reports *p*-values with Romano-Wolf (RW-*p*) correction for multiple hypothesis testing. Standard errors are clustered at the randomization-block level.

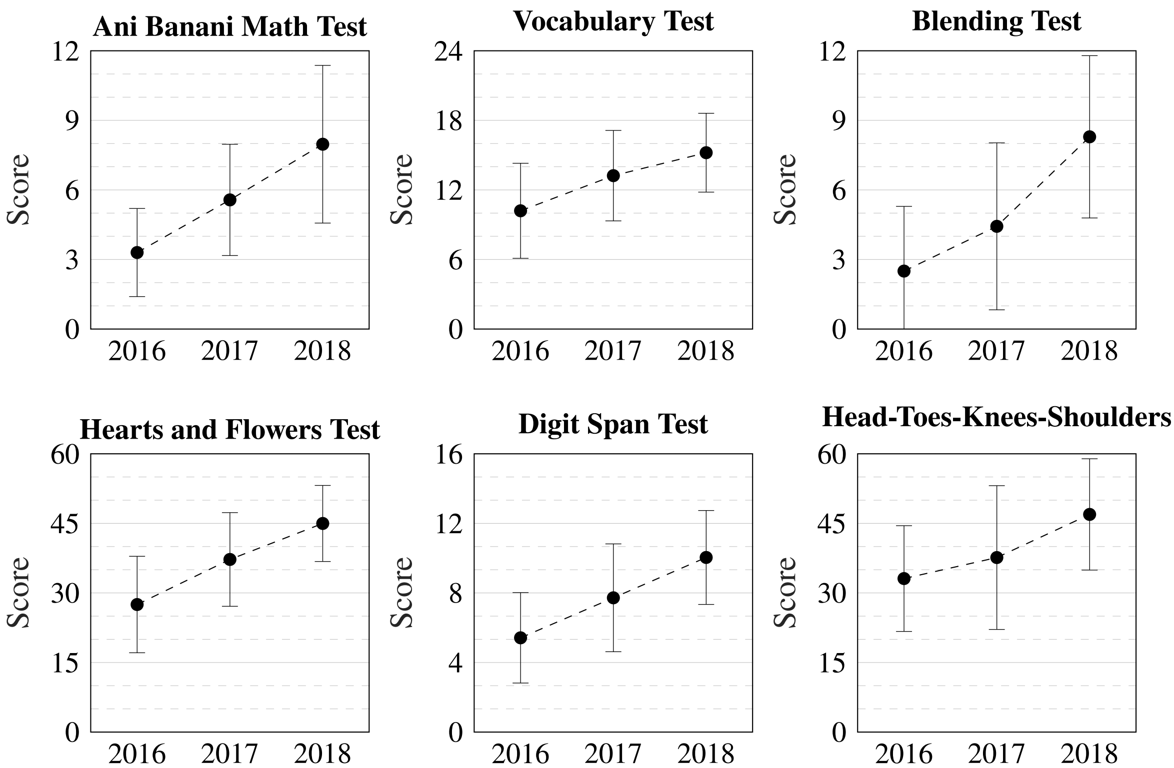

Figure A.1.1. Mean Score and Standard Deviation by Assessment Wave

Table A.1.4. Descriptive Statistics Child and Parental Characteristics

	(1) Mean	(2) *SD*	(3) Obs.
The Child is Female	0.49	0.50	701
Birth Month	6.00^{a}	3.19	701
Education Mother	6.00^{a}	1.67	676
Education Father	4.00^{a}	1.59	666
Non-Western Immigrant	0.15	0.36	682
Income Mother (in NOKs)	330,587	213,546	698
Income Father (in NOKs)	555,051	268,071	683
Late Parental Consent	0.19	0.39	701

Notes. This table reports descriptive statistics for the child and parental characteristics. Education comprises eight categories: (1) primary education; (2) lower secondary education; (3) upper secondary (basic); (4) upper secondary (final year); (5) post-secondary, not higher education; (6) first stage of higher education, undergraduate level; (7) first stage of higher education, graduate-level; and (8) second stage of higher education (postgraduate education).
[a] We report the median rather than the mean.

Table A.1.5. Self-Reinforcement and Cross-Production of Skills

	(1) Mathematics	(2) Language	(3) EFs
Panel A: Developmental Stage 1 (2016 - 2017)			
EFs	0.383	0.185	0.769
	(0.047)	(0.051)	(0.055)
Mathematics	0.600	0.281	0.330
	(0.046)	(0.067)	(0.061)
Language	0.076	0.265	0.135
	(0.046)	(0.060)	(0.043)
Panel B: Developmental Stage 2 (2017 - 2018)			
EFs	0.414	0.260	0.923
	(0.049)	(0.052)	(0.066)
Mathematics	0.594	0.109	0.268
	(0.054)	(0.065)	(0.058)
Language	0.047	0.222	−0.029
	(0.041)	(0.050)	(0.039)
Control Variables	Yes	Yes	Yes

Notes. This table reports the self-productivity parameter estimates for the first developmental stage (Panel A) and the second (Panel B). The columns denote the dependent variables, and the rows denote the independent variables. Consider the first row (i.e., EF skills) in Panel A; the reported results measure the effect of EF skills measured at baseline (2016) on mathematical skills (Column 1), language skills (Column 3), and EFs (Column 5) measured post-intervention (2017). Columns (2), (4), and (6) differ from Columns (1), (3), and (5). Consequently, the table reports estimates from twelve models. We include the following control variables: child sex, birth month, whether or not at least one of the parents is a non-Western immigrant, parental education, family income, an indicator for late parental consent, and randomization-block indicators. All of the reported parameter estimates are given in standard deviations. We standardize the parameter estimates using the "true" variance, not the predicted variance. We present randomization-block clustered standard errors in parentheses. Standard errors are computed using a wild residual bootstrap procedure (1,000 bootstrap samples). All models are estimated with full-information maximum likelihood (total observations 701). Online Appendix A.1 provides an overview of missingness.

