Causes and Consequences of Educational Attainment and Household Decisions - Six Essays in Applied Microeconomics

INAUGURAL-DISSERTATION

zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaft
doctor rerum politicarum
(Dr. rer. pol.)

am Fachbereich Wirtschaftswissenschaft
der Freien Universität Berlin

Freie Universität Berlin

vorgelegt von
Jan Berkes (M.Sc.)

Berlin, 2021

Erstgutachterin:	Prof. Dr. C. Katharina Spieß
	Freie Universität Berlin und DIW Berlin
Zweitgutachter:	Prof. Dr. Lukas Menkhoff
	Humboldt-Universität zu Berlin und DIW Berlin

Disputation: 26. Oktober 2021

Acknowledgments

Firstly, I would like to thank C. Katharina Spiess for her tireless support throughout the execution of this project. Your valuable advice, guidance and flexibility during the entire process is much appreciated. I would also like to thank my second supervisor Lukas Menkhoff for the exchange. I am grateful to my coauthors Frauke Peter, C. Katharina Spiess, Felix Weinhardt, Jonas Jessen, Adrien Bouguen, Deon Filmer, Abbie Raikes and Tsuyoshi Fukao for the fruitful cooperation and inspiring exchanges during fieldwork, analysis and writing stages of this project.

I wrote this dissertation at the Education and Family Department of the DIW Berlin. As a member of the DIW Graduate center and the German Academic Scholarship Foundation I gratefully received financial support which allowed me to pursue this project, including the two stays abroad at the World Bank and University of California, Berkeley. Chapter 2 was partially funded by the German Science Foundation (Geschäftszeichen SP 1091/2-1). Chapters 4-7 were partially written as a short-term consultant at the World Bank and data collections were partially funded by the World Bank's Strategic Impact Evaluation Fund. In addition, I am very grateful to the dedicated team of field researchers for providing excellent data, and all study participants who contributed to this work.

A big thank you goes to all BiFa-colleagues and my fellow GC-cohort for great exchanges during countless lunches, coffee breaks, video conferences and outside activities. I am particularly grateful to Jonas Jessen for the great atmosphere that we shared in the office and beyond. Last but not least, I thank my friends, family and, above all, Olga, for the wholehearted support.

Berlin, May 2021

Contents

Erklärung zu Ko-Autorenschaften

Diese Dissertation besteht aus sechs (Arbeits-)Papieren, von denen fünf in Zusammenarbeit mit Koautoren entstanden sind:

- Jan Berkes, Frauke Peter, C. Katharina Spiess und Felix Weinhardt: ***"Information Provision and Postgraduate Studies"***
- Jan Berkes, Jonas Jessen und Felix Weinhardt: ***"Gender Inequality and the Relation between Market and Household Work"***
- Jan Berkes, Adrien Bouguen, Deon Filmer, Tsuyoshi Fukao: ***"Improving Preschool Provision and Encouraging Demand: Heterogeneous Impacts of a Large-Scale Program"***
- Jan Berkes und Adrien Bouguen: ***"Estimating Preschool Impacts when Counterfactual Enrollment Varies: Bounds and Conditional LATE"***
- Jan Berkes: ***"The Effects of Quality on Child Development in Cambodian Preschools"***
- Jan Berkes, Abbie Raikes, Adrien Bouguen und Deon Filmer: ***"Joint Roles of Parenting and Nutritional Status for Child Development: Evidence from Rural Cambodia"***

Liste der Vorveröffentlichungen

- **Publikationen in referierten Fachzeitschriften**
  - **Vorpublikation von Kapitel 7:**
    - Berkes, J., Raikes, A., Bouguen, A., Filmer, D. (2019). Joint roles of parenting and nutritional status for child development: Evidence from rural Cambodia. *Developmental Science, 22(5), e12874*, https://doi.org/10.1111/desc.12874
- **Beiträge zur Politikberatung**
  - **Teilweise basierend auf Kapitel 2 ist erschienen:**
    - Berkes, J., Peter, F., Spieß, C. K., Weinhardt, F. (2020). Masterstudium ja oder nein? Gezielte Online-Informationen beeinflussen weitere Studienpläne. *DIW Wochenbericht, 87(21), 359-367.*
  - **Teilweise basierend auf Kapitel 4 ist erschienen:**
    - Berkes, J, Bouguen, A., Filmer, D., Fukao, T. (2020). Cambodia : Can New and Improved Preschools Improve Child Development in Rural Communities? *From Evidence to Policy*; World Bank, Washington, DC.
- **Arbeitspapiere**
  - **Vorpublikation von Kapitel 2:**
    - Berkes, J., Peter, F., Spiess, C. K., Weinhardt, F. (2019). Information Provision and Postgraduate Studies. *IZA Discussion Papers (No. 12735).*
  - **Vorpublikation von Kapitel 4:**
    - Berkes, J., Bouguen, A., Filmer, D., Fukao, T. (2019). Improving Preschool Provision and Encouraging Demand: Heterogeneous Impacts of a Large-Scale Program. *Policy Research Working Paper (No. 9070).* World Bank, Washington, DC.
  - **Vorpublikation von Kapitel 5:**

- * Berkes, J., Bouguen, A. (2019). Estimating Preschool Impacts when Counterfactual Enrollment Varies: Bounds, Conditional LATE and Machine Learning. *Working Paper Series No. WPS-088.* Center for Effective Global Action. University of California, Berkeley.

– **Vorpublikation von Kapitel 7:**

- * Berkes, J., Raikes, A., Bouguen, A., Filmer, D. (2019). Joint Effects of Parenting and Nutrition Status on Child Development: Evidence from Rural Cambodia. *Policy Research Working Paper (No. 8529).* World Bank, Washington, DC.

Rechtliche Erklärung

Erklärung gem. §4 Abs. 2 (Promotionsordnung)

Hiermit erkläre ich, dass ich mich noch keinem Promotionsverfahren unterzogen oder um Zulassung zu einem solchen beworben habe, und die Dissertation in der gleichen oder einer anderen Fassung bzw. Überarbeitung einer anderen Fakultät, einem Prüfungsausschuss oder einem Fachvertreter an einer anderen Hochschule nicht bereits zur Überprüfung vorgelegen hat.

Berlin, Mai 2021
Jan Berkes

Erklärung gem. §10 Abs. 3 (Promotionsordnung)

Hiermit erkläre ich, dass ich für die Dissertation folgende Hilfsmittel und Hilfen verwendet habe. Auf dieser Grundlage habe ich die Arbeit selbstständig verfasst.

- Software:
  - Stata Versionen 13, 14 und 15
  - Microsoft Office 2016
  - LaTeX mit Overleaf
- Literatur: siehe Literaturverzeichnis

Berlin, Mai 2021
Jan Berkes

Zusammenfassung

Diese Dissertation besteht aus sechs Kapiteln, die sich mit der ökonomischen Analyse von Bildungs- und Haushaltsentscheidungen befassen. Diesen Kapiteln geht eine Einleitung voraus, in denen gemeinsame und komplementäre Beiträge der Forschungsarbeiten dieser Dissertation zur Literatur dargestellt und die einzelnen Forschungsfragen motiviert werden. Im Schlussteil werden Limitationen der Forschungsarbeiten, offene Forschungsfragen und Handlungsempfehlungen für die Politik diskutiert.

Kapitel 2 befasst sich mit den Effekten von Informationen auf die erwarteten monetäre und nicht-monetären Erträge eines Masterstudiums sowie der beabsichtigten und tatsächlichen Aufnahme eines Masterstudiums. Hierfür wurden im Rahmen eines Experiments Studierenden gezielt Online-Informationen bereitgestellt. Sechs Monate nach der Informationsintervention lassen sich kausale Effekte auf die erwarteten Erträge beobachten. Die Effekte auf erwartete Erträge sowie die Effekte auf die beabsichtigte und tatsächliche Aufnahme eines Masterstudiums unterscheiden sich nach Geschlecht und akademischem Hintergrund. Insbesondere Männer schätzen infolge der Informationsintervention ihre Perspektiven ohne Masterstudium besser ein als zuvor. Ein Masterstudium verliert für sie an Attraktivität. Mithilfe weiterer Erhebungen wird gezeigt wie sich die Intervention auf die tatsächliche Aufnahme eines Masterstudiums auswirkt. Zusammengefasst zeigt dieses Kapitel die Relevanz von Informationen zu monetären und nicht-monetären Erträgen eines Masterstudiums für die Studienentscheidung auf.

Kapitel 3 zeigt mithilfe von Zeitbudgeterhebungen aus der DDR und der wiedervereinigten Bundesrepublik Deutschland auf, dass Frauen in der DDR, und später in den ostdeutschen Bundesländern, mehr Zeit im Arbeitsmarkt und weniger Zeit mit Hausarbeit als westdeutsche Frauen verbringen. Bei der differenzierten Betrachtung dieser Unterschiede wird jedoch deutlich, dass keine Differenzen in der Hausarbeit zwischen ost- und westdeutschen Frauen bestehen, wenn Frauen mit gleichem zeitlichen Beschäftigungsumfang im Arbeitsmarkt verglichen werden. Da Männer eine höhere Arbeitsbelastung der Frauen am Arbeitsmarkt nicht durch die Übernahme von mehr Hausarbeit kompensieren, sind berufstätige Frauen in der DDR und BRD doppelt belastet. Das Kapitel schließt mit einer Diskussion über die Folgen von Geschlech-

ternormen und möglichen Nebeneffekten von politischen Maßnahmen zur Erhöhung der Frauenerwerbsquote ab.

Kapitel 4 untersucht die Effekte eines umfangreichen Programms zur Eröffnung neuer Vorschulen in Kambodscha. Der Bau neuer Vorschulen im Jahr 2016 wurde mit zwei nachfrageseitigen Interventionen zur Stimulierung der Nachfrage nach Vorschulen in einer randomisierten kontrollierten Studie kombiniert. Während Kinder durch den verbesserten Zugang zu Vorschulen häufiger eine Vorschule besuchen, hatten die nachfrageseitigen Interventionen keinen zusätzlichen Effekt. Ein Jahr nach Beginn der Intervention wurden kleine Effekte auf die kognitive Entwicklung (0.04 Standardabweichungen) und sozio-emotionale Entwicklung (0.09 Standardabweichungen) der Kinder gemessen. Mittelfristige Effekte werden für Kinder des oberen Wohlstandsquartils zwei Jahre nach Beginn der Intervention gemessen. Dieses Kapitel zeigt zudem auf, dass das untersuchte Programm zu großen Verbesserungen in der strukturellen Qualität der vorhandenen Vorschulen geführt hat, jedoch nur geringe Effekte auf deren Prozessqualität hatte. Die Ergebnisse legen nahe, dass in Ländern mit niedrigem Einkommen, Kinder aus verhältnismäßig gutsituierten Haushalten stärker von den Vorschulangeboten profitieren. Ein Grund sind komplementäre Effekte des Vorschulbesuchs und bessere kognitive Stimulation und emotionale Unterstützung der Eltern. Die Ergebnisse zeigen auch auf, dass weitere Verbesserungen in der Prozessqualität von Vorschulen notwendig sind um die Entwicklung benachteiligter Kinder zu fördern.

Kapitel 5 untersucht die Effekte des gleichen Vorschulprogramms in Kambodscha mit besonderem Fokus auf die Tatsache, dass die neu geschaffenen Vorschulen mit bestehenden Vorschulangeboten von geringerer Qualität sowie familiärer Betreuung konkurrieren. In diesem Kontext wird aufgezeigt, dass sich die gemessenen Effekte des Vorschulprogramms auf die frühkindliche Entwicklung danach unterscheiden, ob ein Kind ohne die neuen Vorschulen eine andere Vorschule besucht hätte oder in familiärer Betreuung geblieben wäre. Mithilfe der Daten der randomisierten kontrollierten Studie werden verschiedene empirische Schätzmethoden angewandt, um den Effekt auf Kinder, die ohne die neuen Vorschulen in elterlicher Betreuung geblieben wären zu identifizieren. Es wird diskutiert, weshalb die Messung dieses Effekts von zentraler Bedeutung für die Wirkungsmessung von Vorschulangeboten ist und besser als konventionelle Schätzer (z.B. *reduced-form estimates*) für Vergleiche zwischen verschiedenen Studien und Kontexten geeignet ist. Es wird zunächst gezeigt, dass der Effekt auf die kognitiven Fähigkeiten von Kindern in kontrafaktischer elterlicher Betreuung, unter realistischen Annahmen, auf zwischen 0.14 und 0.39 Standardabweichungen eingegrenzt werden kann. Anschließend wird eine Instrumentvariablenschätzung angewandt, um die Effektgröße näher einzugrenzen. Die Ergebnisse zeigen, dass der Effekt in etwa 0.2 Standardabweichungen beträgt. Der Effekt auf Kinder, die ohne das neue Programm eine andere Vorschule besucht hätten, ist klein und nicht statistisch signifikant.

Kapitel 6 untersucht den Zusammenhang verschiedener Qualitätsindikatoren kambodschanischer Vorschulen mit der Entwicklung frühkindlicher kognitiver und sozio-emotionaler Fähigkeiten auf Basis eines value-added Modells. Diese Studie basiert auf einer Stichprobe von 327 Vorschulen und etwa 3,000 Kindern in Kambodscha. Die Qualitätsindikatoren basieren auf detaillierten Unterrichtsbeobachtungen und beinhalten mehrere Dimensionen von Struktur- und Prozessqualität, wie zum Beispiel die pädagogischen Ansätze der Lehrkraft zur Vermittlung des Unterrichtsstoffes und die Berücksichtigung individueller Bedürfnisse von Kindern. Mithilfe einer Faktorenanalyse wird zunächst aufgezeigt, dass häufig verwendete Qualitätsindikatoren in weitere latente Variablen unterteilt werden sollten. Jedoch lässt sich auch mit detaillierten Qualitätsindikatoren kein eindeutiger Zusammenhang zwischen der Lehrqualität und frühkindlicher kognitiver Entwicklung nachweisen. Auch mit der sozio-emotionalen Entwicklung bestehen nur schwache Zusammenhänge. Jedoch zeigt sich deutlich, dass Kinder in Vorschulen mit besonders geringer Strukturqualität, aufgrund eines Mangels an grundlegender Ausstattung, geringere kognitive und sozio-emotionale Fähigkeiten haben. Die Ergebnisse legen nahe, dass im untersuchten Kontext zunächst universelle Qualitätsstandards für alle Vorschulen erreicht werden sollten, bevor eine gezielte Förderung einzelner Schulen angestrebt wird. Der fehlende Zusammenhang zwischen der beobachteten Prozessqualität und frühkindlicher Entwicklung kann erklären, weshalb die Effekte von Lehrerfortbildungen auf Lernerfolge oftmals geringfügig sind und zeigen die Schwierigkeiten bei der Erfassung effektiver Bildungs- und Betreuungsmethoden auf.

Kapitel 7 fasst zunächst anhand bestehender Literatur zusammen, wie die Ernährungssituation von Kindern und deren häusliches Umfeld, einschließlich des Ausmaßes, in dem sie kognitive Stimulation und emotionale Unterstützung von den Eltern erhalten, eine tiefgreifende Rolle für die frühkindlichen Entwicklung spielen. Dieses Kapitel ergänzt die bestehende Literatur mit Evidenz dazu, welche komplementären Einflüsse die frühkindliche Erziehung und Ernährungssituation auf die Entwicklung bei kambodschanischen Kindern im Vorschulalter haben. Hierzu werden Paneldaten aus den Jahren 2016-17 zu elterlicher Erziehung, dem Ernährungsstatus und der frühkindlichen Entwicklung (exekutive Funktionen, sprachliche Entwicklung, frühe Rechenfertigkeiten und sozio-emotionale Entwicklung) von 6,508 kambodschanischen Kindern im Alter von 3 bis 5 Jahren genutzt. Die Ergebnisse zeigen, dass Ungleichheiten in der frühkindlichen Entwicklung, die mit dem sozio-ökonomischen familiären Hintergrund verbunden sind, bereits im Alter von 3 Jahren deutlich zu beobachten sind und sich bei Kindern im Alter

von 4 und 5 Jahren weiter verstärken. Mit Hilfe eines hierarchischen Regressionsmodells lässt sich ein signifikanter Anteil dieser Ungleichheiten durch Unterschiede in der elterlichen Erziehung und dem frühen Ernährungsstatus, gemessen durch Wachstumsstörungen, erklären. Für besser gebildete Eltern lässt sich mehr kognitive Stimulation und emotionale Unterstützung der Kinder beobachten. Allerdings ist die positive Assoziation zwischen elterlicher Erziehung und frühkindlichen kognitiven Fähigkeiten bei Kindern ohne Wachstumsstörungen um 35%-54% stärker und die elterlichen Aktivitäten erklären nur etwa 8%-14% des kognitiven Entwicklungsunterschieds zwischen dem untersten und obersten Wohlstandsquintil. Die Ergebnisse unterstreichen den Bedarf an zusätzlicher Forschung, die die Wechselwirkungen zwischen Einflussfaktoren, die den Wohlstand der Familie mit der Entwicklung des Kindes verbinden, aufzeigt.

Abstract

This dissertation comprises six chapters which contribute to the economic analysis of education and household choice. These chapters are preceded by an introduction which outlines joint and complementary contributions of the chapters and the motivation for the individual research questions. In the concluding chapter, limitations, open research questions and policy implications of the dissertation are discussed.

Chapter 2 experimentally examines effects of information provision on beliefs about pecuniary and non-pecuniary returns of postgraduate education, enrollment intentions and realized enrollment. It is shown that the treatment causally affects beliefs measured six months after treatment. The effects on beliefs differ by gender and academic background, and stated enrollment intentions change accordingly: in particular males significantly adjust their beliefs and intentions to undertake postgraduate studies downward. This is driven by males upward-adjusting earnings expectations with a first degree only. Students are followed further to provide evidence on actual enrollment one and two years after treatment. Taken together, this chapter highlights the relevance of information provision on pecuniary and non-pecuniary labor market returns for postgraduate study decisions.

Chapter 3 uses novel time-use data from the GDR and reunified Germany to show that women in the GDR, and later in East Germany, spend more hours on the labour market and less time doing housework tasks, compared to their West German counterparts. Decomposing these gaps, it is shown that gender gaps between the West and East are statistically identical once individual time-constraints are accounted for. In the absence of partner's reactions and across all regimes studied, working women face the second shift. The chapter concludes with a discussion of implications regarding the nature of gender norms and effects of labour market policy targeted at gender gaps.

Chapter 4 examines impacts of a large-scale program that increased the supply and quality of community preschools in Cambodia. Construction of preschools was paired with two demand-side interventions designed to stimulate additional enrollment. The construction caused an increase in enrollment but demand-side interventions did not. After one year, small impacts on cognitive (0.04 standard deviations) and socio-

emotional development (0.09 standard deviations) are observed. Persistent effects on children from the wealthiest quartile are observed two years after the program started. The chapter provides evidence that the program had large impacts on the quality of preschool infrastructure but only limited impacts on the quality of educational processes. The results indicate that, in low-income countries, less disadvantaged families are better able to take advantage of the preschools by providing a home environment that complements the education received at preschool. The results also suggest that improving the quality of educational processes might be needed to foster the development of disadvantaged children.

Chapter 5 studies the impacts of the same large preschool construction program with a focus on the fact that the newly built preschools compete with lower quality existing preschools as well as home care. In this context, the chapter highlights that impacts are likely to differ between those who would have been enrolled in a preexisting preschool and those who would have been in home care, with expected larger gains among the latter. Using short-term data from an experiment conducted in Cambodia, several empirical techniques to isolate the impact on children who would have stayed at home had they not been enrolled in the newly built preschools are implemented. It is argued that the impact on these children is a central parameter in the preschool literature and is more comparable across studies and contexts than traditional reduced form estimates. First a bounding approach is implemented to show that, under reasonable assumptions, the effect on children who would have stayed at home absent the program is high and significant (between 0.14 and 0.39 standard deviations). Then an instrumental variable approach is implemented to pinpoint the effect on these children. The results include consistent evidence that the impact on these children is around 0.2 standard deviations on a child development aggregate score while the effect on children who would have enrolled in a preexisting preschool (absent the newly built school) is small and insignificant.

Chapter 6 examines the value-added of quality in Cambodian preschools for cognitive and socio-emotional development in early childhood. The study is conducted with a sample of 327 preschools and about 3,000 children in rural Cambodia. The preschool quality measures are based on detailed classroom observations, capturing multiple aspects of structural and process quality, such as the teachers' pedagogical approaches to cover the curriculum and their responsiveness to children's needs. Using a factor analysis, I first show that commonly measured categories of quality are multidimensional and difficult to summarize in few indices. Teacher cognition and training positively predict teaching practices. Despite the detail of the measures, no clear link between teaching practises and child development is found for cognitive skills and only a weak link is found for socio-emotional development. Yet, children in preschools with a particularly low structural quality due to a lack of basic equipment perform significantly worse. Policywise, the results suggest that achieving a universal quality standard for all preschool facilities, instead of following a targeted approach, should be a priority in the studied context. The missing link between observed process quality and child outcomes can also explain why effects of teacher training are generally found to be small and confirm that effective teaching is difficult to measure.

Chapter 7 first summarizes how substantial work has demonstrated that early nutrition and home environments, including the degree to which children receive cognitive stimulation and emotional support from parents, play a profound role in influencing early childhood development. Yet, less work has documented the joint influences of parenting and nutritional status on child development among children in the preschool years living in low-income countries. Using panel data from 2016–17 on the parenting, nutritional status, and early developmental outcomes (executive function, language, early numeracy, and socioemotional problems) of 6,508 Cambodian children ages 3 to 5 years, the findings demonstrate that inequities in early development associated with family wealth are evident at age 3 and increase among children ages 4 and 5 years. Using hierarchical regression analysis, a significant share of these inequalities is explained by differences in parenting and early nutritional status, measured by stunting. Better-educated parents engage in more stimulating and supportive parenting practices. However, the positive association between parenting and language and early numeracy outcomes is 35%-54% stronger for non-stunted children, and parental activities explain only about 8%-14% of the cognitive gap between the lowest and highest wealth quintiles. The results highlight the need for additional research outlining interactions between environmental factors that link family wealth and child development.

List of Tables

List of Figures

CHAPTER 1

Introduction and Overview

1.1 Motivation

Investments into education and training as a source of utility are at the core of classical human capital theory (Mincer, 1958; Schultz, 1961; Becker, 1962). The positive relationship between education and labor income, alongside other welfare measure such as health or life satistfaction, has been documented by countless empirical economic studies (e.g. Psacharopoulos, 1985; Blundell et al., 1999; Conti et al., 2010; Gunderson and Oreopolous, 2020; Patrinos and Psacharopoulos, 2020). Costs and benefits of investments into human capital occur across several life stages and are often not directly observable to individuals. As a consequence, even rational decision makers might not be able to maximize their lifetime utility through optimal investments into their human capital, e.g. due to financial constraints or information asymmetries (e.g. Delavande and Zafar, 2019).[1] A big part of research in education economics – and this dissertation, as one small puzzle piece of many – aims to deepen the understanding around education choices to support individuals and policy makers in making better choices for humans' economic and personal development. In six essays, this dissertation focuses on open questions that can be placed in four areas: the home environment, early childhood education and care (ECEC) services, tertiary education and the labor market.

Human skill formation processes are characterized by "skills begetting skills", i.e. the existence of dynamic complementarities between skills obtained at some point in life and the productivity of later investments. In addition, skills itself are self-productive, i.e. a skill learned at some point in time enables to further improve this and other skills (Cunha and Heckman, 2007). The plurality of skills and their interconnectedness over time is relevant for learning processes at each point in human life, but is particularly

[1] In addition, the field of behavioral economics has emerged and offers various explanations for seemingly irrational long-run education decisions (for an overview, see Lavecchia et al., 2016).

striking in early childhood development. Researchers in the fields of economics, neurobiology and developmental psychology independently come to the conclusion that the early years in life are crucial for the development of cognitive and social skills (Knudsen et al., 2006; Heckman, 2006). The human brain is particularly receptive at young age and later remediation of missed development opportunities is almost impossible (Shonkoff et al., 2000). In addition, early skills have a particularly long time to pay off and beget further learning, which makes early childhood experiences particularly influential for education, labour market and health outcomes (e.g. Heckman and Masterov, 2007; Currie and Almond, 2011). This justifies early interventions from an equity perspective, as children are born into very different home environments and socio-economic gradients in child development are visible at age 5 and earlier, as shown by many studies (e.g. Bradbury et al., 2015; Rubio-Codina et al., 2015). This includes **Chapter 7** of this dissertation, which sheds additional light on *associations between the home environment and child development* in a low-income county context. Early childhood interventions are also justified from an economic, efficiency perspective, as they can be highly cost-effective (Heckman, 2006).

One promising instrument to foster early development and mitigate the role of socio-economic differences in the home environment are ECEC services, such as preschool education. In particular, studies from the United States, which demonstrate strong positive effects of early childhood education for socially disadvantaged children, suggest an equalizing effect. Elango et al. (2015) describe in a comprehensive review the effects of compensatory preschool programs such as the Perry Preschool Project or Head Start (see also Currie and Thomas, 1995; Kline and Walters, 2016). This makes ECEC services a particularly promising instrument for low-income countries, where a particularly large share is at risk of not fulfilling their developmental potential (Black et al., 2017). Yet, the existing evidence on such programs from low-income countries is scarce and with contrasting findings. While non- and quasi-experimental study designs generally point to positive effects of preschool attendance (Berlinski et al., 2008, 2009; Engle et al., 2011; Rao et al., 2014), experimental evidence on large-scale ECEC services often find insignificant or even negative effects of preschool education (Brinkman et al., 2017; Bouguen et al., 2018a; Blimpo et al., 2019; Bernal et al., 2019).[2] More evidence is needed to identify feasible program designs that meet the promises of ECEC services. This dissertation adds evidence to this research field by studying *impacts of a large-scale public preschool program in Cambodia on child development* (**Chapters 4 and 5**).

[2] Also for high income-countries, large and long-lasting results of preschool education (Belfield et al., 2006; Carneiro and Ginja, 2014) coexists with insignificant (Puma et al., 2012) and even negative results (Baker et al., 2008, 2019).

Studies on ECEC services from high-income countries show that the effect of ECEC depends on multiple factors, such as the quality of the program (Blau and Currie, 2006; Datta Gupta and Simonsen, 2016) and continuous follow-up by later investments (Cunha and Heckman, 2007). Several factors can be at play when ECEC services do not meet their promises, e.g. by not improving cognitive or non-cognitive skills of children. The high quality child outcomes produced by targeted programs described in the literature (e.g. Grantham-McGregor et al., 1991; Elango et al., 2015) are often based on very cost-intensive programs of much higher quality than provided in studies with insignificant or even negative findings (see e.g. Bouguen et al. (2018a), Blimpo et al. (2019), or for a summary Spiess (2017)). Recent experimental evidence has shown that process quality of ECEC services is particularly predictive of child development (Andrew et al., 2019). Further research is required to identify properties of feasible, affordable and effective large-scale programs. This process includes the *identification of important quality dimensions for effective ECEC services* and the *monitoring of existing programs to ensure adherence of existing quality standards*, which, for a low-income country context, is a contribution of **Chapter 6**.

Previous research has shown, that the provision of information about the return to education can affect education choices in high-income countries (e.g. Bettinger et al., 2012; Oreopoulos and Dunn, 2013; Bettinger and Baker, 2014; Wiswall and Zafar, 2015; Peter and Zambre, 2017; Peter et al., 2018; Kerr et al., 2020) and low-income countries (Nguyen, 2008; Jensen, 2010). Yet, the effectiveness of information provision varies greatly between studies with most of the evidence looking at secondary school and initial college enrollment. More evidence is needed to identify other areas with binding information constraints and successful approaches to mitigate those constraints. This dissertation studies *the role of information for enrollment into ECEC services versus home care* (**Chapter 4**) in a low-income country context, Cambodia, and *the role of information for attaining postgraduate education versus entering the labor market* with an undergraduate degree only (**Chapter 2**) in a high-income country context, Germany.

Related to the role of perceived returns is also the role of social norms, as another set of beliefs which can change through learning processes and affect education and household choices (Bicchieri et al., 2018). For example, how members of households, such as wife and husband, allocate their (time) resources has been described as an outcome of an intra-household bargaining process which can be heavily affected by social (gender) norms (e.g. Bertrand et al., 2015). Gender inequality in non-market work, such as child care or housework, and market work can be amplified through strong gender role attitudes and mitigated through effective labor market and ECEC policies (see e.g. Olivetti and Petrongolo, 2017). Associations between these factors, which are highly relevant to gauge the full effects of policies in one of these areas, are not yet fully understood. **Chapter 3** studies *associations between market work and non-market* work among couples in the GDR, as well as East and West Germany. This contributes to the understanding of norms and their evolution under different political regimes. It also highlights potential detrimental consequences of labor market policies that target an increase in female employment.

1.2 Summary and Overview

This dissertation consists of six empirical research papers. **Chapters 2 and 3** cover independent questions of education and household choice. While all chapters are self-contained, **Chapters 4 to 7** cover distinct questions using data from the same survey and study sample. The interdisciplinary nature of early childhood development science is also reflected in this dissertation, as **Chapter 7** takes a more psychological perspective on aspects of parenting, nutrition and early development. The chapters are summarized in the following. The individual main research questions, main findings and methodological approaches are also briefly summarized in Tables 1.1 and 1.2.

Chapter 2 adds to the evidence on the role of information for education choice. Based on quasi-experimental and experimental studies, it is well established that information, costs and beliefs all play important roles in explaining college decisions. Yet, so far no experimental evidence existed that looks at the role of information for postgraduate study decisions – despite their relevance due to documented substantial returns to postgraduate education (Lindley and Machin, 2016; Altonji et al., 2016). Using a randomized experiment, the effects of information provision on beliefs about pecuniary and non-pecuniary returns of postgraduate education are explored. In addition, the effects of information provision on enrollment intentions and realized enrollment are measured. The study is based on a panel of German students (PostGrad-Best Up) which are close to graduating with an undergraduate degree. The treatment consists of information about pecuniary and non-pecuniary returns of postgraduate versus undergraduate degrees in the labor market, based on data of existing employees in Germany. It is shown that the treatment causally affects beliefs measured six months after the treatment. The effects on beliefs differ by gender and academic background, and stated enrollment intentions change accordingly: in particular males significantly adjust their beliefs and intentions to undertake postgraduate studies downward. This is driven by males upward-adjusting earnings expectations with a bachelor's degree only. Students are followed further to provide evidence on actual enrollment one and two years after treatment. Taken together, this chapter highlights the relevance of information provision on pecuniary and non-pecuniary labor market returns for postgraduate study decisions. The significant effects on males document that information frictions exist even for students already enrolled in an undergraduate degree.

Chapter 3 closer investigates a key prediction of canonical theories of within-household allocation, which is that the time an individual has to spend doing household tasks depends on the labour market options of each partner. To examine if this is the case, this chapter makes use of novel time-use data from the GDR and reunified Germany. This setting is of particular interest because full-time employment was the government-enforced norm in the GDR, in stark contrast to West Germany. It is first documented descriptively that women in the GDR, and later in East Germany, indeed spend more hours in the labour market and less time doing housework tasks, compared to their West German counterparts. It is then shown that this inverse relation between time spend on market and non-market work emerges as a mechanical product of individual time-constraints alone. No evidence for impacts of the partner's labour supply is found. Decomposing persistent gaps between East and West German households after reunification using a Gelbach decomposition (Gelbach, 2016), the results show that gender gaps are statistically identical once individual time-constraints are accounted for. As a result, working women in the GDR, and in Germany until today, face the second shift. In the absence of partner reactions, this is because women fail to reduce their time spent in non-market work accordingly when increasing time spent on the labour market. The chapter concludes with a discussion of implications regarding the nature of gender norms for market and non-market work and effects of labour market policies targeted at gender gaps.

Chapters 4 to 7 are based on data from a large randomized controlled trial in Cambodia. While **Chapter 4** analyzes all treatment arms and waves and provides a details overview on the experiment, **Chapter 5** focuses on a specific aspect of treatment effect heterogeneity. **Chapters 6 and 7** utilize the rich data to analyze additional, complementary research questions using non-experimental methods.

Chapter 4 evaluates the impacts of a large-scale preschool construction program on school participation and child development in a low-income country, Cambodia, using a randomized controlled trial. The construction of new preschools, which entails the training of a designated teacher, was complemented with two different demand-side interventions that aimed to increase preschool enrollment by increasing caregiver's awareness about the availability and importance of preschool education and to strengthen the connection of the preschool through community mobilization. The construction caused an increase in enrollment of children at the age 3–5 but demand-side interventions had no effect. The impacts on enrollment are independent of socio-economic background. After one year, small average impacts on cognitive (0.04 standard deviations) and socio-emotional development (0.09 standard deviations) are driven by large effects on children from the wealthiest quartile. Persistent effects on children from the wealthiest quartile are also observed two years after the program started. The chapter provides evidence that the program had large impacts on aspects of structural preschool quality, such as the availability of basic equipment, but only limited impacts on process quality in preschool classes, such as the quality of teacher-child interactions. The results indicate that, in low-income countries, comparatively advantaged families are better able to take advantage of the preschools by providing a home environment that complements the education received at preschool. The results also suggest that improving the quality of educational processes might be needed to foster the development of disadvantaged children.

Chapter 5 studies the impacts of the same large preschool construction program. Focusing only on the first wave and the preschool construction treatment, the chapter discusses the fact that the newly built preschools compete with lower quality existing preschools as well as home care. In this context, the chapter highlights that impacts are likely to differ by counterfactuals, i.e. between those who would have been enrolled in a preexisting preschool and those who would have been in home care, with expected larger gains among the latter. Treatment effects of many other empirical studies have an underlying heterogeneity of this type. However, implications are often not discussed as the counterfactual care mode is not observable at the individual level. Using short-term data from an experiment conducted in Cambodia, several empirical techniques to isolate the impact on children who would have stayed at home had they not been enrolled in the newly built preschools are implemented. It is argued that the impact on these children is a central parameter in the preschool literature and is more comparable across studies and contexts than traditional reduced form estimates. First a bounding approach is discussed and implemented to show that, under reasonable assumptions, the effect on children who would have stayed at home absent the program is high and significant (between 0.14 and 0.39 standard deviations). Then an instrumental variable approach combined with lasso regressions is implemented to pinpoint the effect on these children. The results include consistent evidence that the impact on these children is around 0.2 standard deviations on a child development aggregate score while the effect on children who would have enrolled in a preexisting preschool (absent the newly built school) is small and insignificant. The paper concludes with a discussion of challenges in the planning of experiments and policies when counterfactuals matter for impacts of a program.