Table A.1.6. Treatment Effects and Decomposition

	(1) Mathematics	(2) Language	(3) EFs
Panel A: Post-Intervention and Follow-Up Treatment Effects			
Post-Intervention	0.176	0.052	0.174
	(0.089)	(0.103)	(0.066)
Follow-Up	0.196	0.042	0.072
	(0.097)	(0.114)	(0.080)
Panel B: Decomposition of Follow-Up Treatment Effect			
EFs	0.049	0.031	0.102
	(0.020)	(0.013)	(0.045)
	[25.0%]	[73.8%]	[64.2%]
Mathematics	0.028	0.005	0.013
	(0.014)	(0.003)	(0.007)
	[14.3%]	[11.9%]	[8.2%]
Language	0.001	0.005	−0.001
	(0.002)	(0.009)	(0.002)
	[0.5%]	[11.9%]	[0.6%]
Unmeasured Variables	0.118	0.001	−0.043
	(0.093)	(0.113)	(0.068)
	[60.2%]	[2.4%]	[27.0%]
Control Variables	Yes	Yes	Yes

Notes. This table reports the parameter estimates for the treatment effect decomposition. Panel A reports the post-intervention treatment effect (*i.e.*, differences between treated and non-treated children post-intervention) and the follow-up treatment effect (*i.e.*, differences between treated and non-treated children at the follow-up). Panel B decomposes the follow-up treatment effect in measured and unmeasured variables. The columns denote the dependent variables, and the rows denote the independent variables. Consider the first row (*i.e.*, post-intervention treatment effect) in Panel A; the reported results measure the effect of the structured curriculum on mathematical skills (Column 1), language skills (Column 3), and EF skills (Column 5) measured post-intervention (2017). Columns (2), (4), and (6) differ from Columns (1), (3), and (5). Consequently, the table reports estimates from twelve models. We include the following control variables: child sex, birth month, whether or not at least one of the parents is a non-Western immigrant, parental education, family income, an indicator for late parental consent, and randomization-block indicators. All of the reported parameter estimates are in standard deviations. We standardize the parameter estimates using the "true" variance, not the predicted variance. We present randomization-block clustered standard errors in parentheses. Standard errors are computed using a wild residual bootstrap procedure (1,000 bootstrap samples). In brackets, we report the relative contributions. We compute these relative contributions by taking the absolute value of the estimate divided by the sum of absolute values of all estimates multiplied by 100. All models are estimated with full-information maximum likelihood (total observations 701). Online Appendix A.1 provides an overview of missingness.

A.2. Further Notes on Measurement Models

We first identify a measurement model in which the manifest variables are continuous. We next consider a measurement model where the manifest variables are categorical (*e.g.*, binary, ordinal). For identification, we assume, as is the case in our sample, a minimum of two *valid* measures (*i.e.*, manifest variables) in each period and a minimum of two periods.

A.2.1. Continuous Scales

Consider the following measurement model,

$$\mathrm{M}_{l,i,t} = \mu_{l,t} + \lambda_{l,t}\mathrm{S}_{i,t} + \zeta_{l,i,t}, \tag{A.2.1}$$

where $\mathrm{M}_{l,i,t}$ denotes the lth manifest variable (for $l = 1,2$) for child i (for $i = 1,\dots,N$) and time t (for $t = 0,1$), $\mu_{l,t}$ denotes the intercept, $\lambda_{l,t}$ denotes the factor loading, $\mathrm{S}_{i,t}$ denotes the unobserved skill (*i.e.*, the common factor), and $\zeta_{l,i,t}$ denotes the error term (*i.e.*, the unique factor). We use the following normalizations: To set the location, we normalize the mean of the the common factor to 0, $\mathbb{E}(\mathrm{S}_{i,t}) = 0\ \forall\ t = 0,1$, where $\mathbb{E}(\cdot)$ denotes the expectation operator. To set the scale, we normalize a factor loading (say the first) to 1, $\lambda_{1,t} = 1\ \forall\ t = 0,1$. Additionally, we made the following assumptions: (i) $\mathrm{Cov}(\zeta_{l,i,t}, \mathrm{S}_{i,t}) = 0\ \forall\ l = 1,2, t = 0,1$, (ii) $\mathrm{Cov}(\zeta_{l,i,t}, \zeta_{l',i,t'}|\mathrm{S}_{i,t}) = 0\ \forall\ l,l' = 1,2,\ l \neq l',\ t = 0,1$, and (iii) the unique factor is independent across children, where $\mathrm{Cov}(\cdot,\cdot)$ denotes the covariance operator. Below, we write the unknown parameters (right-hand side) as a function of known (or identified) parameters (left-hand side). We can write the following covariances,

$$\mathrm{Cov}(\mathrm{M}_{1,i,t}, \mathrm{M}_{1,i,t+1}) = \mathrm{Cov}(\mathrm{S}_{i,t}, \mathrm{S}_{i,t+1}), \tag{A.2.2}$$

$$\mathrm{Cov}(\mathrm{M}_{1,i,t}, \mathrm{M}_{2,i,t+1}) = \lambda_{2,t+1}\mathrm{Cov}(\mathrm{S}_{i,t}, \mathrm{S}_{i,t+1}), \tag{A.2.3}$$

$$\mathrm{Cov}(\mathrm{M}_{2,i,t}, \mathrm{M}_{1,i,t+1}) = \lambda_{2,t}\mathrm{Cov}(\mathrm{S}_{i,t}, \mathrm{S}_{i,t+1}). \tag{A.2.4}$$

We can identify the factor loading, $\lambda_{2,t+1}$, by taking the ratio of Equation (A.2.3) to Equation (A.2.2) and $\lambda_{2,t}$ by taking the ratio of Equation (A.2.4) to Equation (A.2.2).

With the factor loadings identified, we can identify the common factor variance from,

$$\frac{\mathrm{Cov}(\mathrm{M}_{1,i,t}, \mathrm{M}_{2,i,t})}{\lambda_{2,t}} = \frac{\lambda_{2,t}\mathrm{Var}(\mathrm{S}_{i,t})}{\lambda_{2,t}} = \mathrm{Var}(\mathrm{S}_{i,t}), \tag{A.2.5}$$

for $t = 0,1$. With the common factor variance identified, we can identify the unique factor variance from,

$$\mathrm{Var}(\mathrm{M}_{l,i,t}) - (\lambda_{l,t})^2\mathrm{Var}(\mathrm{S}_{i,t}) = \mathrm{Var}(\zeta_{l,i,t}), \tag{A.2.6}$$

for $t = 0,1$ and $l = 1,2$. We next identify the mean structure. We can identify the intercepts from the expectations,

$$\mathbb{E}(\mathrm{M}_{l,i,t}) = \mu_{l,t} + \lambda_{l,t}\mathbb{E}(\mathrm{S}_{i,t}) = \mu_{l,t}, \tag{A.2.7}$$

for $t = 0,1$ and $l = 1,2$.

With the factor loadings identified, we can (nonparametrically) identify the distribution of common and unique factors using Kotlarski's lemma (see Lemma 1 and Remark 4 in Kotlarski, 1967, pp. 70–73). However, these (nonparametric) identification results no longer hold when measures have an ordinal scale.

A.2.2. Categorical Scales

Consider again the measurement model in Equation (A.2.11)

$$\mathrm{M}^{*}_{l,i,t} = \mu_{l,t} + \lambda_{l,t}\mathrm{S}_{i,t} + \zeta_{l,i,t}, \tag{A.2.8}$$

but the observed manifest variable, $\mathrm{M}^{*}_{l,i,t}$ is now a latent response variable related to observed (categorical) responses through a threshold function. Consider the Head-Toes-Knees-Shoulders Test. Per item, children received two points when they performed the task correctly, one point when they carried out an incorrect move but ended with a correct response, and zero points for incorrect responses. We can write the following threshold function,

$$\mathrm{M}_{l,i,t} = \begin{cases} 0 & \text{if } \mathrm{M}^{*}_{l,i,t} < \tau_{1,l,t} \\ 1 & \text{if } \tau_{1,l,t} < \mathrm{M}^{*}_{l,i,t} < \tau_{2,l,t} \\ 2 & \text{if } \tau_{2,l,t} < \mathrm{M}^{*}_{l,i,t} \end{cases}$$

where $\tau_{1,l,t}$ and $\tau_{2,l,t}$ are threshold parameters that provide a mapping from the common factor, $\mathrm{S}_{i,t}$, and unique factor, $\zeta_{l,i,t}$, to the observed ranks.

Consider the case in which we assume the factors are normally distributed. Doing so, we can write the probabilities associated with a child achieving a particular score as follows:

$$\Pr(\mathrm{M}_{l,i,t} = 0) = \Phi\left(\frac{\tau_{1,l,t}}{\sqrt{(\lambda_{l,t})^2 + \mathrm{Var}(\zeta_{l,i,t})}}\right)$$

$$\Pr(\mathrm{M}_{l,i,t} = 1) = \Phi\left(\frac{\tau_{1,l,t} - \tau_{2,l,t}}{\sqrt{(\lambda_{1,l,t})^2 + \mathrm{Var}(\upsilon_{l,i,t})}}\right) - \Phi\left(\frac{\tau_{1,l,t}}{\sqrt{(\lambda_{l,t})^2 + \mathrm{Var}(\zeta_{l,i,t})}}\right)$$

$$\Pr(\mathrm{M}_{l,i,t} = 2) = 1 - \Phi\left(\frac{\tau_{1,l,t} - \tau_{2,l,t}}{\sqrt{(\lambda_{l,t})^2 + \mathrm{Var}(\zeta_{l,i,t})}}\right)$$

where $\Phi(\cdot)$ denotes the cumulative normal distribution.