Chapter 6 utilizes the unusually rich data on children and preschool quality from the randomized controlled trial in Cambodia for additional analyses. While the data on preschool quality is used for descriptive statistics in **Chapters 4 and 5**, associations between preschool quality and child outcomes were not further explored. This chapter examines the value-added of different quality dimensions in Cambodian preschools for cognitive and socio-emotional development in early childhood. The study is conducted with a sample of 327 preschools, which includes almost 100 preexisting preschools, i.e. preschools which were not constructed under the randomized experiment. The preschool quality measures are based on detailed classroom observations, capturing multiple aspects of structural and process quality, such as the teachers' pedagogical approaches to cover the curriculum and their responsiveness to children's needs. Using factor analyses and principal component analyses, it is first shown that commonly measured categories of quality are multidimensional and difficult to summarize in few indices. Teacher cognition and training positively predict teaching practices. Despite the detail of the measures, no clear link between teaching practices and child development is found for cognitive skills and only a weak link is found for socio-emotional development. Yet, children in preschools with a particularly low structural quality due to a lack of basic equipment perform significantly worse. Policywise, the results suggest that achieving a universal quality standard for all preschool facilities, instead of following a targeted approach, should be a priority in the studied context. The missing link between observed process quality and child outcomes can also explain why effects of teacher training are generally found to be small and confirm that effective teaching is difficult to measure.

Chapter 7 complements the analysis of **Chapters 4 to 6** by a more detailed analysis of associations between the home environment and child outcomes. In addition, the chapter includes an in-depth discussion of the validity of the child development measures, which are also used in the previous chapters. It first summarizes how substantial work has demonstrated that early nutrition and home environments, including the degree to which children receive cognitive stimulation and emotional support from parents, play a profound role in influencing early childhood development. Yet, less work has documented the joint influences of parenting and nutritional status on child development among children in the preschool years living in low-income countries. Using panel data from 2016–17 on parenting, nutritional status, and early developmental outcomes (executive function, language, early numeracy, and socioemotional problems) of 6,508 Cambodian children ages 3 to 5 years, the findings demonstrate that inequities in early development associated with family wealth are evident at age 3 and increase among children ages 4 and 5 years. Using hierarchical regression analysis, a significant share of these inequalities is explained by differences in parenting and early nutritional status, measured by stunting. Better-educated parents engage in more stimulating and supportive parenting practices. However, the positive association between parenting and language and early numeracy outcomes is 35%-54% stronger for non-stunted children, and parental activities explain only about 8-14% of the cognitive gap between the lowest and highest wealth quintiles. The results highlight the need for additional research outlining interactions between environmental factors that link family wealth and child development. The results also complement the findings of **Chapter 4** by highlighting that dynamic complementarities of early stimulation, such as parenting or preschool education, can be weakened by pre-existing conditions, such as stunting.

Table 1.1: Overview and summary of the individual chapters 2-4

	Chapter 2	**Chapter 3**	**Chapter 4**
Title	Information Provision and Postgraduate Studies	Gender Inequality and the Relation between Market and Household Work	Improving Preschool Provision and Encouraging Demand: Heterogeneous Impacts of a Large-Scale Program
Research question	What are the effects of information provision on students' beliefs about pecuniary and non-pecuniary returns of postgraduate education?	How does the amount of housework of East and West German women depend on the own and spouses' market work?	What are the effects of a large preschool construction program in Cambodia on preschool enrollment and early childhood development?
Main finding	Males significantly adjust their beliefs and intentions to undertake postgraduate studies downward. This is driven by males upward-adjusting earnings expectations with a first degree only.	Time spent on housework is identical for West and East German women once individual time-constraints are accounted for. Men do not spend more time on housework when women work more.	The construction caused an increase in enrollment and small impacts on cognitive and socio-emotional development after one year.
Country	Germany	GDR, reunified Germany	Cambodia
Analyzed groups	Students	Households, couples	Children
Data	Berliner-Studienberechtigten-Panel (Best Up), PostGrad-Best Up	GDR time budget study, German Time-Use Survey, Multinational Time Use Study (MTUS), World Values Survey	Survey data
Methodological approach	RCT	Descriptive and correlational methods, Gelbach decomposition	RCT

Source: Own illustration.

Table 1.2: Overview and summary of the individual chapters 5-7

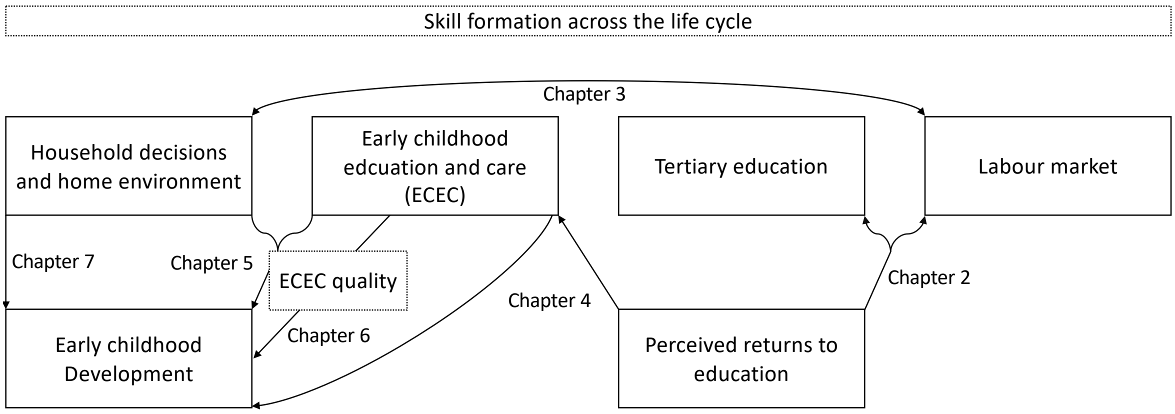
	Chapter 5	**Chapter 6**	**Chapter 7**
Title	Estimating Preschool Impacts when Counterfactual Enrollment Varies: Bounds and Conditional LATE	The Effects of Quality on Child Development in Cambodian Preschools	Joint Roles of Parenting and Nutritional Status for Child Development: Evidence from Rural Cambodia
Research question	How do the effects of a preschool construction program differ by counterfactual enrollment of children?	What are associations between preschool quality indicators and child development in rural Cambodian preschools?	What are the joint associations of parenting and nutritional status with child development in Cambodia?
Main finding	The effect on children who would have stayed in home care absent the new program is high and significant. The effect on children who would have enrolled in a preexisting preschool is small and insignificant.	No clear link between teaching practices and child development is found for cognitive skills. Yet, children perform significantly worse when schools are lacking basic equipment.	A significant share of inequalities in child development is explained by differences in parenting and early nutritional status. Associations are stronger for non-stunted children.
Country	Cambodia	Cambodia	Cambodia
Analyzed groups	Children	Preschool teachers, children	Parents, children
Data	Survey data	Survey data	Survey data
Methodological approach	RCT, bounding, instrumental variable methods, lasso regression	Value-added modeling, factor analysis, principal component analysis	Factor analysis, hierarchical regression models

Source: Own illustration.

1.3 Contributions

Chapters 2 to 6 make independent, but complementary contributions to the economic literature. **Chapter 7** makes additional contributions at the intersection of the fields of economics and developmental psychology. While all chapters make content-related contributions, the main contributions of **Chapter 5** are more methodological and particularly **Chapters 2, 6 and 7** document important data-related contributions. The main content-related, methodological and data-related contributions of this dissertation are outlined below. Figure 1.1 summarizes how the individual chapters are related.

Figure 1.1: Overview of the connections between different chapters



Source: Own illustration.

1.3.1 Content-related contribution

Information has been shown to effectively increase attendance and change the social composition of students (e.g. Jensen, 2010; Oreopoulos and Dunn, 2013). This dissertation contributes to the research on the role of information for education choice by providing experimental evidence in two very different contextual settings, where substantial socio-economic gaps in educational attainment are prevalent, possibly caused by information constraints. **Chapter 2** provides evidence on the role of information about pecuniary and non-pecuniary returns among German undergraduate students for their perceived returns and attainment of postgraduate education. Studies have documented the increasing variance in earnings within the group of college-educated workers, and estimated substantial returns to postgraduate education (Altonji et al., 2016; Lindley and Machin, 2016). Hence, whether to enter the labour market with an undergraduate degree only or pursure postgraduate education is a decision with important economic implications. Due to substantial socio-economic gaps in the attainment of postgraduate education in Germany, even conditional on obtaining a bachelor's degree, determining the role of information for the postgraduate education decision can have important policy implications. The results of this chapter document that information frictions exist even for students already enrolled in undergraduate degrees.

In a different setting, **Chapter 4** studies the role of information for preschool enrollment in rural Cambodia. The main treatment in this randomized controlled trial, the construction of preschools, is paired with two demand-side intervention. Under the demand-side interventions, parents are educated about the new preschools, enrollment procedures and about the importance of early childhood education in general. These demand-side approaches were not enough to mobilize a significant amount of additional enrollment over simply building preschools, suggesting that other factors than information constraints determine the decision to send children to preschool in this setting.

The socio-economic background, i.e. particularly the home-environment of a person, is also characterized by the internalization of norms. These norms play an important role in explaining a multitude of economic outcomes (e.g. Fogli and Veldkamp, 2011; Fernández, 2013; Bertrand et al., 2015). Multiple studies have highlighted how the gender-egalitarian policies of the GDR regime have affected female labour market participation and child care usage, as well as their related norms (Lippmann et al., 2020; Sprengholz et al., 2020; Campa and Serafinelli, 2019; Zoch, 2021; Bauernschuster and Rainer, 2012). **Chapter 3** extends this discussion by highlighting that gender norms regarding market work need not be closely related to gender norms regarding housework. The chapter highlights that there are no male reactions to different levels in female market work regarding their household work. This complements recent findings that males do not adjust their own labour supply when their wives start working (Knowles, 2013).

The promising evidence on highly effective targeted ECEC services (see e.g. Grantham-McGregor et al., 1991; Elango et al., 2015) exists alongside more mixed evidence on the effectiveness of large-scale ECEC services for child development, particularly from low-income countries (see e.g. Martinez et al., 2017b; Brinkman et al., 2017; Bouguen et al., 2018a; Blimpo et al., 2019; Bernal et al., 2019). In addition to compromises in quality, implementation issues, including low construction rates, lifespan of programs and lack of individual take-up, plague some of these studies. **Chapters 4 and 5** add to this field with evidence on a public preschool construction program in rural Cambodia, with detailed information on both children and on the quality of the new preschool services provided. The program is particularly large, well-implemented and with high take-up in comparison to pre-existing studies. With particular focus on heterogeneous effects and their determinants, the effectiveness of the program on cognitive and non-cognitive outcomes and areas for further improvements of the program are discussed.

1.3.2 Methodological and data-related contributions

Chapter 5 focuses on a particular mechanism behind the heterogeneous effects of the evaluation preschool program in Cambodia. While subgroup analyses typically focus on impacts of an intervention conditional on observable characteristics, this chapter focuses on effects conditional on the counterfactual care arrangement – which is unobservable at the individual level. The chapter builds on methodological contributions by (Kline and Walters, 2016; Feller et al., 2016; Hull, 2018) by discussing their importance for early childhood education research and applying them to the preschool program in Cambodia. In addition a bounding approach is discussed. The results show that the identification of "subLATEs" can be applied easily using realistic assumptions. Further, the chapter highlights that incorporating the method in subgroup analyses of similar studies would lead to an improved comparability of study results.

Finally, several unique, well-documented and publicly accessible data sets are used and discussed for the first time in this dissertation. The PostGrad-Best Up study, including the collection of 5 waves of data, is described in detail in **Chapter 2**. The chapter not only studies an information intervention conducted with a unique sample, but also serves as a documentation for future studies. The data of the trial in Cambodia, used in **Chapters 4 to 7**, is particularly extensive and the outcome of over 6 months of field work, distributed over three data collection waves, by a team of several dozens of dedicated field workers. The rich data allows an unusually comprehensive analysis of the ECEC infrastructure in rural Cambodia. In addition to answering own research questions, **Chapters 6 and 7** also serve as a documentation of the validity of the parenting measures, tests for cognitive and non-cognitive development of over 7000 children and detailed measures of preschool quality of 327 preschool, which were developed and adapted under this study. For **Chapter 3**, access to the 1985 and 1990 waves of the GDR time budget study (Zeitbudgeterhebung) was obtained at the German Federal Archives (BArch). The data has, to the best of my knowledge, not been used for research in economics before. It allows for unique insights in this dissertation and offers opportunities for further research.

CHAPTER 2

Information Provision and Postgraduate Studies[1]

2.1 Introduction

Numerous studies in the economic literature document that compared to vocational training or high school education, returns to college education are high, with Katz and Murphy (1992) being a well-known early example. Consequently, the individual decision to enroll in college or not, has been widely studied. Based on quasi-experimental and experimental studies, we know that information, costs and beliefs all play important roles in explaining college decisions.[2] Recently, studies have documented the increasing variance in earnings within the group of college-educated workers, and estimated substantial returns to postgraduate education (Lindley and Machin, 2016; Altonji et al., 2016). This suggests that not only the initial decisions to enroll in college, but also postgraduate enrollment decisions matter.[3] Yet, comparatively little is known about factors that influence individual decisions to pursue postgraduate education. In a recent study, Boneva et al. (2019) show that pecuniary and non-pecuniary factors play a role by using a choice model, but, to the best of our knowledge, experimental evidence on factors that affect the postgraduate education decisions does not exist.

[1]This chapter is joint work with Frauke Peter (German Centre for Higher Education Research and Science Studies (DZHW), DIW Berlin), C. Katharina Spiess (DIW Berlin Freie Universität Berlin), Felix Weinhardt (DIW Berlin, European University Viadrina, CESifo, IZA, CEP/LSE). We thank for comments by Eric Hanushek, Guido Imbens, Dorothea Kübler, Nicolas Salamanca Acosta, Marta Golin, Katharina Wrohlich and participants of the 2019 VfS annual conference, of the Potsdam Workshop in Empirical Economics, and of the EALE annual conference 2019. We gratefully acknowledge funding from the German Science Foundation (SP 1091/2-1), and Felix Weinhardt acknowledges additional funding through the German Science Foundation (CRC TRR 190). The study is registered at the AEA RCT registry (AEARCTR-0002446), and we obtained ethical approval. No third party had the right to preview our results. All views and remaining errors are our own.

[2]See detailed literature review below.

[3]Note we use “postgraduate education” in the European sense, i.e. including master’s degrees, which is called “graduate education” in the US.

This study starts to fill this gap by studying effects of information provision about pecuniary and non-pecuniary returns to postgraduate education to undergraduate college students close to completion of their bachelor's degrees, in a randomized controlled trial (henceforth, RCT). We study effects of our randomized treatment on beliefs about pecuniary and non-pecuniary returns and how this affects postgraduate enrollment intentions six months later. Moreover, we can provide evidence on realized enrollment in postgraduate education one and two years after treatment.

The treatment consists of information about pecuniary and non-pecuniary returns of postgraduate versus undergraduate degrees in the labor market, based on empirical data of existing employees. The treatment is delivered at the end of an online survey to a randomly selected subgroup of our sample. Note, we do not present any information on costs and benefits of the student experience as such. Our target population already has first-hand experience on these through their undergraduate studies. In this regard, the information set available to students who decide about postgraduate enrollment differs to the information and decisions about initial college-going at the end of high school. Rather than providing students with information about the student experience, our treatment gives information about pecuniary and non-pecuniary returns on the labor market depending on undergraduate or graduate degrees, of which undergraduate students have no first-hand experience.

The study population was recruited out of an existing experimental panel study that focused on the initial college-going decision of high school students of the 2014 graduation cohort in the Berlin area, Germany (Peter and Zambre, 2017; Peter et al., 2018). Our focus on the 446 students presumably enrolled in their final years of the undergraduate program in 2017 resulted in a number of benefits, including access to information on pre-baseline characteristics that were collected in the past. In particular, pre-baseline information on postgraduate enrollment intentions was available and we used this, together with more background variables, to implement a randomization design based on pair-wise matching. Moreover, we believe the fact that the targeted students have had experiences in a previous panel study might explain the very low rates of attrition in the three follow-up surveys of this experiment.

We use the information collected in the intervention and the three follow-up surveys that we conducted six months, one year and two years after treatment in four steps.

First, we present correlations between postgraduate enrollment intentions and pecuniary and non-pecuniary returns for our control population. This confirms the relevance of both sets of factors, in line with existing research that stresses that non-pecuniary factors matter in addition to –and potentially more than– pecuniary factors for postgraduate education (Boneva et al., 2019).

Second, we examine how the treatment has shifted individual beliefs about pecuniary and non-pecuniary returns of postgraduate education. This is interesting both, to understand later effects on intentions and enrollment, but also in its own right as it sheds light on belief updating. This is because the treatment consisted of objective information on a range of attributes of jobs, for example average earnings for different occupations – and so depends not only on existing beliefs but also on how students place themselves in the categories that we have presented. The main finding here is that many students previously either held very accurate beliefs about pecuniary and non-pecuniary differences between graduate and postgraduate jobs, or did not significantly update their beliefs due to our online information intervention. The largest, and statistically significant, updating of beliefs occurs for males, who downward adjust their expected postgraduate earnings premium. This is driven by higher expectations about earnings with a first degree, rather than lower earning expectations with a second degree. We explore reasons for this gender heterogeneity and discuss that it can be related to differences in initial beliefs or differences in updating, i.e. processing of the new information.

Third, we examine how the treatment affected postgraduate enrollment intentions stated six months later. Here, we find effects that mirror the effects on belief updating documented above: males are significantly less likely to state the intention to directly enroll for a postgraduate degree following the successfully completion of their undergraduate studies. We find further heterogeneity along parental background, which however are rarely significant at conventional levels of statistical significance.

Fourth and finally, we estimate effects on postgraduate enrollment one and two years after initial treatment. Here, we again find the largest and negative estimates for male students. These estimates are not statistically significant, yet, growing in magnitude and almost statistically significant after two years when more students in our sample completed their undergraduate studies and faced the decision to enroll into a postgraduate program.