Since variances are redundant in polychoric and tetrachoric correlation matrices, we require additional normalizations as there would be 13 unknowns and only 10 knowns. If we additionally normalize the unique factor variance to 1, $\mathrm{Var}(\zeta_{l,i,t}) = 1 \;\forall\; l = 1, 2, t = 0, 1$, we can establish (local) identification. See, for example, Skrondal and Rabe-Hesketh (2004, pp. 135–158). We can use our known and unknown parameters to demonstrate the local identification. Let $\boldsymbol{\vartheta}$ denote the parameter vector with unknown parameters and let $\mathbf{m}(\boldsymbol{\vartheta})$ denote the vector with reduced-form thresholds and covariances. We can then compute the $(10, 13)$-dimensional Jacobian, $\mathbb{J}(\boldsymbol{\vartheta}) = \boldsymbol{\partial}\mathbf{m}(\boldsymbol{\vartheta})/\boldsymbol{\partial\vartheta}$. The matrix rank of the Jacobian is 9, which is equal to the number of unknown parameters, so the model is locally identified (if $\boldsymbol{\vartheta}$ is a regular point) (Wald, 1950; Skrondal and Rabe-Hesketh, 2004).

A.2.3. Estimating the Measurement Models

We first estimate "lower-level" measurement models (results available on request). Specifically, we start by estimating a measurement model for the Head-Toes-Knees-Shoulders Test (30 items), the Hearts and Flowers Test (60 items), the Norwegian Vocabulary Test (20 items) and the Blending Test (12 items). We then use each estimated measurement model to assign values to the common factors. Next, we estimate the measurement models for EF skills, mathematical skills, and language skills using these predicted factor scores. For children's language skills, we only have two measures: the Norwegian Vocabulary Test and the Blending Test. Any prediction error would become part of the error term.

Before we estimated the lower-level measurement models, we had to take a position regarding the distribution and link function. Since we have no prior, we selected a model based on Akaike's Information Criterion (AIC: Akaike, 1987) and the Bayesian Information Criterion (BIC: Schwarz, 1978). We considered (i) a Gaussian distribution and identity link function, (ii) a binomial (or ordinal) distribution, and (iii) logit link function, and a binomial (ordinal) distribution and probit link function.

Table A.2.7 presents the AIC and BIC values. In Panel A, we document the AIC and BIC for the Ani Banani Math Test. We observe that a Gaussian distribution with identity link function results in a (comparatively) better fit in the first (August 2016) and third (March 2018) assessment waves. We observe that a binomial distribution with a logit link function fits the second assessment wave better. However, to maintain consistency in estimating this measurement model, we choose a Gaussian distribution with an identity link function in each assessment wave. In Panel B, we document the AIC and BIC for the Head-Toes-Knees-Shoulders Test. We observe that an ordinal distribution with logit link function results in a better fit in the first (August 2016) and second (June 2017) assessment waves. In the third wave, a probit link function produces a better fit. We choose an ordinal distribution with a logit link function in each assessment wave to maintain consistency. In Panel C, we document the AIC and BIC for the Hearts and Flowers Test.

We observe that a binomial distribution with a logit link function fits each assessment wave better. Therefore, we use a binomial distribution with a logit link function when estimating the measurement model for further analysis. In Panel D, we document the AIC and BIC for the Norwegian Vocabulary Test. We observe that a Gaussian distribution with an identity link function fits each assessment wave better. Therefore, we use a Gaussian distribution with an identity link function when estimating the measurement model for further analysis. Lastly, in Panel E, we document the AIC and BIC for the Blending Test. We observe that a Gaussian distribution with an identity link function produces a better fit in the first (August 2016) and last (March 2018) measurement waves. A binomial distribution with a probit link function in the second wave results in a better fit. We use a Gaussian distribution with an identity link function in each wave to maintain consistency.

Table A.2.8, Table A.2.9, and Table A.2.10 present the estimates of the measurement models for EFs, mathematical skills, and language skills, respectively. The Head-Toes-Knees-Shoulders Test and Hearts and Flowers Test in Table A.2.8 and the Norwegian Vocabulary Test and Blending Test in Table A.2.10 are the predicted factor scores from the lower-level measurement models. Table A.2.11 reports the variance-covariance matrix for EF skills, mathematical skills, and language skills. Table A.2.12 reports the Pearson correlation matrix for EF skills, mathematical skills, and language skills. These measures of association are based on predicted (Bartlett) factor scores (Bartlett, 1937; Thomson, 1938).

Lastly, Table A.2.13 and Table A.2.14 report the main results. The difference between Table A.2.13 and Table A.2.14, compared with Table 1 and Table 2, is that we do not use a measurement model. Instead, we standardize the individual tests first, compute a simple arithmetic average, and standardize again so that the composite has mean zero and standard deviation one. The results presented in these tables show that we lose a great deal of precision by not accounting for measurement error. Nonetheless, we find that program-induced improvements in EFs in preschool lead to improvements in mathematical skills and language skills in primary school (statistically significant at the ten percent level).

Table A.2.7. Akaike and Bayesian Information Criteria

	August 2016			June 2017			March 2018		
	(1) Iden.	(2) Probit	(3) Logit	(4) Iden.	(5) Probit	(6) Logit	(7) Iden.	(8) Probit	(9) Logit
Panel A: Ani Banani Math Test									
Obs.	665	665	665	651	651	651	661	661	661
AIC	5,702	6,714	6,714	7,795	7,632	7,625	6,821	7,131	7,129
BIC	5,810	6,813	6,810	7,907	7,730	7,724	6,920	7,230	7,228
Panel B: Head-Toes-Knees-Shoulders Test									
Obs.	644	644	644	645	645	645	659	659	659
AIC	35,028	19,921	19,905	40,678	22,369	22,367	34,775	21,469	21,497
BIC	35,430	20,323	20,307	41,081	22,771	22,769	35,179	21,874	21,901
Panel C: Hearts and Flowers Test									
Obs.	642	642	642	635	635	635	660	660	660
AIC	44,965	43,008	42,992	40,141	39,490	39,462	30,636	33,553	33,542
BIC	45,474	43,517	43,501	40,638	39,998	39,970	31,148	34,065	34,054
Panel D: Norwegian Vocabulary Test									
Obs.	647	647	647	648	648	648	659	659	659
AIC	13,043	13,183	13,178	11,372	12,317	12,319	6,904	10,538	10,535
BIC	13,223	13,362	13,357	11,551	12,496	12,497	7,084	10,718	10,715
Panel E: Blending Test									
Obs.	648	648	648	645	645	645	658	658	658
AIC	4,561	5,550	5,568	7,292	7,134	7,146	6,133	6,768	6,747
BIC	4,669	5,657	5,675	7,399	7,241	7,253	6,241	6,875	6,855

Notes. This table reports the number of observations, Akaike's Information Criterion (AIC: Schwarz, 1978), and Bayesian Information Criterion (BIC: Schwarz, 1978). We considered a Gaussian distribution and identity (Iden.) link functions (Columns 1, 4, and 7), a binomial (or ordinal) distribution and probit link function (Columns 2, 5, and 8), and a binomial (ordinal) distribution and logit link function (Columns 3, 6, and 9). All models are estimated with maximum likelihood.

Table A.2.8. Measurement Model Parameter Estimates: Executive Functions

	August 2016		June 2017		March 2018	
	(1)	(2)	(3)	(4)	(5)	(6)
	$\lambda_{l,0}$	$\mathrm{Var}(\zeta_{l,i,0})$	$\lambda_{l,1}$	$\mathrm{Var}(\zeta_{l,i,1})$	$\lambda_{l,2}$	$\mathrm{Var}(\zeta_{l,i,2})$
Digit Span Test	1	0.60	1	0.53	1	0.60
		(0.05)		(0.04)		(0.05)
Head-Toes-Knees-Shoulders	0.77	0.75	0.66	0.47	0.75	0.63
	(0.09)	(0.06)	(0.07)	(0.03)	(0.10)	(0.05)
Hearts and Flowers Test	0.92	0.59	1.03	0.48	1.14	0.59
	(0.10)	(0.05)	(0.09)	(0.04)	(0.13)	(0.05)

Notes. This table reports the measurement model parameter estimates for children's EFs. The common factor is EF skills. The manifest variables are the (Forward/Backward) Digit Span Test, the Head-Toes-Knees-Shoulders Test, and the Hearts and Flowers Test. We standardized the manifest variables to be mean zero and standard deviation one. For this reason, we omit the measurement model intercept estimates. We anchor children's EF skills in the Digit Span Test. Columns (1), (3), and (5) present factor loading estimates. Columns (2), (4), and (6) present unique factor variance estimates. We present the Huber-White standard errors in parentheses. All models are estimated with full-information maximum likelihood (total observations 701). The log-likelihood is $-7,313.07$.

Table A.2.9. Measurement Model Parameter Estimates: Mathematical Skills

	August 2016			June 2017			March 2018		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Items	$\mu_{l,0}$	$\lambda_{l,0}$	Var($\zeta_{l,i,0}$)	$\mu_{l,1}$	$\lambda_{l,1}$	Var($\zeta_{l,i,1}$)	$\mu_{l,2}$	$\lambda_{l,2}$	Var($\zeta_{l,i,2}$)
1	0.24	1	0.12	0.59	1	0.17	0.88	1	0.09
	(0.02)		(0.01)	(0.02)		(0.01)	(0.01)		(0.01)
2	0.10	0.15	0.09	0.26	0.72	0.15	0.47	2.18	0.20
	(0.01)	(0.06)	(0.01)	(0.02)	(0.09)	(0.01)	(0.02)	(0.41)	(0.01)
3	0.08	0.45	0.06	0.29	1.04	0.13	0.59	3.25	0.13
	(0.01)	(0.08)	(0.01)	(0.02)	(0.11)	(0.01)	(0.02)	(0.56)	(0.01)
4	0.37	0.58	0.21	0.60	0.42	0.23	0.82	0.27	0.15
	(0.02)	(0.10)	(0.01)	(0.02)	(0.07)	(0.01)	(0.01)	(0.17)	(0.01)
5	0.43	0.68	0.22	0.70	0.48	0.20	0.83	0.88	0.13
	(0.02)	(0.10)	(0.01)	(0.02)	(0.07)	(0.01)	(0.01)	(0.19)	(0.01)
6	0.13	0.43	0.11	0.32	0.62	0.19	0.57	1.44	0.22
	(0.01)	(0.07)	(0.01)	(0.02)	(0.07)	(0.01)	(0.02)	(0.28)	(0.01)
7	0.71	0.63	0.18	0.87	0.34	0.11	0.93	0.47	0.06
	(0.02)	(0.10)	(0.01)	(0.01)	(0.06)	(0.01)	(0.01)	(0.14)	(0.01)
8	0.66	0.68	0.20	0.84	0.43	0.12	0.87	0.91	0.14
	(0.02)	(0.10)	(0.01)	(0.01)	(0.06)	(0.01)	(0.01)	(0.19)	(0.01)
9	0.42	0.69	0.22	0.60	0.57	0.22	0.82	1.12	0.14
	(0.02)	(0.09)	(0.01)	(0.02)	(0.08)	(0.01)	(0.02)	(0.24)	(0.01)
10	0.11	0.62	0.08	0.37	1.07	0.15	0.70	2.09	0.16
	(0.01)	(0.06)	(0.01)	(0.02)	(0.07)	(0.01)	(0.02)	(0.29)	(0.01)
11	0.03	0.19	0.03	0.15	0.71	0.09	0.49	3.08	0.15
	(0.01)	(0.05)	(0.01)	(0.01)	(0.09)	(0.01)	(0.02)	(0.55)	(0.01)