Taken together, we present causal evidence that an information treatment on pecuniary and non-pecuniary returns can have long-run consequences. We show significant effects on enrollment intentions measured six months later, and supportive evidence on enrollment one and two years after treatment. In addition, we document that the treatment has lead to a different updating of beliefs of students, with male students significantly downward adjusting expectations of postgraduate wage premia. These differences in belief updating from the same treatment are in line with the heterogeneity in the effects that we document on direct postgraduate study intentions and enrollment. This study therefore has two main contributions: first and foremost, we provide the first causal evidence of the role of information for postgraduate enrollment decisions. Moreover, we document that the heterogeneity that we find in the treatment effect of receiving information is in line with the heterogeneity that we find in belief updating. This means that despite our finding that male students react strongly to our information treatment, and female students do not, this does not imply that males and females place a different importance on information when making decisions. Differences in belief updating and information processing presents an alternative explanation to heterogeneity in treatment effects of information treatments.

This study is related to the large literature on the role of financial constraints or on the lack or effectiveness of information about actual costs and future monetary returns for the college enrollment decisions (see for example, Dynarski, 2002; Dynarski and Scott-Clayton, 2006; Bettinger et al., 2012; Oreopoulos and Dunn, 2013; Wales, 2013; Bettinger and Baker, 2014; Castleman et al., 2014; Kerr et al., 2020; Wiswall and Zafar, 2015; Castleman and Long, 2016; Oreopoulos and Ford, 2016; Carrell and Sacerdote, 2017; Dynarski et al., 2018). In the German context, Peter et al. (2018) and Peter and Zambre (2017) study the effects of providing information about returns and financing possibilities for college education to high school students. One key finding is that students of non-academic background, in particular those with intentions to enroll, are more likely to pursue college education if they have received information about its benefits. Moreover, an existing literature on individuals' beliefs about returns to educational investment shows that besides pecuniary, especially non-pecuniary returns can explain educational decisions (Boneva and Rauh, 2017; Belfield et al., 2019). This paper differs from this literature because we study postgraduate education decisions. In an important and recent paper, Boneva et al. (2019) show that both pecuniary and non-pecuniary returns also matter for postgraduate education decisions. We complement this literature by providing first experimental evidence on these, as well as on the role of information.

The remainder of this paper is structured as follows. The next section describes the institutional context and data. In Section 2.3 we describe the treatment, randomization and compliance. Section 2.4 describes our estimation strategy and outcome variables. Section 2.5 presents the estimates on stated beliefs, enrollment intentions and enrollment. Here, we also provide descriptive evidence on the association between pecuniary and non-pecuniary postgraduate returns and enrollment intentions. Section 2.6 concludes.

2.2 Institutional context and data

2.2.1 Institutional Context

Almost 20 years after the Bologna-process[4], Germany has a well-established two-tier setting where students first enroll for a bachelor's degree that typically lasts for four years.[5] No exlusively long first degree programs exist and fast track programs to earn a doctoral degree without a master's degree are not often used. Overall, about 60 percent of bachelor graduates move on to study for an additional two years to earn a master's degree (Spangenberg and Quast, 2016). From ten students who continue almost seven do the master's program at the same higher education institution where they earned their bachelor's degree (Fabian et al., 2016).[6] Both percentages are higher for university students compared to those at universities of applied sciences, which usually offer longer and more practically oriented degrees. Moreover, a higher percentage of students with academically educated parents and relatively more male students continue with a master. For these groups the transition rates are between 70 and 80 percent, while the others have lower rates (50 to 60 percent). However, compared to other countries, the share of 25-34 year olds with a bachelor's degree in 2019 is with 16% in Germany still lower than in other countries with a much longer tradition of a two-tiered system, such as the UK with 24%. The share of those with a master's degree is 12% in both countries (OECD, 2020).

Bachelor students usually apply for a master's program in the last semester of their studies, which usually lasts for three to four years. Most master's programs start in the winter-term, which means that students have to apply in the early summer before. Most bachelor students who continue with a master's program do this without an interruption. Only about 20 percent perform or plan a transition after a short interruption (Spangenberg and Quast, 2016). The main reason for an interruption among university students are internships: In a survey among bachelor degree students this is stated by 36 percent. The reported main reason among other students is the intention to gain work experience: 42 percent. The main reasons for no transition to a master's program are attractive job offers or intended work experience (Autorengruppe Bildungsberichterstattung, 2018). In general, higher education in Germany (at public institutions) is free of charge, with students paying only a small administrative fee each term. There are no fee-differences between bachelor's and master's programs.

[4]The Bologna process has created the European Higher Education Area. The Bologna declaration was signed by education ministers from 29 European countries in 1999.

[5]Before the Bologna-process in the 2000s Germany had a system of longer single-tier degrees.

[6]In the sample we base our analysis on, these are 64% who continue at the same higher education institution. Moreover, 76% continue their studies in the same city.

This does not mean that a master's program is free of costs, once living costs and opportunity costs are considered.

Although the two-tier system is well established in the higher education system, it is obvious that there is little if not no knowledge on the life-long career prospects of finishing tertiary education with a bachelor's or master's degree in Germany. After 20 years, there are no graduates from bachelor's or master's programs who reached the retirement age and subsequently there is no empirical evidence on lifetime earnings of these degrees. Thus, any empirically based information on non-pecuniary and pecuniary returns of a master's degree are limited to mid-term returns. Even this type of information is relatively new and not widely discussed in the public so far. Thus, an information treatment on the returns of a bachelor's and master's degree might be more effective compared to other countries with a longer tradition with these two degrees. Furthermore, an effective information ideally has to be placed in a period where the decision is not finalized, but students actively consider whether to transition to a master's program. Thus for the average student, who starts a master's program in October and thus has to sent out admissions around June, a period of 5-7 months before might be suitable for an information treatment.

2.2.2 Data

A central design feature of this RCT is that we sample the students from a population of students who are likely to pursue postgraduate studies or to enter the labor market after their undergraduate degree. We exploit existing knowledge about students from the *Berliner-Studienberechtigten-Panel* (Best Up) to sample our study population. This panel study provides us the necessary target population, as it comprises vast information about students starting from their enrollment in undergraduate studies until students' early intentions of postgraduate enrollment.[7] The Best Up data contains very detailed information about students of the cohort that graduated from high school either in summer 2014 or one year later. These students come from a relatively homogeneous environment and are followed from the last year prior to high school graduation (*Abitur* in German) to the first two years of college or vocational training. Although the Best Up data provides us with undergraduate students from the same high school cohort, not all directly enroll in college after high school graduation in 2014. Around 30% of the Best Up participants take a gap year after high school. Thus students in our sample are progressing at different speeds through their undergraduate studies. In addition, the speed varies, because students enrolling in universities of

[7]For further information on the Best Up study and data see Ehlert et al. (2017); Peter and Zambre (2017); Peter et al. (2018).

applied sciences take on average one year longer to finish their bachelor's degree due to different program structures compared to university majors.

In the Best Up data, a majority of students start their final year of undergraduate studies in winter 2017.[8] Thus, in winter 2018 these students are likely to transit to postgraduate studies or to enter the labor market. Out of the Best Up data we identify 446 students who are likely to be studying in winter 2017 as our target population of which 371 students (83%) participate in the baseline survey (see Section 2.3 for more information about the RCT and the survey).

In Table 2.1 we provide further descriptive evidence looking at students from the first survey of our study. We show means of all matching (pre-trial) variables as well as baseline study and background characteristics. The sample consists of a majority of students from a non-academic background, who are slightly less likely to enroll in postgraduate degrees.[9] Students are on average 23 years old and in their fifth semester.[10] This shows that the majority of students in our sample is nearly at the end of their undergraduate degree, as students on average study 7.2 semesters to compete a bachelor's degree (Autorengruppe Bildungsberichterstattung, 2020). As described in Section 2.2.1 some majors are still organized under the old degree system prior to the change to the two-tier structure. In our sample about 6% of students are enrolled in such a major. The majority of students (78%) are enrolled in a bachelor's degree. 46% students in our sample intend to directly enroll in postgraduate studies in December 2017 (baseline) and 48% in May 2018 (first follow-up). In December 2018, 26% of those students who participated in the baseline and one year follow-up of our study (N=293) are enrolled in postgraduate studies. In February 2020, this number has increased to 41% (see last rows in Table 2.1).

A first comparison of our initial target sample with a nationwide representative study already shows, that students are fairly comparable in terms of age, final GPA, intentions to enroll in postgraduate studies to the average German student (see Table 2.5 in the Appendix 2.B.1).[11] Students from the NEPS SC4 cohort also graduated from high school in 2014 and 50% come from a nonacademic background, i.e. are first generation students, compared to 59% in our baseline sample. In the NEPS SC5 cohort, students were sampled in the winter term 2010/2011 at German universities and universities of applied sciences. In this sample 63% of students are first generation students and 83%

[8]See also Figure 2.3 for more information on the stylized timeline to an undergraduate degree in Germany in relation to our trial timeline.

[9]Students are considered to come from a non-academic parental background if none one of their parents holds a college degree.

[10]In Germany one year of college is divided in two terms called semester.

[11]We compare our sample to two so-called starting cohorts of the National Educational Panel Study (NEPS): SC4 and SC5 (see Blossfeld et al., 2011, for more information).

are enrolled in a bachelor program, compared to 78% in our baseline sample (see Table 2.5 in Appendix 2.B.1).

2.3 Details of intervention

We conducted five online surveys to accompany bachelor students from the *Best Up panel* at the transition to postgraduate studies or the labor market. In the baseline survey in December 2017/January 2018 we routed students according to their treatment status and presented to those in the treatment group a series of screens with information about realized pecuniary and non-pecuniary returns on the labor market differentiated by college degree. The following section introduces the treatment in detail and discusses timeline and set up of the randomized controlled trial.

2.3.1 Treatment

The information treatment consists of an online learning module that informed students about different aspects relevant for the postgraduate decision. The learning module comprised visual and audio information and addressed three topics: realized pecuniary and non-pecuniary labor market returns by college degree, e.g. earning levels and differentials for different occupations and sectors, and funding options for postgraduate studies.

The online survey was programmed and administered by a survey institute (Kantar Public) to ensure a professional interaction with survey participants. Students were invited via email to participate in an online survey providing them with an individual link. The link worked with smartphones, tablets, and PCs, as the online survey was mobile-ready. In this online survey, the presented information allowed students to place themselves and to update their beliefs about their individual returns. This is important since providing students with a single number, like average postgraduate earnings, might be misleading. The information provision was based on visual and audio material. Students were shown some informative graphs with explanatory text and helpful audio explanation transporting the depicted information. After each information slide, students were asked to answer a comprehensive question about the previous screen. Students could not continue to the next screen without listening to the audio file and without answering the short comprehensive question. However, students could go back to the previous screen. Moreover, we presented information about a range of non-pecuniary labor market returns depending on degree (bachelor's vs master's degree), such as on the likelihood of working in a high-skilled occupation.

In Figure 2.1 we show two exemplary visual information that students received in the online learning module.[12] The figure at the top of Figure 2.1 presents a pecuniary example of realized labor market returns by college degree type, whereas the bottom figure depicts a non-pecuniary example, namely probability to work in high-skilled occupations. For all measures depicting pecuniary returns examples we used data from the Microcensus (*Mikrozensus* in German).[13] The measures for the non-pecuniary examples were constructed using another large nationwide household survey the German Socio-Economic Panel Study (SOEP).[14] Using large representative data sets to construct the measures for the treatment allowed us to tailor the information to students close to finish their bachelor's degree and to provide students with information not widely available. Providing information by different college qualifications is important, as students might not observe pecuniary and non-pecuniary returns in their environment. Numbers on realized labor market returns are not widely available in newspapers or on the web for different college degree types given that the two tier system is still young.

The realized earnings by college degree and occupational group shown in Figure 2.1, for example, depict average realized earnings of bachelor graduates in jobs in natural sciences as well as those of master graduates and explicitly mark the difference between both qualifications. Earnings over time and separated by gender were shown in other graphs. Although the two tier system is still young, master graduates have been around long enough in the labor market to describe realized wage differences in the first years of employment. These first five to ten years after college graduation are exactly those years of realized labor market returns we are interested in to support near-bachelor-graduates at the transition to postgraduate studies or the labor market with relevant information.

For non-pecuniary labor market returns the online learning module comprised for example information about the likelihood to work in high-skilled occupations (see bottom figure in Figure 2.1). This visual material shows the percentage of employees with bachelor or master's degree working in highly skilled occupations, i.e. in team/department leading positions or managerial position.

Apart from pecuniary and non-pecuniary realized labor market returns the online learning module also comprised information about funding possibilities of postgraduate studies. We informed students about different funding sources in Germany and high-

[12]Examples of programmed screens as seen by students are included in Appendix 2.B.2.

[13]The Microcensus is an annual household survey providing nationwide representative statistics on the population and the labour market in Germany. It surveys 1% of the population in Germany.

[14]The SOEP has been carried out since 1984 and in 2017 more than 30,000 individuals in approximately 17,000 households participated in (see Wagner et al., 2007).

lighted for example that students can also apply for student aid (*BAföG (Bundesausbildungsförderungsgesetz)* in German) for a master degree, as many eligible students tend to believe the support covers only the first degree. Appendix 2.B.2 comprises example screenshots of programmed visual material shown in the online learning module. Together with the field institute, we worked out ways to provide the information on multiple screens to make the content reader-friendly and to monitor students' behavior: they were only able to continue to the next screen after listening to the figure-guiding audio message and answering a simple knowledge-based question. These questions were implemented to ensure that students had looked at the material and understood the visualized information, as it is otherwise very difficult to know for certain that students looked at the information with online or handout based provision compared to information provided in person (see for example, Oreopoulos and Dunn, 2013; Peter and Zambre, 2017).

The information treatment consists of various pecuniary and non-pecuniary labor market returns, providing also information by different fields and by gender, as we do not know what type of returns the individual will experience. Showing realized returns differentiated by college degree helps students to place themselves using their own best guess.[15]

2.3.2 Implementation, timing of intervention

We implemented in total five online surveys which were optimized to smartphones, tablets and computers for easy access to participation. In a first very short pre-trial survey we assessed how many students would still be studying in the winter term 2017. This pre-trial survey took place from August to September 2017 (see Figure 2.3 in Appendix 2.B.1). The following four online surveys took place from December 2017 to January 2018, from May to June 2018, from December 2018 to January 2019, and in February 2020. From the pre-trial survey in fall 2017 we received a target population of 446 potential students still studying for the bachelor's degree in winter 2017. Out of these 446 students, response rates in all three trial surveys are very high and lie always clearly above 80% (see number of participants per survey in Figure 2.3).[16]

The baseline survey in December 2017/January 2018 was conducted about 7-8 months before final year students would typically graduate with an undergraduate degree. At the end of this first online survey treated students were routed to the online learning module (see also Section 2.3.1). The first follow-up survey was 6 months later in

[15]The study has been approved by an IRB (see for more information the AEA RCT registry entry under `https://doi.org/10.1257/rct.2446-2.0`).

[16]Compared to other response rate of similar RCTs, this response rate is very high and satisfactory.

May/June 2018. With this first follow-up we were able to measure students intentions to enroll in postgraduate studies. These intentions measured up to 6 months after the first survey are comparatively long run intentions and most likely coincide for the majority of students with their application process for postgraduate studies. The second follow-up survey was conducted 12 months after treatment in December 2018. With this second follow-up we asked students about their actual enrollment. With the winter term 2018, we expected most students to have graduated from their undergraduate studies and have directly enrolled in master's program. As our data from 12 months after treatment shows, this second follow-up was still a little bit early to detect the full effect on actual enrollment, as students are still more likely to be enrolled in undergraduate programs and less likely to be enrolled in postgraduate studies. For this reason an additional follow-up was conducted in February 2020 where postgraduate enrollment rates were much higher (Table 2.1).

2.3.3 Randomization and compliance

The randomization of students into treatment and control groups was implemented using pair-wise matching. Bruhn and McKenzie (2009) show that in small samples, other methods than pure randomization can improve the degree of balance among relevant pre-treatment characteristics and follow-up outcomes. Pair-wise matching allowed us to balance treatment and control students matching on many variables predictive of the outcome variables and thereby increasing efficiency and power of hypothesis testing. We applied the greedy pair-wise matching algorithm mentioned and provided by Bruhn and McKenzie (2009). Since we utilize data from the *Best Up panel*, we had enough information and time to perform randomization using matching techniques, as information about, for example, pre-trial postgraduate intentions, GPA from high school graduation or gender was already available. Pair-wise matching using baseline characteristics would not have been feasible, as the treatment took place immediately after the baseline data collection. After having selected "statistical twins" based on a rich set of pre-treatment characteristics, we randomized participants in each pair into treatment and control groups.

Table 2.2 shows the balancing of covariates in the pre-trial survey (Aug/Sept 2017) and the baseline survey (Dec 2017/Jan 2018). We separately regress balancing variables on a treatment group dummy to calculate raw treatment group differences. To account for the ex-ante balance approach, we further regress balancing variables on a treatment group dummy and pair fixed-effects, dropping all the observations from incomplete pairs. The actual difference between treatment and control group means is not statistically significant for the pair-wise matching variables and the variables on intentions, enrollment, as well as background characteristics. Yet, a statistically significant higher share of treated students has a migration background. The same picture emerges for the treatment group differences controlling for pair fixed effects.

The attrition rate is never significantly related to the treatment and equals 11.9% in the control group at the first follow-up, 19.5% at the second follow-up and 14.6% at the third follow-up. Attrition is also not related to most matching variables and important predictors of postgraduate enrollment intentions in neither the control nor treatment groups. Although attrition is small and does not differ between treatment and control groups, we see a small statistically significant difference between male and female participants (see Table 2.6 in Appendix 2.B.1). Females are less likely to drop out at the second follow-up. While this does not imply that treatment effect estimates are biased, we acknowledge that it might limit representativeness of our estimates for our baseline sample. We therefore also run separate regressions for females and males in the analysis below.

2.4 Empirical analysis

2.4.1 Estimation specification

The main model for the estimation of treatment effects is

$$Y_i^{post} = \alpha + \delta T_i + \gamma W_i + \beta X_i + \epsilon_i, \tag{2.1}$$

where Y_i^{post} is the post-treatment outcome of a student i and T_i is a binary treatment group indicator. In our main specification, we control for variables used for pair-wise matching W_i to account for the randomization procedure. In addition, we control for a set of baseline characteristics X_i to improve power (direct and general enrollment intentions and postgraduate enrollment at baseline).

As shown by Bruhn and McKenzie (2009), the most accurate way to account for the randomization procedure is to run a regression on the treatment group indicator and pair fixed effects. Otherwise, standard error estimates tend to be overly conservative. However, this leads to the omission of observations from pairs with only one follow-up observation. As this lowers the effective sample size – particularly for subgroup analyses – we control for matching variables W_i instead. In Table 2.7, we show that our main results are robust to using different sets of control variables. In Table 2.12, we provide evidence that our results are robust to using only complete randomization pairs and randomization inference based p-values (Young, 2019).

2.4.2 Outcome variables and effect heterogeneity

A first set of outcome variables consists of students' beliefs about pecuniary and non-pecuniary labor market returns by degree type. We asked students to rate the answers to the following question[17]: "Please think about the time in the near future when you are 30-35 years old. Further assume, you are working full-time then. Certain aspects of your life might depend on whether you graduated with a bachelor's degree or with a master's degree. How likely do you think that you will . . . ". We provided students with the following five pecuniary and non-pecuniary aspects for both bachelor and master's degree and asked to rate these on a likert-type scale from 1 (very unlikely) to 7 (very likely): (1) to earn above average income, (2) to do intellectually challenging work, (3) to be able to combine work and family life, (4) to work in a highly-skilled job or with managerial responsibility, or (5) that parents are satisfied with their job.[18] Table 2.14 shows how students on average rate these five dimensions for each degree type. For example, students on average rate the probability to work in a highly-skilled job or with managerial responsibility with 4.1 with a bachelor's degree and 5.5 with a master's degree.[19] We construct the perceived postgraduate return measures as the difference in perceived probabilities between a master's and bachelor's degree.

The second set of outcome variables comprises students' *intentions* to pursue postgraduate studies and their *actual enrollment* in master's programs. Given the German context, measuring students' postgraduate application behavior is somewhat difficult, similar to measuring undergraduate application (see Peter et al., 2018, for a discussion regarding bachelor's programs). Not all study programs require students to apply. In many programs, they can just enroll without any further requirements. We therefore focus in particular on students' intentions to enroll *directly* after obtaining a bachelor's degree. We measure postgraduate enrollment intentions using a binary variable measuring *direct* transition intentions. We define *direct intentions* as intending to enroll in a postgraduate program immediately after completion of the bachelor's program. We code all students for those the question does not apply due to permanent study termination as 0 and students already enrolled into a postgraduate or 5-year program as 1.

Enrollment in postgraduate studies is defined for students who completed their bachelor's degree and are enrolled in a master's program at the second follow up survey.

[17]See Section 2.B.6 for the full German and English version of the question.

[18]We use these particular categories to elicit students' beliefs, as they have been shown to matter by Boneva and Rauh (2017), which allows us to compare our findings to the emerging literature on pecuniary and non-pecuniary returns and educational choices.

[19]Table 2.14 shows mean values for the control group only, as well as standard deviations and differences between groups. For reporting treatment effects on perceived probabilities, standardized variables are used to make results easier to interpret.

They are coded as 1 and the bachelor graduates no longer enrolled in higher education, either at universities or universities of applied sciences, are coded as 0. As the scope for finding effects 12 months after treatment is limited if students progress slower through their studies than the population average we conducted an additional follow-up survey two years after the treatment. This is particularly important for nonacademic background students, who are less likely to directly enroll in postgraduate studies (see Section 2.2.1).

Besides a potential deferral of enrollment, nonacademic students might differ in their beliefs about labor market returns to a postgraduate degree compared to students from an environment where college returns are observable (see Boneva and Rauh, 2017). In addition, the effects might also vary by gender. Studies show that the expected returns to a bachelor or master degree differ for male and female students (see for example Reuben et al., 2017; Zambre, 2018). Due to a relatively large number of missing observations for students' fields of study and the observed distribution over the different subjects (see also Table 2.1) we are not able to estimate effects by fields of study without getting too small sample sizes. Yet, we run our main regression with study subject fixed effects as a robustness check.

2.5 Results

2.5.1 Beliefs about pecuniary and non-pecuniary returns

In order to elicit students' beliefs about different returns for either bachelor's or master's degrees, we asked them to rate five pecuniary and non-pecuniary aspects separately for each degree type on a likert-type scale from 1 to 7 (see Section 2.4.2 for more details). In the following, we show associations between postgraduate enrollment intentions and a range of pecuniary and non-pecuniary return factors for the control group at the first follow-up. Table 2.3 presents estimates of a new regression equation in each column, with differing sets of covariates: column (1) shows that the perceived probability to earn an above average income is significantly correlated with intentions to study a master's degree. Columns (2) to (5) introduce the non-pecuniary factors. Interestingly, parental satisfaction with job does not matter, possibly because these bachelor students are already somewhat detached from their parental background.[20] All other factors are significant predictors in the association between non-pecuniary beliefs and direct enrollment intentions. Next, in column (6) we jointly estimate the associations between

[20]This finding is different to Boneva et al. (2019) who find a large and statistically significant effect for parental support. However, it might be possible that they measure another aspect of parental support.

perceived pecuniary and non-pecuniary returns and enrollment intentions. Last but not least, in column (7) we combine the non-pecuniary factors into a preference-weighted index of non-pecuniary returns.[21] As before, and in line with the existing literature, we find that these two measures of perceptions of pecuniary and non-pecuniary returns matter (Boneva and Rauh, 2017; Belfield et al., 2019).

2.5.2 Causal effects on perceived returns

In Figure 2.2 we show treatment effects on students' beliefs about pecuniary and non-pecuniary postgraduate returns. The upper panel shows the overall effect for the preferred specification, the middle panel effects by gender, and the bottom panel effects by academic background.[22] Each sub-figure in Figure 2.2 lists the results for the one pecuniary return measure ("earn above average income") and the four non-pecuniary return measures and their summary index, as described in Sections 2.4.2 and 2.5.1.

Figure 2.2 shows that treated students increase their beliefs about a bachelor's degree six months after treatment (see also Table 2.13 in appendix 2.B.1). Treated students mainly update their beliefs about monetary returns with an undergraduate degree, but also increase the returns to a bachelor regarding non-pecuniary returns, such as "to do more intellectually challenging work" and "to work in a highly skilled job". In particular treated male students and students with at least one parent with a university degree increase their pecuniary beliefs about a bachelor's degree, making a master's degree relatively less attractive. The differences in the distribution of perceived probabilities of males (Figure 2.4) show that post treatment males are more likely to report probabilities above 4 on the Likert scale, in particular for earning above average income with a BA. Similar updating cannot be observed for females (Figure 2.5).[23]

Returning to the standardized impacts, treated male students increase their belief "to earn above average income" with a bachelor's degree by 0.42 sd. This leads to a similar decrease in the difference between both degrees (-0.42 points, see also column 3 in Table 2.13 in the appendix). Students from an academic background also significantly adjust their beliefs about monetary returns of a bachelor's degree. Similar to male treated students, the difference between the beliefs by degree significantly decreases by 0.36 points.

[21] Students were asked to rate the importance of each category on a scale from 1 (not important at all) to 5 (very important). We constructed the non-pecuniary return index by weighting each of the 4 non-pecuniary return measures with the relative importance reported by the respective student.

[22] The estimates are also summarized in table format in Table 2.13 in appendix 2.B.1, which includes regressions on the difference between probabilities with a bachelor's and master's degree.

[23] For the exact phrasing of the German original question as well as for the English-translation see Appendix 2.B.6.

Female students and students from non-academic backgrounds do not significantly update their beliefs about either degree due to the information treatment. In addition, we show the absolute levels of perceived returns in the control group in Table 2.14 and distributions of perceived probabilities for the treatment and control group in Figures 2.4 and 2.5. Males and females in the control group have fairly similar perceptions of bachelor's and master's degrees. Yet, untreated males tend to assess bachelor's degrees somewhat worse than untreated females, which can be seen as suggestive evidence that males are more likely to underestimate the returns to a bachelor's degree, but these differences are hardly statistically significant. This pattern is not observed for students from academic vs. non-academic background.

Looking at preferences about pecuniary and non-pecuniary returns, Table 2.15 in the appendix shows that for male students it is slightly more important to earn above average income, doing intellectually challenging work and working in highly skilled jobs than for female. Whereas female students place higher values on the work-life balance aspect. Yet, only the gender difference regarding work-life balance and to work in a highly-skilled jobs are statistically significant. The information treatment provided more information on aspects which are particularly important to men, which is one potential explanation for the gender-differences in treatment effects on perceived returns. This conclusion is also supported by literature from several disciplines which points towards gender differences in work values or career preferences. Women have been found to attach higher value to things not covered in this information treatment than men, such as enjoying the work at their jobs, helping others and interacting with people (Busch-Heizmann, 2015; Diekman et al., 2010; Lippa, 2005; Weisgram and Bigler, 2006; Weisgram et al., 2011; Zafar, 2013).

To summarize the effects on perceived returns, the most significant finding is that males expect higher BA returns. In principle, such gender heterogeneity could reflect underlying differences in baseline beliefs, i.e. that males correct their beliefs post treatment, or alternatively gender-differences in updating could explain the heterogeneity in the effect. Given that we fail to document systematic differences in expectations, in particular with respect to earnings, based on male and female control group students, we believe that this result is most likely explained by differential processing of information. Interestingly, among the few studies that analyze belief updating after an information intervention is the study by Kerr et al. (2020), and similar to our results on postgraduate decisions, they only find evidence for men adjusting their beliefs after the intervention.

Note that we asked the students again in the second and third follow-up, taking place 12 and about 26 month post our information treatment respectively. In both, we do not find any evidence for causal long-term effects of our intervention on perceived returns, highlighting the importance of timing in relation to when actual enrollment decisions are taken.[24]

2.5.3 Effects on direct enrollment intentions

In Table 2.4 we first present treatment effects on students' intentions to enroll into a postgraduate program measured 6 months after the treatment, i.e. in the first follow-up survey. The first column shows the overall effect, column 2 and 3 effects differentiated by gender, and columns 4 and 6 by students' academic background. We estimate all effects in Table 2.4 controlling for matching variables, direct enrollment intentions and postgraduate enrollment prior to treatment (at baseline).

Table 2.4 shows that students direct enrollment intentions decreases by 0.042 in the overall sample (see column 1). Compared to students' in the control group, where 50% intend to directly enroll in postgraduate studies, treated students are 4 percentage points (pp) less likely to pursue a master's degree directly after graduating from their bachelor's degree. Looking at the treatment effect separately by gender shows that this reduction in intentions is driven by male students. Treated male students are 16 pp less likely to intend to enroll in postgraduate studies after the treatment (see column 3 in Table 2.4). Considering that enrollment intentions of males in the control group are 26 pp higher than those of females, the treatment led to a reduction in the gender gap of enrollment intentions. This effect mirrors the effects found in Figure 2.2, as treated male students increase their belief of pecuniary returns to a bachelor's degree. We also estimate the treatment effects separately by academic background and continuing generation students are also less likely to intend to directly enroll in master's programs by 5 pp.

All effects on direct intention to enroll in postgraduate studies are statistically insignificant at conventional levels, apart from the effects on male students. Yet, the size of the overall effect is not small. We also studied the frequency of upward and downward changes in direct transition intentions between baseline and the first follow-up (Table 2.11). No significant effect of the intervention on the frequency of observed changes in enrollment intentions are found and the point estimate is close to zero. The intervention has reduced the number of upward changes and increased the number of downward changes by similar amounts – again, not statistically significant - which leads to a net effect very close to zero.

[24]In the second and third follow-up we also asked participants for expected min, mean and max earnings. Using these measures we confirm the finding that effects have faded out after twelve month.

While the results presented above refer to our preferred specification, we also provide results for five alternative control variable specifications in Tables 2.7 - 2.10 in appendix 2.B.1. The first result shows that the mean difference in the outcome between treatment and control group is -0.038 points. To control for finite sample imbalances, we are gradually adding sets of control variables to the regression. Controlling for baseline enrollment intentions and enrollment on top of the matching variables lowers standard errors considerably and increases the effect size to about -0.042. In a next step, we add background characteristics, e.g. controlling for the finite-sample imbalance in migration background, with negligible changes in point estimates and standard errors. Finally, we also control for three groups of study subjects. Again, this only leads to negligible changes in the results.

When accounting for randomization pair dummies in Table 2.12, as suggested by Bruhn and McKenzie (2009), estimates and standard errors only change slightly. We further show that traditional standard errors and standard errors based on randomization inference, following the approach by Young (2019), are almost the same.

2.5.4 Effects on actual postgraduate enrollment

In Table 2.4 we also present results on the last of our main outcomes and look at treatment effects on actual postgraduate enrollment one and two years after the information treatment. Other than the results on postgraduate enrollment intentions, these results are likely to be very dependent on the exact timing of the intervention. Initially, when the intervention was planned, it was assumed that a large share of students were likely to complete their undergraduate degree between the first and second follow-up. Yet, as the control group mean for postgraduate enrollment at the second follow-up shows, a relatively small share of students already and directly transitioned to postgraduate studies (28.2%). One year after the baseline survey, the majority of students are still enrolled in an undergraduate program, e.g. because they started studying later, they are more likely to be enrolled in programs of universities of applied sciences, switched majors or need longer than the population average for other reasons.[25] Yet, while the enrollment shares are still fairly low, there is a clear correlation between enrollment intentions and actual enrollment. Of the control group students who did not report direct enrollment intentions at baseline, only 6% were enrolled in a master's program after one year, while among those who reported direct enrollment intentions, the share is 37%. After two years, the respective shares are 21% and 53%. The correlation shows that expecting a link between impacts on intentions and enrollment is justified.

[25]See Figure 2.3 for a stylized visualisation of time to bachelor's degree and application period for master programs.

Treated students are 4 pp less likely to be enrolled in postgraduate studies in the winter term 2018. This translates into a reduction in enrollment by 15% compared to the control group. While not negligible in size, effects on actual enrollment after one year shown in Table 2.4 are far from being statistically significant.

While the pattern from the previous findings – stronger effects on males and students from academic family background – is visible one year after the treatment, the impact on actual postgraduate enrollment of men only becomes visible after two years, where the effect is much more pronounced (-12.4pp), yet still not statistically significant at conventional levels (p-value=0.12) .

In the last part of Table 2.4, we also look at the two last follow-ups jointly to see if pooling the sample increases statistical power. Yet, no relevant differences in standard errors can be observed. Here the outcome is defined as being enrolled in a postgraduate program at least once during the one and two year follow-ups.

2.6 Conclusion

This is the first study to present estimates for effects of information provision on beliefs about postgraduate returns, enrollment intentions, and realized enrollment. We show that students significantly updated their beliefs about postgraduate returns half a year later. In particular males downward-adjust expectations regarding the postgraduate premium. Moreover, we document corresponding changes in enrollment intentions six months after treatment. Moreover, we provide suggestive evidence that the effects of information on intention materialised into differences in realised postgraduate enrollment two years after initial treatment.

These results are important as they document that information frictions exist even for students already enrolled in undergraduate degrees. Moreover, the online-treatment could be scaled up at low costs.

On the other hand, we show that only groups of students for whom we find significant effects of the treatment on beliefs also show significant reactions in our enrollment measures. This highlights a general difficulty in providing systematic information about the role of beliefs in an experimental setup where research is bound in the analysis by the ethical requirement to only present truthful information to the students. One implication is that effects of information can only be estimated for groups where significant belief updating takes place. Our setting is fortuitous in this context since the two-tier system, while well established at the university-level, was still quite novel for the labour market. As a consequences, little information on long-run postgraduate earnings premia of older cohorts could be observed, thus providing a setting that gives scope for updating to take place. RCTs on the role of information for belief updating and postgraduate decisions as a result have particularly high demands on sample size to shed light on heterogeneity on the role of information, which requires significant belief updating across groups.

2.A Tables

Table 2.1: Descriptive statistics

Variable	Mean	SD	N
Matching variables (pre-inquiry)			
General intention	0.771	0.421	371
BestUp treatment group	0.305	0.461	371
Female	0.623	0.485	371
Pre-inquiry enrollment	0.865	0.342	371
GPA (categorical)	1.911	0.795	371
Baseline covariates			
Direct transition	0.461	0.499	371
General intention	0.768	0.423	371
Non-academic background	0.589	0.493	365
Migration background	0.467	0.500	368
Age (Juni 2018)	23.442	0.945	371
High School (Gymnasium)	0.302	0.460	371
Integrat. compreh. school	0.369	0.483	371
Vocational high school	0.329	0.470	371
GPA	2.327	0.593	335
Degree: not enrolled	0.102	0.304	371
Degree: bachelor	0.779	0.415	371
Degree: staatsexamen/diplom	0.057	0.231	371
Degree: master	0.046	0.209	371
Degree: art/n.a.	0.016	0.126	371
Total semester enrolled	5.466	1.837	356
1. Follow-up			
University	0.464	0.499	371
Applied University	0.259	0.439	371
Lehramt (teaching)	0.097	0.296	371
Subj.: Law, Business, Social Sci.	0.310	0.463	371
Subj.: Natural Sci., Engineering	0.253	0.436	371
Subj.: Other	0.173	0.378	371
Direct transition	0.478	0.500	322
General intention	0.739	0.440	322
2. Follow-up			
Postgraduate enrollment	0.256	0.437	293
3. Follow-up			
Postgraduate enrollment	0.408	0.492	311

Source: *Berliner-Studienberechtigten-Panel*, 2013-2020.