Notes. This table reports the measurement model parameter estimates for children's mathematical skills. The common factor is a mathematical skill. The manifest variables are the 11 (dichotomous) items presented in the rows. We anchor children's mathematical skills in the first item. This item asks children to copy a pattern. Note that the ordering of items is not the same as the ordering in the original Ani Banani Math Test. Columns (1), (4), and (7) present the measurement model intercepts. Columns (2), (5), and (8) present the factor loadings. Columns (3), (6), and (9) present unique factor variance estimates. We present the Huber-White standard errors in parentheses. All models are estimated with full-information maximum likelihood (total observations 701). The log-likelihood is $-9,757.48$.

Table A.2.10. Measurement Model Parameter Estimates: Language Skills

	August 2016		June 2017		March 2018	
	(1) $\lambda_{l,0}$	(2) $\text{Var}(\zeta_{l,i,0})$	(3) $\lambda_{l,1}$	(4) $\text{Var}(\zeta_{l,i,1})$	(5) $\lambda_{l,2}$	(6) $\text{Var}(\zeta_{l,i,2})$
Blending Test	1	0.96	1	0.87	1	0.85
		(0.06)		(0.05)		(0.05)
Vocabulary Test	3.71	0.23	2.37	0.24	2.48	0.17
	(0.69)	(0.09)	(0.26)	(0.05)	(0.30)	(0.06)

Notes. This table reports the measurement model parameter estimates for children's language skills. The common factor is language skills. The manifest variables are the Blending Test and the Norwegian Vocabulary Test. We standardized the manifest variables to be mean zero and standard deviation one. For this reason, we omit the measurement model intercept estimates. We anchor children's language skills in the Blending Test. Columns (1), (3), and (5) present factor loading estimates. Columns (2), (4), and (6) present unique factor variance estimates. We present the Huber-White standard errors in parentheses. All models are estimated with full-information maximum likelihood (total observations 701). The log-likelihood is $-4,775.79$.

Table A.2.11. Variance-Covariance Matrix Children's Skills

	August 2016			June 2017			March 2018		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. EF Skills	0.61	0.12	0.07	0.36	0.12	0.08	0.29	0.04	0.07
2. Mathematical Skills	0.12	0.10	0.02	0.11	0.06	0.03	0.10	0.02	0.03
3. Language Skills	0.07	0.02	0.08	0.06	0.02	0.03	0.05	0.01	0.02
4. EF Skills	0.35	0.11	0.06	0.52	0.14	0.11	0.31	0.05	0.07
5. Mathematical Skills	0.12	0.06	0.02	0.14	0.10	0.04	0.11	0.02	0.02
6. Language Skills	0.08	0.03	0.03	0.11	0.04	0.17	0.07	0.01	0.04
7. EF Skills	0.29	0.10	0.05	0.31	0.11	0.07	0.44	0.04	0.05
8. Mathematical Skills	0.04	0.02	0.01	0.05	0.02	0.01	0.04	0.01	0.01
9. Language Skills	0.07	0.03	0.02	0.07	0.02	0.04	0.05	0.01	0.14

Notes. This table reports the variance-covariance matrix for children's EF skills, mathematical skills, and language skills in each of the three assessment waves. We calculate these estimates based on predicted (Bartlett) factor scores (Bartlett, 1937; Thomson, 1938). The numbers in the columns refer to the numbers in the rows (and the corresponding skills). For instance, in Panel A, the first row (i.e., (1) EF skills) and the second column (2) present the covariance between children's EFs and mathematical skills.