Table 2.2: Balance in baseline and pre-trial covariates

	Full sample			Pair fixed-effects	
	Control group mean	Treatment group difference	N	Treatment group difference	N
Matching variables (pre-inquiry)					
General intention	0.784	-0.026	371	-0.019	310
BestUp treatment group	0.308	-0.007	371	-0.026**	310
Female	0.627	-0.009	371	0.000	310
Pre-inquiry enrollment	0.870	-0.010	371	-0.006	310
GPA (categorical)	1.908	0.006	371	-0.013	310
Enrollment intentions (baseline)					
Direct transition	0.449	0.024	371	0.039	310
General intention	0.757	0.023	371	0.006	310
Background (baseline)					
Non-academic background	0.575	0.029	365	0.066	302
Migration background	0.404	0.125**	368	0.112**	304
Age (Juni 2018)	23.403	0.079	371	0.135	310
High School (Gymnasium)	0.286	0.031	371	-0.006	310
Integrat. compreh. school	0.405	-0.072	371	-0.077	310
Vocational high school	0.308	0.041	371	0.084	310
GPA	2.319	0.017	335	0.010	268
Enrollment (baseline)					
Degree: not enrolled	0.114	-0.022	371	-0.019	310
Degree: bachelor	0.757	0.044	371	0.052	310
Degree: staatsexamen/diplom	0.054	0.005	371	-0.013	310
Degree: master	0.054	-0.016	371	-0.013	310
Degree: art/n.a.	0.022	-0.011	371	-0.006	310
Total semester enrolled	5.474	-0.016	356	0.104	288

Notes: Treatment group differences using pair fixed-effects are based on a regression with pair fixed-effects omitting observations from incomplete pairs. Source: *Berliner-Studienberechtigten-Panel*, 2013-2018. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.3: Associations of enrollment intentions and perceived return

	Dependent variable: direct transition intention						
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Pecuniary aspect:							
Above avgerage income	0.14***					0.10**	0.10**
	(0.04)					(0.05)	(0.04)
Non-pecuniary aspects:							
Intell. challenging work		0.10**				0.07*	
		(0.04)				(0.04)	
Work-life balance			0.10**			0.10**	
			(0.05)			(0.05)	
Highly-skilled/managerial				0.08**		0.03	
				(0.04)		(0.04)	
Parents satisfied with job					0.01	-0.09	
					(0.05)	(0.06)	
Non-pecuniary ret. index							0.10*
							(0.06)
Others:							
Age (June 2018)	-0.05	-0.05	-0.06	-0.03	-0.05	-0.05	-0.04
	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)
Female	-0.23***	-0.24***	-0.19**	-0.22***	-0.22**	-0.23***	-0.23***
	(0.08)	(0.08)	(0.09)	(0.08)	(0.09)	(0.08)	(0.08)
Migration background	0.17**	0.18**	0.14*	0.17**	0.16*	0.19**	0.18**
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Non academic fam. backgr.	-0.08	-0.06	-0.08	-0.09	-0.08	-0.07	-0.07
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
F-test (pvalue): Joint significance of return measures						0.00	
Constant	1.74*	1.69	1.95*	1.42	1.76	1.73*	1.61
	(0.99)	(1.08)	(1.11)	(1.02)	(1.09)	(1.02)	(0.98)
N	147	147	147	147	147	147	147

Notes: This table shows the effects of step-wise regressions for direct transition intentions on perceived pecuniary and non-pecuniary returns. Dependent variable and return measures are from the first follow-up survey (6 months after treatment). Regressions are based on control group only. Robust standard errors in parentheses. Source: *Berliner-Studienberechtigten-Panel*, 2013-2018. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.4: Treatment effects on postgraduate enrollment intentions and enrollment

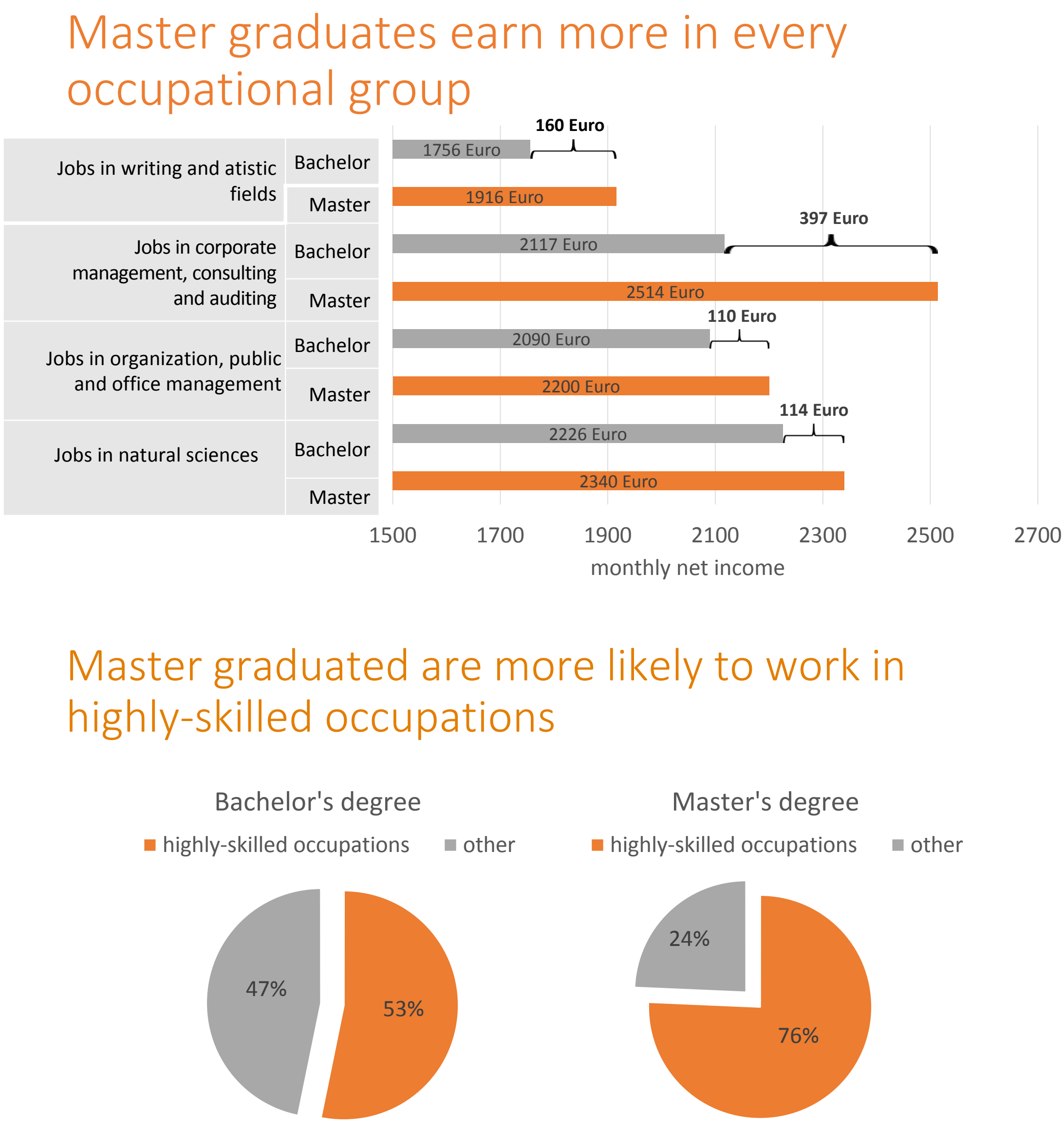
	Total sample	**Female**	**Male**	**Non-academic**	**Academic**
1. Follow-up (after 6 months): Direct transition intentions					
Treatment effect	-0.042	0.026	-0.156**	-0.003	-0.053
	(0.041)	(0.050)	(0.077)	(0.059)	(0.060)
Control group mean	*0.497*	*0.404*	*0.661*	*0.457*	*0.567*
N	322	206	116	189	130
2. Follow-up (after 1 year): Postgraduate enrollment					
Treatment effect	-0.043	-0.042	-0.063	-0.026	-0.039
	(0.042)	(0.052)	(0.074)	(0.056)	(0.073)
Control group mean	*0.282*	*0.273*	*0.300*	*0.276*	*0.305*
N	293	192	101	176	112
3. Follow-up (after 2 years): Postgraduate enrollment					
Treatment effect	-0.058	-0.029	-0.124	-0.034	-0.014
	(0.047)	(0.060)	(0.079)	(0.060)	(0.082)
Control group mean	*0.437*	*0.408*	*0.483*	*0.421*	*0.475*
N	311	192	119	187	120
2. or 3. Follow-up (within 2 years): Postgraduate enrollment					
Treatment effect	-0.035	-0.001	-0.105	-0.027	-0.004
	(0.045)	(0.057)	(0.075)	(0.058)	(0.078)
Control group mean	*0.408*	*0.373*	*0.469*	*0.404*	*0.431*
N	344	216	128	203	136
Pair fixed effects	No	No	No	No	No
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	Yes	Yes	Yes	Yes	Yes
Controls: Background	No	No	No	No	No
Controls: Study subject	No	No	No	No	No

Notes: All regressions control for matching variables (see Table 2.2) and direct and general enrollment intentions and postgraduate enrollment at baseline. We deal with missing information in control variables by setting these variables to a constant value and including a binary variable indicating missing values in control variables. Robust standard errors in parentheses. Source: *Berliner-Studienberechtigten-Panel*, 2013-2020. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix

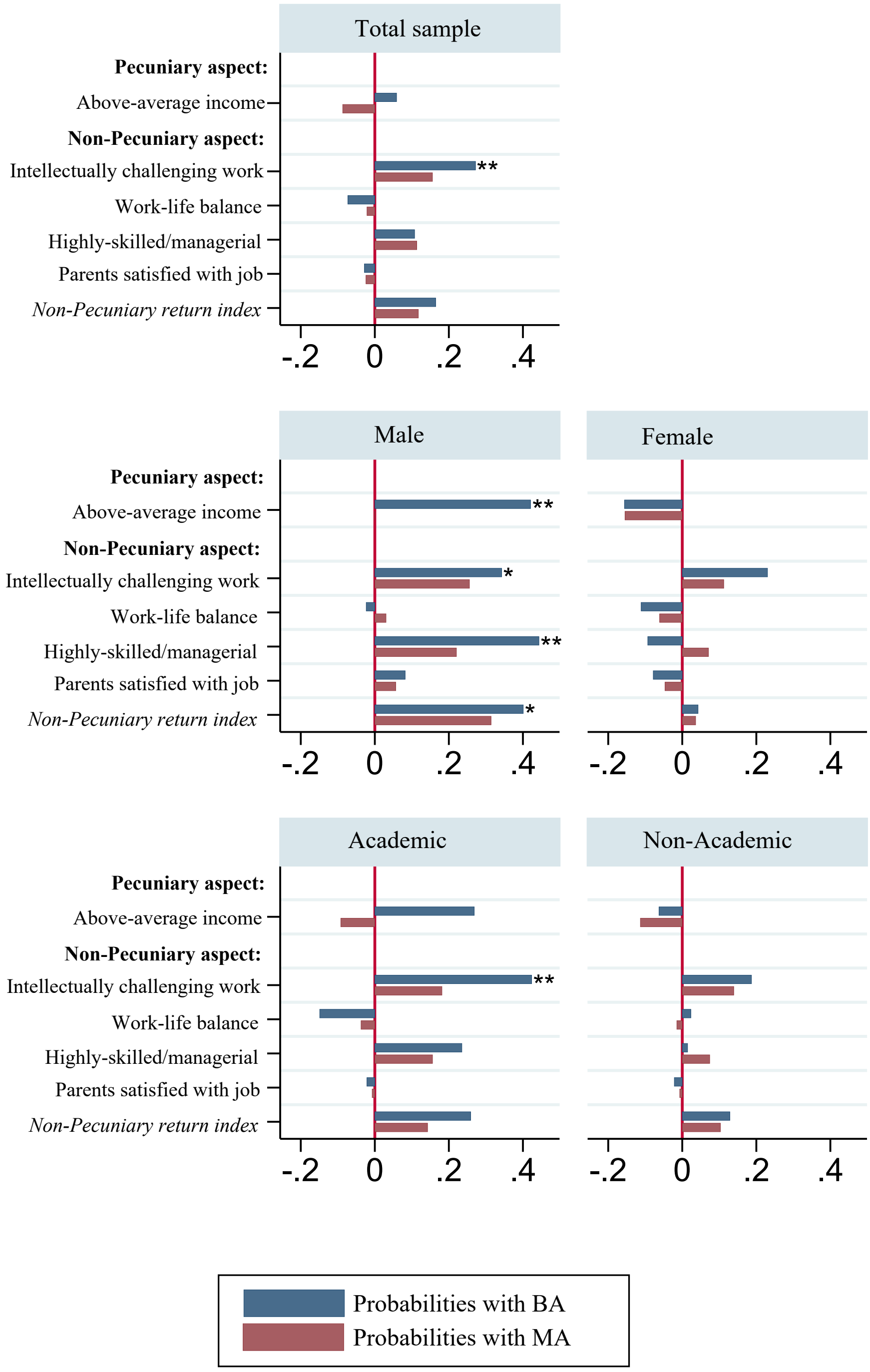
2.A.1 Figures

Figure 2.1: Example slides of online information module



Notes: This figure provides examples of the slides used in the online information module and shows two out of ten illustrative screens. The top figure shows income by field of education and degree type and the bottom figure the share of people working in a highly-skilled occupation by degree type. Both slides are translated from German. Examples of the original screens seen by students are included in appendix 2.B.2.

Figure 2.2: Treatment effects on standardized perceived probabilities with a bachelor's degree only and with a master's degree

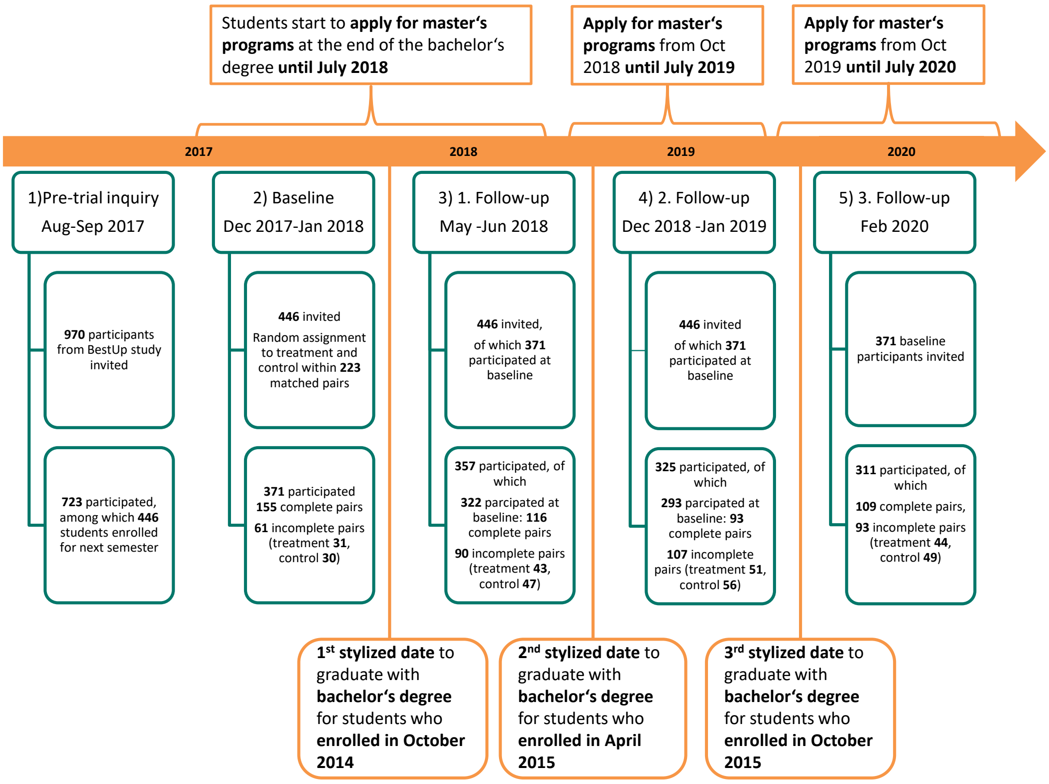


Notes: All outcome measures taken at first follow-up (6 months after treatment). All outcome measures are standardized using the control group. Figure shows treatment effects from a regression of the outcome measure on a treatment group indicator, also controlling for matching variables (see Table 2.2), direct and general enrollment intentions and postgraduate enrollment at baseline. Source: *Berliner-Studienberechtigten-Panel*, 2013-2018. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