Table A.2.12. Pearson Correlation Matrix Children's Skills

	August 2016			June 2017			March 2018		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. EF Skills	1.00	0.49	0.32	0.65	0.50	0.24	0.56	0.44	0.22
2. Mathematical Skills	0.49	1.00	0.21	0.50	0.60	0.23	0.46	0.50	0.22
3. Language Skills	0.32	0.21	1.00	0.28	0.23	0.30	0.26	0.17	0.21
4. EF Skills	0.64	0.50	0.28	1.00	0.61	0.36	0.65	0.61	0.27
5. Mathematical Skills	0.50	0.60	0.23	0.61	1.00	0.31	0.52	0.65	0.20
6. Language Skills	0.24	0.23	0.30	0.36	0.31	1.00	0.26	0.25	0.27
7. EF Skills	0.56	0.46	0.26	0.65	0.52	0.26	1.00	0.54	0.19
8. Mathematical Skills	0.44	0.50	0.17	0.61	0.65	0.25	0.51	1.00	0.19
9. Language Skills	0.22	0.22	0.21	0.27	0.20	0.27	0.19	0.19	1.00

Notes. This table reports the Pearson correlation matrix for children's EF skills, mathematical skills, and language skills in each of the three assessment waves. We calculate these estimates based on predicted (Bartlett) factor scores (Bartlett, 1937; Thomson, 1938). The numbers in the columns refer to the numbers in the rows (and the corresponding skills). For instance, in Panel A, the first row (i.e., (1) EF skills) and the second column (2) present the covariance between children's EFs and mathematical skills.

Table A.2.13. Self-Reinforcement and Cross-Production of Skills using Arithmetic Averages

	Mathematics		Language		EFs	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Developmental Stage 1 (2016 - 2017)						
EF	0.219	0.206	0.103	0.114	0.534	0.535
	(0.038)	(0.038)	(0.036)	(0.032)	(0.043)	(0.043)
Mathematics	0.425	0.415	0.112	0.119	0.194	0.196
	(0.029)	(0.030)	(0.041)	(0.043)	(0.042)	(0.043)
Language	0.140	0.145	0.556	0.543	0.079	0.070
	(0.040)	(0.041)	(0.078)	(0.043)	(0.030)	(0.032)
Panel B: Developmental Stage 2 (2017 - 2018)						
EFs	0.287	0.296	0.171	0.138	0.568	0.557
	(0.034)	(0.033)	(0.039)	(0.033)	(0.038)	(0.038)
Mathematics	0.426	0.412	0.060	0.083	0.187	0.187
	(0.041)	(0.043)	(0.047)	(0.035)	(0.044)	(0.043)
Language	0.066	0.081	0.543	0.499	0.029	0.025
	(0.034)	(0.037)	(0.035)	(0.034)	(0.035)	(0.033)
Control Variables	No	Yes	No	Yes	No	Yes

Notes. This table reports the self-productivity parameter estimates. The columns denote the dependent variables, and the rows denote the independent variables. The table reports estimates from twelve models. The control variables include: child sex, birth month, whether or not at least one of the parents is a non-Western immigrant, parental education, family income, an indicator for late parental consent, and randomization-block indicators. Bootstrap (clustered) standard errors in parentheses (1,000 repetitions). All models are estimated with full-information maximum likelihood (701 obs.). Online Appendix A.1 provides an overview of missingness.

Table A.2.14. Treatment Effects and Decomposition using Arithmetic Averages

	Mathematics		Language		EFs	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Post-Intervention and Follow-Up Treatment Effects						
Post-Intervention	0.167	0.152	0.015	0.022	0.108	0.113
	(0.091)	(0.089)	(0.078)	(0.077)	(0.060)	(0.058)
Follow-Up	0.227	0.221	0.010	0.051	0.063	0.071
	(0.060)	(0.058)	(0.089)	(0.085)	(0.053)	(0.055)
Panel B: Decomposition of Total Treatment Effect						
EFs	0.031	0.033	0.019	0.016	0.057	0.059
	(0.018)	(0.017)	(0.010)	(0.008)	(0.036)	(0.033)
	[13.7%]	[14.9%]	[27.9%]	[30.8%]	[50.4%]	[56.2%]
Mathematics	0.070	0.062	0.010	0.013	0.031	0.028
	(0.039)	(0.036)	(0.005)	(0.007)	(0.017)	(0.017)
	[30.8%]	[28.1%]	[14.7%]	[25.0%]	[27.4%]	[26.7%]
Language	0.001	0.002	0.010	0.016	0.000	0.001
	(0.005)	(0.006)	(0.043)	(0.039)	(0.001)	(0.002)
	[0.4%]	[0.9%]	[14.7%]	[30.8%]	[0.0%]	[1.0%]
Unmeasured Variables	0.125	0.124	−0.029	0.007	−0.025	−0.017
	(0.080)	(0.076)	(0.083)	(0.082)	(0.061)	(0.063)
	[55.1%]	[56.1%]	[42.6%]	[13.5%]	[22.1%]	[16.2%]
Control Variables	No	Yes	No	Yes	No	Yes

Notes. This table reports the parameter estimates for the treatment effect decomposition. Panel A reports the post-intervention treatment effect (*i.e.*, differences between treated and non-treated children post-intervention) and the follow-up treatment effect (*i.e.*, differences between treated and non-treated children at the follow-up). Panel B decomposes the follow-up treatment effect in measured and unmeasured variables. The columns denote the dependent variables, and the rows denote the independent variables. We include the following control variables: child sex, birth month, whether or not at least one of the parents is a non-Western immigrant, parental education, family income, an indicator for late parental consent, and randomization-block indicators. Standard errors (in parentheses) are computed using a wild residual (clustered) bootstrap procedure (1,000 bootstrap samples). In brackets, we report the relative contributions. We compute these relative contributions by taking the absolute value of the estimate divided by the sum of absolute values of all estimates multiplied by 100. All models are estimated with full-information maximum likelihood (total observations 701). Online Appendix A.1 provides an overview of missingness.