2.B Appendix

2.B.1 Additional Figures

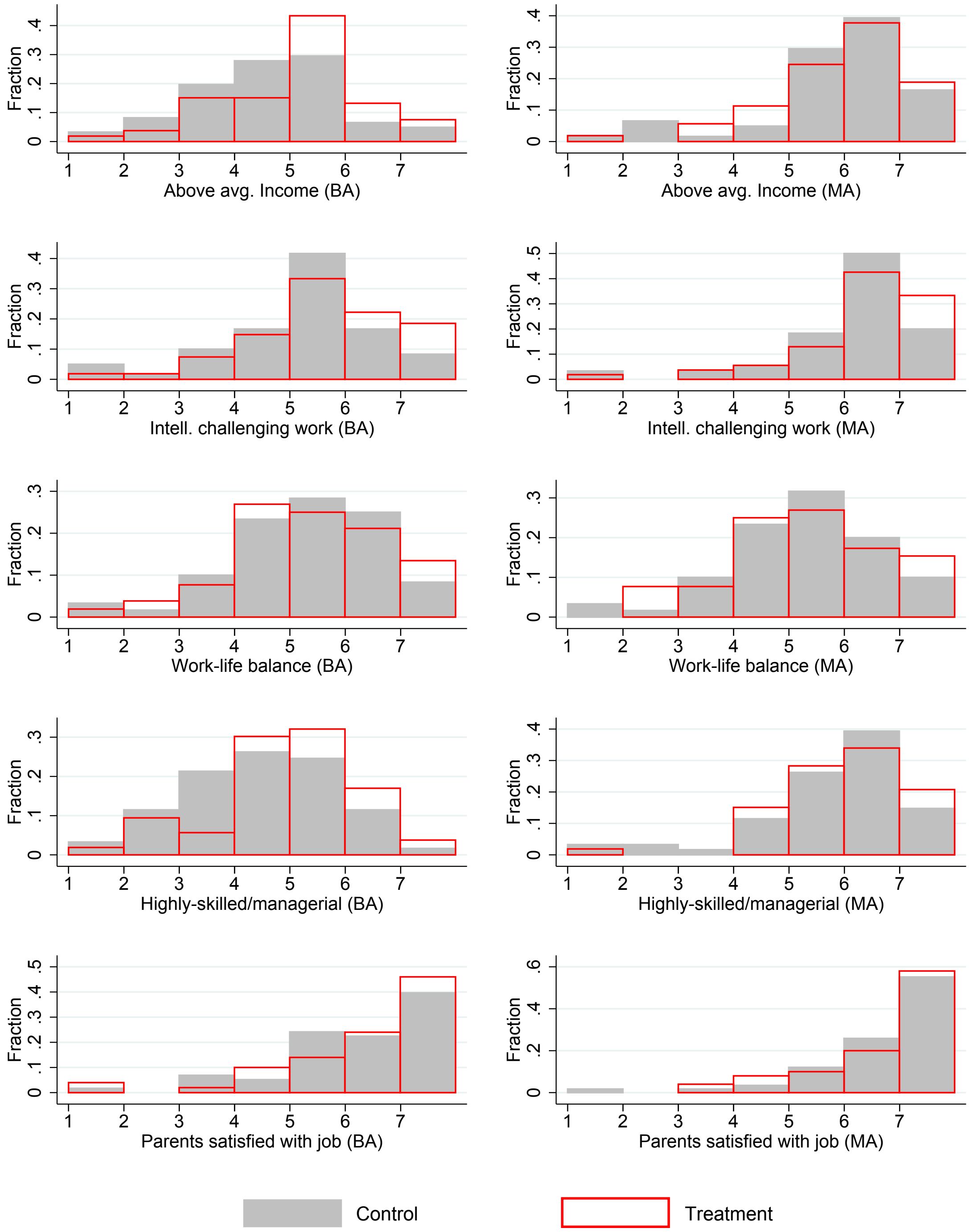
Figure 2.3: Trial profile and timeline including stylized bachelor graduation dates



Notes: This figure presents participation rates in each survey of the *Berliner-Studienberechtigten-Panel* from 2017-2020. We report participation rates related to the utilized randomization method pair-wise matching. We also report the number of participants per wave and in relation to baseline participation. For example, in the second follow-up 325 persons participated and of those 293 also participated in the baseline survey. The latter equals a response rate of 79% compared to baseline and 91% compared to the imminent wave (N=322). In contrast to the baseline and follow-up surveys, students did not receive any incentives to participate in the pre-trial inquiry in 2017.

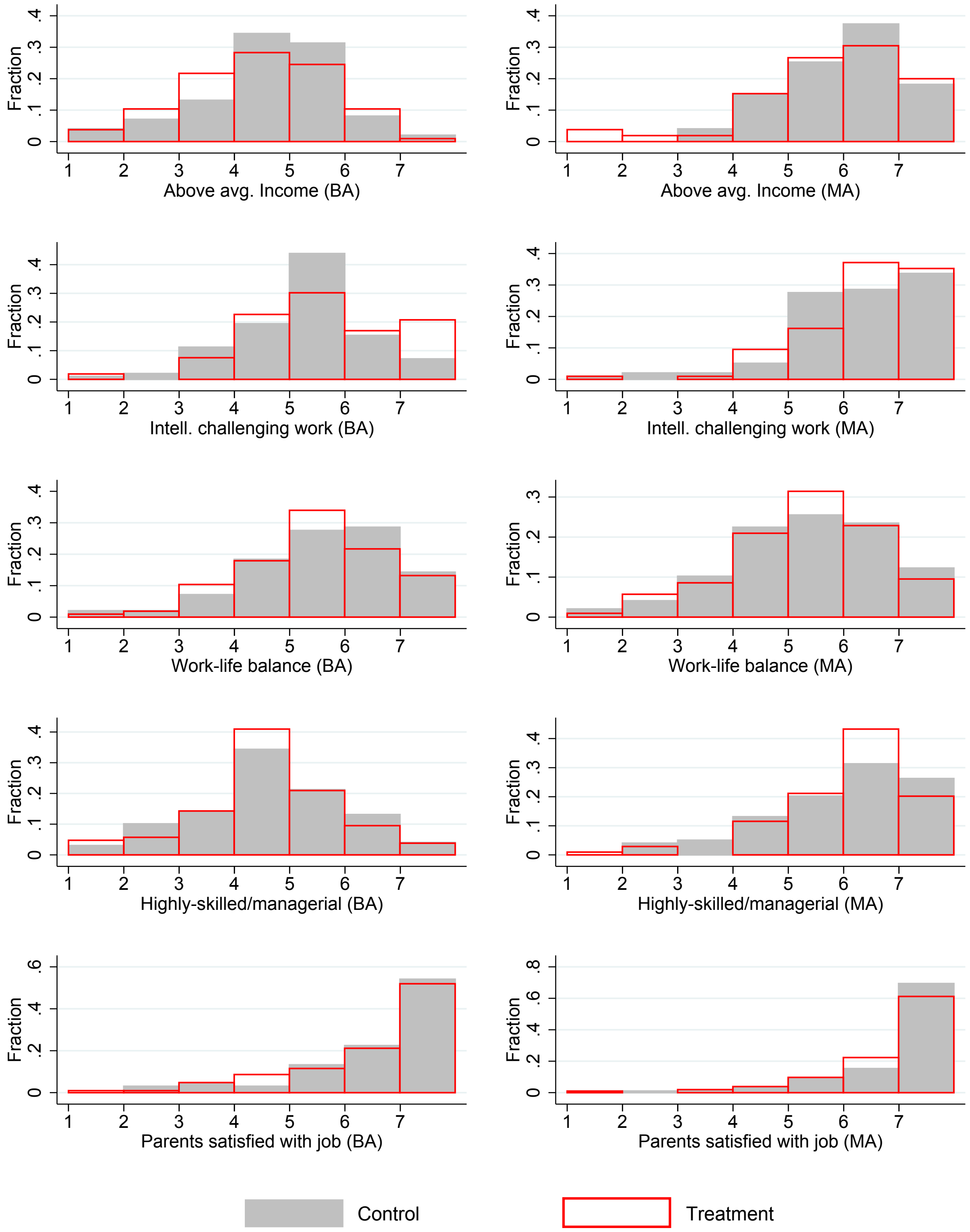
In addition, we include stylized dates where students given their enrollment date would graduate with a bachelor's degree in the German system, as the average time to degree in Germany comprises 7.2 semesters (Autorengruppe Bildungsberichterstattung, 2020). Given that the majority of students graduated high school in 2014 and either enrolled directly in fall 2014 or within one year in fall 2015, the stylized graduation dates are March 2018-March 2019.

Figure 2.4: Histograms of males perceived probabilities in treatment and control group



Notes: This figure shows distributions of non-standardized perceived probabilities measured on a 7-point likert-type scale from 1 (very unlikely) to 7 (very likely). The probabilities are measured separately for having a bachelor's degree only (BA) and having a master's degree (MA) for the categories to earn above average income (row 1), to do intellectually challenging work (row 2), to be able to combine work and family life (row 3), to work in a highly-skilled job or with managerial responsibility (row 4), and that parents are satisfied with their job (row 5). All measures are taken from the first follow-up survey (6 months after treatment). Source: *Berliner-Studienberechtigten-Panel*, 2018.

Figure 2.5: Histograms of females perceived probabilities in treatment and control group



Notes: This figure shows distributions of non-standardized perceived probabilities measured on a 7-point likert-type scale from 1 (very unlikely) to 7 (very likely). The probabilities are measured separately for having a bachelor's degree only (BA) and having a master's degree (MA) for the categories to earn above average income (row 1), to do intellectually challenging work (row 2), to be able to combine work and family life (row 3), to work in a highly-skilled job or with managerial responsibility (row 4), and that parents are satisfied with their job (row 5). All measures are taken from the first follow-up survey (6 months after treatment). Source: *Berliner-Studienberechtigten-Panel*, 2018.

2.B.2 Additional Tables

Table 2.5: Comparison of baseline sample with NEPS

	Best Up Baseline (2017/2018)		NEPS SC 4		NEPS SC 5	
Variable	Mean	N	Mean	N	Mean	N
Direct transition	0.46	371			0.44	1129
General intention	0.77	371	0.75	631	0.82	1129
Non-academic background	0.59	365	0.50	2360	0.63	1129
Migration background	0.47	368	0.22	2360	0.19	1129
Age (June 2018)	23.44	371	22.42	2360	24.21	1129
Academic high school	0.30	371	0.78	2360	0.75	1129
Comprehensive high school	0.37	371	0.07	2360	0.03	1129
Vocational academic high school	0.33	371	0.14	2360	0.13	1129
GPA	2.33	371	2.20	2360	2.24	1129
Degree: not enrolled	0.10	371	0.00	2360	0	1129
Degree: bachelor	0.78	371	0.84	2360	0.83	1129
Degree: *Staatsexamen/Diplom*	0.06	371	0.16	2360	0.06	1129
Degree: master	0.05	371	0.00	2360	0.11	1129
Degree: art/n.a.	0.02	371	0.00	2360	0	1129
Total semester enrolled	5.47	356	3.23	2360		

Notes: Source: This table uses data from the *Berliner-Studienberechtigten-Panel*, 2017-2018 and the National Educational Panel Study (NEPS): Starting Cohort Grade 9, doi:10.5157/NEPS:SC4:10.0.0 and Starting Cohort First-Year Students, doi:10.5157/NEPS:SC5:12.0.0, own calculations. From 2008 to 2013, NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.
The starting cohort 4 of the NEPS follows the educational pathway of students in grade 9 and higher into either university or vocational training. This sample consists of students who graduated from high school in 2014 or 2015. The variable “General intention” to start a master uses a restricted sample, namely that of bachelor students only, hence the smaller sample size. The starting cohort 5 comprises first-year students who started studying at a higher education institution in 2010. This sample was restricted to students who graduated from high school in 2010 and who could have finished their bachelor in 2013 at the earliest.

Table 2.6: Attrition of students to follow-up surveys

	1. Follow-up		2. Follow-up		3. Follow-up	
T: Treatment group	0.026	0.085	0.031	0.137	0.031	0.035
	(0.035)	(0.177)	(0.042)	(0.200)	(0.038)	(0.162)
Direct transition		-0.061		-0.067		-0.060
		(0.050)		(0.060)		(0.052)
BestUp treatment group		0.013		-0.028		0.090
		(0.053)		(0.060)		(0.062)
Female		-0.050		-0.143**		0.024
		(0.052)		(0.063)		(0.054)
GPA		0.061		0.048		0.025
		(0.046)		(0.057)		(0.041)
Degree: not enrolled		-0.059		0.080		0.026
		(0.072)		(0.109)		(0.090)
T*Direct transition		0.034		0.041		0.074
		(0.073)		(0.086)		(0.076)
T*BestUp treatment group		0.060		-0.015		-0.087
		(0.080)		(0.090)		(0.084)
T*Female		-0.033		0.050		-0.015
		(0.077)		(0.091)		(0.077)
T*GPA		-0.038		-0.069		-0.015
		(0.067)		(0.077)		(0.063)
T*Degree: not enrolled		0.156		0.122		0.363**
		(0.124)		(0.165)		(0.155)
Control group mean attrition	0.119		0.195		0.146	
Joint F-tests (p-values):						
Baseline controls		0.373		0.133		0.413
T interactions with baseline controls		0.737		0.881		0.150
N	371	371	371	371	371	371

Notes: This table shows OLS regressions with attrition at first, second and third follow-up as dependent variable using all baseline participants. Missing baseline control variables are replaced by the control group mean. Robust standard errors in parentheses. Source: *Berliner-Studienberechtigten-Panel*, 2013-2020. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.7: Robustness checks for treatment effects on direct transition intentions

	Total sample	**Female**	**Male**	**Non-academic**	**Academic**
1. Follow-up (after 6 months): Direct transition intentions					
Treatment effect	-0.038	0.057	-0.205**	-0.005	-0.091
	(0.056)	(0.069)	(0.091)	(0.073)	(0.088)
Controls: Matching variables	No	No	No	No	No
Controls: Enrolment intentions	No	No	No	No	No
Controls: Background	No	No	No	No	No
Controls: Study subject	No	No	No	No	No
Treatment effect	-0.030	0.063	-0.206**	-0.000	-0.029
	(0.053)	(0.065)	(0.090)	(0.071)	(0.087)
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	No	No	No	No	No
Controls: Background	No	No	No	No	No
Controls: Study subject	No	No	No	No	No
Treatment effect	-0.042	0.026	-0.156**	-0.003	-0.053
	(0.041)	(0.050)	(0.077)	(0.059)	(0.060)
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	Yes	Yes	Yes	Yes	Yes
Controls: Background	No	No	No	No	No
Controls: Study subject	No	No	No	No	No
Treatment effect	-0.038	0.026	-0.145*	0.012	-0.049
	(0.042)	(0.052)	(0.080)	(0.060)	(0.062)
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	Yes	Yes	Yes	Yes	Yes
Controls: Background	Yes	Yes	Yes	Yes	Yes
Controls: Study subject	No	No	No	No	No
Treatment effect	-0.038	0.022	-0.151*	0.007	-0.048
	(0.042)	(0.051)	(0.080)	(0.061)	(0.062)
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	Yes	Yes	Yes	Yes	Yes
Controls: Background	Yes	Yes	Yes	Yes	Yes
Controls: Study subject	Yes	Yes	Yes	Yes	Yes
N	322	206	116	189	130

Notes: Table compares treatment effects using different sets of control variables. See Table 2.2 for list of matching variables which are used as controls. Enrollment intentions are controlled for by controlling for direct and general enrollment intentions and postgraduate enrollment at baseline. Background control variables are migration background, non-academic family background, gender, age, high school type and GPA. To control for study subjects the three groups listed in Table 2.1 are used. We deal with missing information in control variables by setting these variables to a constant value and including a binary variable indicating missing values in control variables. Robust standard errors in parentheses. Source: *Berliner-Studienberechtigten-Panel*, 2013-2018. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.8: Robustness checks for treatment effects on postgraduate enrollment (2. follow-up)

	Total sample	Female	Male	Non-academic	Academic
2. Follow-up (after 1 year): Postgraduate enrollment					
Treatment effect	-0.053	-0.036	-0.084	-0.040	-0.098
	(0.051)	(0.063)	(0.088)	(0.066)	(0.083)
Controls: Matching variables	No	No	No	No	No
Controls: Enrolment intentions	No	No	No	No	No
Controls: Background	No	No	No	No	No
Controls: Study subject	No	No	No	No	No
Treatment effect	-0.051	-0.039	-0.094	-0.037	-0.048
	(0.048)	(0.059)	(0.085)	(0.065)	(0.088)
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	No	No	No	No	No
Controls: Background	No	No	No	No	No
Controls: Study subject	No	No	No	No	No
Treatment effect	-0.043	-0.042	-0.063	-0.026	-0.039
	(0.042)	(0.052)	(0.074)	(0.056)	(0.073)
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	Yes	Yes	Yes	Yes	Yes
Controls: Background	No	No	No	No	No
Controls: Study subject	No	No	No	No	No
Treatment effect	-0.040	-0.053	-0.069	-0.016	-0.041
	(0.042)	(0.053)	(0.090)	(0.056)	(0.074)
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	Yes	Yes	Yes	Yes	Yes
Controls: Background	Yes	Yes	Yes	Yes	Yes
Controls: Study subject	No	No	No	No	No
Treatment effect	-0.038	-0.058	-0.081	-0.013	-0.049
	(0.044)	(0.055)	(0.094)	(0.060)	(0.078)
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	Yes	Yes	Yes	Yes	Yes
Controls: Background	Yes	Yes	Yes	Yes	Yes
Controls: Study subject	Yes	Yes	Yes	Yes	Yes
N	293	192	101	176	112

Notes: See notes of Table 2.7.