A.3. Further Notes on Croon's Correction Method

We provide further notes on Croon's correction method (Croon, 2002) in this appendix. Let,

$$\varphi_{h,t} = (\boldsymbol{\lambda}'_{h,t}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\zeta},h,t}\boldsymbol{\lambda}_{h,t})^{-1}\boldsymbol{\lambda}'_{h,t}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\zeta},h,t},$$

denote the (Bartlett) factor scoring matrix, where $\boldsymbol{\lambda}_{h,t}$ is a L_h-dimensional vector with factor loadings, and $\boldsymbol{\Sigma}_{\zeta,h,t}$ is a (L_h, L_h)-dimensional matrix with unique factor variances. We can then write the factor scores as,

$$\tilde{\mathrm{S}}_{h,i,t} = \varphi_{h,t}(\mathbf{M}_{h,i,t} - \boldsymbol{\mu}_{h,t}).$$

Consider data in which we want to estimate the relationship between skill h and skill k, $h \neq r$, using the predicted factor scores,

$$\tilde{\mathrm{S}}_{h,i} = \beta_0 + \beta_1\tilde{\mathrm{S}}_{r,i} + \varepsilon_{h,i}.$$

We can write,

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(\tilde{\mathrm{S}}_{h,i}, \tilde{\mathrm{S}}_{r,i})}{\mathrm{Var}(\tilde{\mathrm{S}}_{r,i})}.$$

We can write the covariance, $\mathrm{Cov}(\tilde{\mathrm{S}}_{h,i}, \tilde{\mathrm{S}}_{r,i})$, as,

$$\begin{aligned}
\mathrm{Cov}(\tilde{\mathrm{S}}_{h,i}, \tilde{\mathrm{S}}_{r,i}) &= \mathbf{Cov}(\varphi_h\mathbf{M}_h, \varphi_k\mathbf{M}_r) \\
&= \varphi_h\mathbf{Cov}(\mathbf{M}_h, \mathbf{M}_r)\varphi'_r \\
&= \varphi_h\mathbf{Cov}(\boldsymbol{\lambda}_h\mathrm{S}_{h,i} + \boldsymbol{\zeta}_{h,i}, \boldsymbol{\lambda}_r\mathrm{S}_{r,i} + \boldsymbol{\zeta}_{r,i})\varphi'_r \\
&= \varphi_h\boldsymbol{\lambda}_h\mathbf{Cov}(\mathrm{S}_{h,i} + \boldsymbol{\zeta}_{h,i}, \mathrm{S}_{r,i} + \boldsymbol{\zeta}_{r,i})\boldsymbol{\lambda}'_r\varphi'_r \\
&= \varphi_h\boldsymbol{\lambda}_h\mathrm{Cov}(\mathrm{S}_{h,i}, \mathrm{S}_{r,i})\boldsymbol{\lambda}'_r\varphi'_r.
\end{aligned}$$

We can write the variance as follows,

$$\begin{aligned}
\mathrm{Var}(\tilde{\mathrm{S}}_{r,i}) &= \mathbf{Var}(\varphi_r\mathbf{M}_r) \\
&= \varphi_r\mathbf{Var}(\mathbf{M}_r)\varphi'_r \\
&= \varphi_r\mathbf{Var}(\boldsymbol{\lambda}_r\mathrm{S}_{r,i} + \boldsymbol{\zeta}_{r,i})\varphi'_r \\
&= \varphi_r\boldsymbol{\lambda}_r(\mathrm{Var}(\mathrm{S}_{r,i}) + \mathbf{Var}(\boldsymbol{\zeta}_{r,i}))\boldsymbol{\lambda}'_r\varphi'_r.
\end{aligned}$$

It follow, then, that,

$$\begin{aligned}
\hat{\beta}_1 = \frac{\mathrm{Cov}(\tilde{\mathrm{S}}_{h,i}, \tilde{\mathrm{S}}_{r,i})}{\mathrm{Var}(\tilde{\mathrm{S}}_{r,i})} &= \frac{\varphi_h\boldsymbol{\lambda}_h\mathrm{Cov}(\mathrm{S}_{h,i}, \mathrm{S}_{r,i})\boldsymbol{\lambda}'_k\varphi'_r}{\varphi_r\boldsymbol{\lambda}_r(\mathrm{Var}(\mathrm{S}_{r,i}) + \mathbf{Var}(\boldsymbol{\zeta}_{r,i}))\boldsymbol{\lambda}'_r\varphi'_r} \\
&= \text{Attenuation Factor} \cdot \frac{\mathrm{Cov}(\mathrm{S}_{h,i}, \mathrm{S}_{r,i})}{\mathrm{Var}(\mathrm{S}_{r,i})},
\end{aligned}$$

where,

$$\text{Attenuation Factor} = \frac{\varphi_h\boldsymbol{\lambda}_h\mathrm{Var}(\mathrm{S}_{r,i})\boldsymbol{\lambda}'_r\varphi'_r}{\varphi_r\boldsymbol{\lambda}_k(\mathrm{Var}(\mathrm{S}_{r,i}) + \mathbf{Var}(\boldsymbol{\zeta}_{r,i}))\boldsymbol{\lambda}'_r\varphi'_r}$$

Since all terms in the attenuation factor are obtained from estimating the measurement model, we can divide $\hat{\beta}_1$ by the attenuation factor to obtain the corrected parameter estimates.