Table 2.9: Robustness checks for treatment effects on postgraduate enrollment (3. follow-up)

	Total sample	**Female**	**Male**	**Non-academic**	**Academic**
3. Follow-up (after 2 years):	**Postgraduate enrollment**				
Treatment effect	-0.058	-0.015	-0.127	-0.051	-0.069
	(0.056)	(0.071)	(0.090)	(0.072)	(0.091)
Controls: Matching variables	No	No	No	No	No
Controls: Enrolment intentions	No	No	No	No	No
Controls: Background	No	No	No	No	No
Controls: Study subject	No	No	No	No	No
Treatment effect	-0.051	-0.023	-0.126	-0.044	-0.005
	(0.051)	(0.064)	(0.086)	(0.067)	(0.090)
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	No	No	No	No	No
Controls: Background	No	No	No	No	No
Controls: Study subject	No	No	No	No	No
Treatment effect	-0.058	-0.029	-0.124	-0.034	-0.014
	(0.047)	(0.060)	(0.079)	(0.060)	(0.082)
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	Yes	Yes	Yes	Yes	Yes
Controls: Background	No	No	No	No	No
Controls: Study subject	No	No	No	No	No
Treatment effect	-0.043	-0.029	-0.111	-0.022	0.005
	(0.047)	(0.060)	(0.088)	(0.061)	(0.078)
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	Yes	Yes	Yes	Yes	Yes
Controls: Background	Yes	Yes	Yes	Yes	Yes
Controls: Study subject	No	No	No	No	No
Treatment effect	-0.044	-0.030	-0.118	-0.028	0.009
	(0.048)	(0.061)	(0.088)	(0.062)	(0.080)
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	Yes	Yes	Yes	Yes	Yes
Controls: Background	Yes	Yes	Yes	Yes	Yes
Controls: Study subject	Yes	Yes	Yes	Yes	Yes
N	311	192	119	187	120

Notes: See notes of Table 2.7.

Table 2.10: Robustness checks for treatment effects on postgraduate enrollment (2. or 3. follow-up)

	Total sample	**Female**	**Male**	**Non-academic**	**Academic**
2. or 3. Follow-up (within 2 years): Postgraduate enrollment					
Treatment effect	-0.037	0.005	-0.109	-0.039	-0.056
	(0.053)	(0.066)	(0.087)	(0.069)	(0.085)
Controls: Matching variables	No	No	No	No	No
Controls: Enrolment intentions	No	No	No	No	No
Controls: Background	No	No	No	No	No
Controls: Study subject	No	No	No	No	No
Treatment effect	-0.034	0.001	-0.115	-0.035	-0.002
	(0.049)	(0.061)	(0.082)	(0.065)	(0.084)
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	No	No	No	No	No
Controls: Background	No	No	No	No	No
Controls: Study subject	No	No	No	No	No
Treatment effect	-0.035	-0.001	-0.105	-0.027	-0.004
	(0.045)	(0.057)	(0.075)	(0.058)	(0.078)
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	Yes	Yes	Yes	Yes	Yes
Controls: Background	No	No	No	No	No
Controls: Study subject	No	No	No	No	No
Treatment effect	-0.024	-0.012	-0.100	-0.012	0.015
	(0.045)	(0.056)	(0.088)	(0.059)	(0.075)
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	Yes	Yes	Yes	Yes	Yes
Controls: Background	Yes	Yes	Yes	Yes	Yes
Controls: Study subject	No	No	No	No	No
Treatment effect	-0.024	-0.014	-0.108	-0.013	0.009
	(0.046)	(0.057)	(0.087)	(0.060)	(0.076)
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	Yes	Yes	Yes	Yes	Yes
Controls: Background	Yes	Yes	Yes	Yes	Yes
Controls: Study subject	Yes	Yes	Yes	Yes	Yes
N	344	216	128	203	136

Notes: See notes of Table 2.7.

Table 2.11: Treatment effects on changes in direct transition intentions

	Total sample	**Female**	**Male**	**Non-academic**	**Academic**
1. Follow-up (after 6 months)					
Direct transition intentions					
Treatment effect	-0.030	0.063	-0.206**	-0.000	-0.029
	(0.053)	(0.065)	(0.090)	(0.071)	(0.087)
Control group mean	*0.497*	*0.404*	*0.661*	*0.457*	*0.567*
Direct transition intentions: upward change					
Treatment effect	-0.023	-0.007	-0.048	-0.013	-0.052
	(0.031)	(0.034)	(0.063)	(0.044)	(0.056)
Control group mean	*0.0982*	*0.0673*	*0.153*	*0.0851*	*0.119*
Direct transition intentions: downward change					
Treatment effect	0.032	0.007	0.071	0.009	0.009
	(0.030)	(0.040)	(0.050)	(0.045)	(0.032)
Control group mean	*0.0675*	*0.0865*	*0.0339*	*0.0957*	*0.0299*
Direct transition intentions: upward or downward change					
Treatment effect	0.009	-0.001	0.023	-0.004	-0.043
	(0.042)	(0.051)	(0.075)	(0.059)	(0.062)
Control group mean	*0.166*	*0.154*	*0.186*	*0.181*	*0.149*
N	322	206	116	189	130
Pair fixed effects	No	No	No	No	No
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	No	No	No	No	No
Controls: Background	No	No	No	No	No
Controls: Study subject	No	No	No	No	No

Notes: Table compares treatment effects on changes in direct transition intentions between baseline and the first follow-up. The first panel shows effects on the binary variable for direct transition intentions. The second (third) variable shows effects on a binary variable measuring only upward (downward) transitions. The fourth variable shows effects on a binary variable for changes in any direction. See Table 2.2 for list of matching variables which are used as controls. Robust standard errors in parentheses. Source: *Berliner-Studienberechtigten-Panel*, 2013-2018. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.12: Treatment effects using complete pairs only

	Total sample	Female	Male	Non-academic	Academic
1. Follow-up (after 6 months): Direct transition intentions					
Treatment effect	-0.033	0.034	-0.152*	0.084	-0.007
	(0.050)	(0.060)	(0.087)	(0.088)	(0.105)
p-value	*0.510*	*0.573*	*0.090*	*0.348*	*0.950*
RI based p-value	*0.497*	*0.567*	*0.085*	*0.330*	*0.930*
N used	232	150	82	82	36
N dropped	90	56	34	107	94
2. Follow-up (after 1 year): Postgraduate enrollment					
Treatment effect	-0.038	-0.047	-0.063	-0.068	-0.068
	(0.052)	(0.070)	(0.082)	(0.100)	(0.076)
p-value	*0.466*	*0.506*	*0.447*	*0.501*	*0.395*
RI based p-value	*0.453*	*0.499*	*0.438*	*0.508*	*0.630*
N used	186	126	60	60	30
N dropped	107	66	41	116	82
3. Follow-up (after 2 years): Postgraduate enrollment					
Treatment effect	-0.034	0.021	-0.103	-0.044	0.023
	(0.058)	(0.081)	(0.079)	(0.096)	(0.151)
p-value	*0.561*	*0.797*	*0.200*	*0.651*	*0.883*
RI based p-value	*0.559*	*0.797*	*0.213*	*0.684*	*0.874*
N used	218	134	84	74	34
N dropped	93	58	35	113	86
2. or 3. Follow-up (within 2 years): Postgraduate enrollment					
Treatment effect	-0.025	0.033	-0.115	-0.075	0.068
	(0.054)	(0.074)	(0.073)	(0.101)	(0.141)
p-value	*0.638*	*0.654*	*0.124*	*0.466*	*0.638*
RI based p-value	*0.638*	*0.643*	*0.127*	*0.498*	*0.615*
N used	266	168	98	84	42
N dropped	78	48	30	119	94
Pair fixed effects	Yes	Yes	Yes	Yes	Yes
Controls: Matching variables	Yes	Yes	Yes	Yes	Yes
Controls: Enrolment intentions	Yes	Yes	Yes	Yes	Yes
Controls: Background	No	No	No	No	No
Controls: Study subject	No	No	No	No	No

Notes: This table shows regressions using pair fixed effects and controlling for matching variables and baseline enrollment intentions. Observations with incomplete pairs are dropped from the regressions. See Table 2.2 for list of matching variables which are used as controls. Enrollment intentions are controlled for by including direct and general enrollment intentions and postgraduate enrollment at baseline. Robust standard errors in parentheses. Corresponding conventional p-values are denoted as *p-value*. *RI (randomization inference) based p-values* are randomization-t p-values using the Stata command randcmd (Young, 2019) with 2000 randomization iterations. Source: *Berliner-Studienberechtigten-Panel*, 2013-2020. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.13: Treatment effects on standardized perceived probabilities and returns

	Total sample	**Female**	**Male**	**Non-academic**	**Academic**
How do you rate the probability...					
. . . to earn an above-average income with. . .					
bachelor's degree?	0.06	-0.16	0.42**	-0.06	0.27
master's degree?	-0.09	-0.15	-0.00	-0.11	-0.09
Difference	-0.15	0.00	-0.42**	-0.05	-0.36**
. . . to do intellectually challenging work with. . .					
bachelor's degree?	0.27**	0.23	0.34*	0.19	0.42**
master's degree?	0.16	0.11	0.26	0.14	0.18
Difference	-0.12	-0.12	-0.09	-0.05	-0.24
. . . to be able to combine work and family life with. . .					
bachelor's degree?	-0.07	-0.11	-0.02	0.02	-0.15
master's degree?	-0.02	-0.06	0.03	-0.01	-0.04
Difference	0.05	0.05	0.05	-0.04	0.11
. . . to work in a highly-skilled job or with managerial responsibility with. . .					
bachelor's degree?	0.11	-0.09	0.44**	0.01	0.24
master's degree?	0.11	0.07	0.22	0.07	0.16
Difference	0.01	0.16	-0.22	0.06	-0.08
. . . that parents will be satisfied with your job with. . .					
bachelor's degree?	-0.03	-0.08	0.08	-0.02	-0.02
master's degree?	-0.02	-0.05	0.06	-0.01	-0.01
Difference	0.00	0.03	-0.03	0.01	0.01
Non-precuniary return index					
bachelor's degree?	0.16	0.04	0.40*	0.13	0.26
master's degree?	0.12	0.04	0.31	0.10	0.14
Difference	-0.05	-0.00	-0.09	-0.03	-0.12
N	322	206	116	189	130

Notes: All outcome measures are taken from the first follow-up survey (6 months after treatment). Table shows treatment effects from a regression of the outcome measure on a treatment group indicator, also controlling for matching variables (see Table 2.2), direct and general enrollment intentions and postgraduate enrollment at baseline. Source: *Berliner-Studienberechtigten-Panel*, 2013-2018. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.14: Control group means and standard deviations of perceived probabilities by degree

	Total sample	**Female**	**Male**	**Diff: Male -Female**	**Non-academic**	**Academic**	**Diff: Academic -Non-academic**
How do you rate the probability...							
(1 very unlikely - ... - 7 very likely)							
...to earn an above-average income with...							
bachelor's degree?	4.10	4.12	4.07	-0.05	4.04	4.15	0.11
	[1.28]	[1.26]	[1.32]	*(.832)*	[1.28]	[1.28]	*(.609)*
master's degree?	5.45	5.46	5.43	-0.03	5.40	5.52	0.12
	[1.18]	[1.10]	[1.31]	*(.861)*	[1.15]	[1.22]	*(.542)*
...to do intellectually challenging work with...							
bachelor's degree?	4.77	4.80	4.72	-0.08	4.79	4.72	-0.07
	[1.28]	[1.21]	[1.39]	*(.709)*	[1.15]	[1.43]	*(.735)*
master's degree?	5.71	5.77	5.60	-0.17	5.73	5.73	0
	[1.28]	[1.26]	[1.32]	*(.465)*	[1.14]	[1.40]	*(.982)*
...to be able to combine work and family life with...							
bachelor's degree?	5.04	5.17	4.83	-0.34	4.89	5.22	0.33
	[1.35]	[1.34]	[1.34]	*(.150)*	[1.26]	[1.45]	*(.150)*
master's degree?	4.85	4.90	4.75	-0.15	4.70	5.03	0.33
	[1.40]	[1.43]	[1.36]	*(.551)*	[1.43]	[1.34]	*(.161)*
...to work in a highly-skilled job or with managerial responsibility with...							
bachelor's degree?	4.08	4.22	3.85	-0.37	4.06	4.10	0.04
	[1.36]	[1.35]	[1.35]	*(.115)*	[1.34]	[1.39]	*(.871)*
master's degree?	5.43	5.52	5.28	-0.24	5.45	5.39	-0.06
	[1.35]	[1.33]	[1.39]	*(.306)*	[1.41]	[1.31]	*(.803)*
...that parents will be satisfied with your job with...							
bachelor's degree?	5.97	6.11	5.73	-0.38	5.95	6.02	0.07
	[1.33]	[1.28]	[1.37]	*(.097)*	[1.28]	[1.38]	*(.769)*
master's degree?	6.31	6.40	6.16	-0.24	6.33	6.29	-0.04
	[1.18]	[1.15]	[1.22]	*(.240)*	[1.05]	[1.35]	*(.857)*
Non-precuniary return index							
bachelor's degree?	-0.08	0.04	-0.29	-0.33	-0.12	-0.04	0.08
	[0.98]	[0.87]	[1.13]	*(.058)*	[0.87]	[1.09]	*(.647)*
master's degree?	-0.06	0.06	-0.25	-0.31	-0.09	0.00	0.09
	[1.00]	[0.82]	[1.24]	*(.086)*	[0.95]	[1.09]	*(.618)*
Observations	146	92	54		83	61	

Notes: Perceived probabilities are measured on a 7-point likert-type scale from 1 (very unlikely) to 7 (very likely). The measures are not standardized and based on control group only. Standard deviations in square brackets. P-values from t-test for significance of differences in parentheses. All measures are taken from first follow-up survey (6 months after treatment). Source: *Berliner-Studienberechtigten-Panel*, 2018.

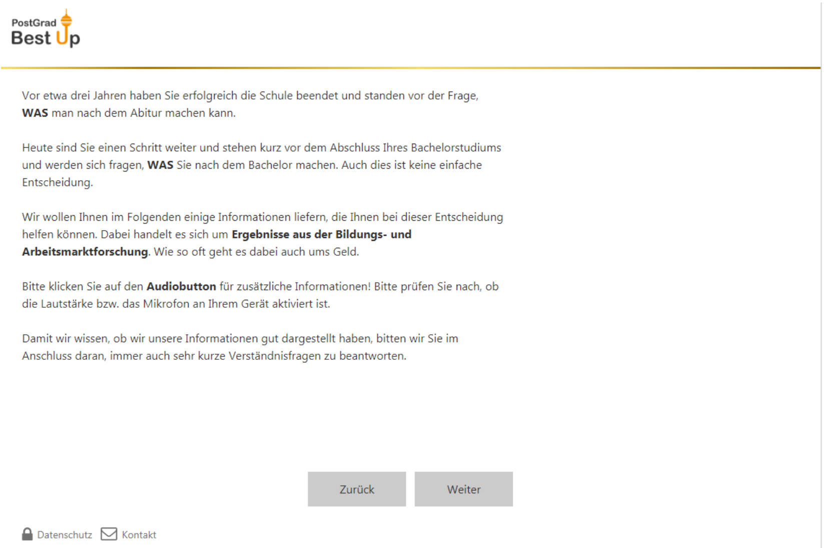
Table 2.15: Control group levels of self-reported preferences

	Female	**Male**	**Difference**	**p-value**
How important is it for you to have...				
(1 not at all-...-5 very much)				
Pecuniary aspect:				
Above avg. Income	3.536	3.727	-0.191	0.217
	(1.020)	(0.937)		
Non-pecuniary aspects:				
Intell. challenging work	3.691	3.788	-0.097	0.532
	(1.020)	(0.953)		
Work-life balance	4.518	4.045	0.473	0.001
	(0.854)	(0.983)		
Highly-skilled/managerial	3.018	3.455	-0.436	0.014
	(1.204)	(0.980)		
Parents satisfied with job	2.518	2.273	0.245	0.221
	(1.311)	(1.235)		
N	110	66		

Notes: Results are based on control group only. All measures are taken from first follow-up survey (6 months after treatment). Standard deviations in parentheses. Source: *Berliner-Studienberechtigten-Panel*, 2018.

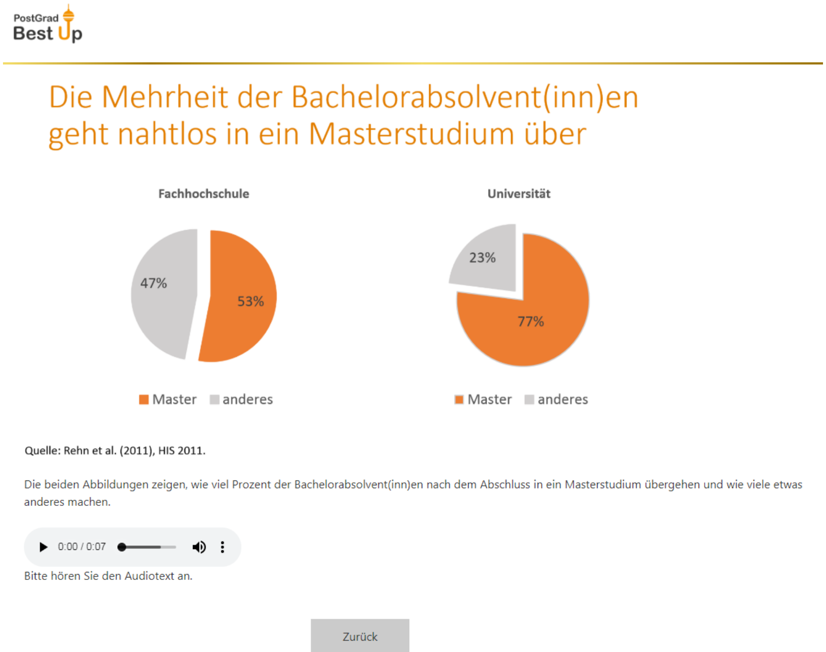
2.B.3 Information Treatment Slides and Text

Figure 2.6: Slide 1 of the information shown to the treatment group



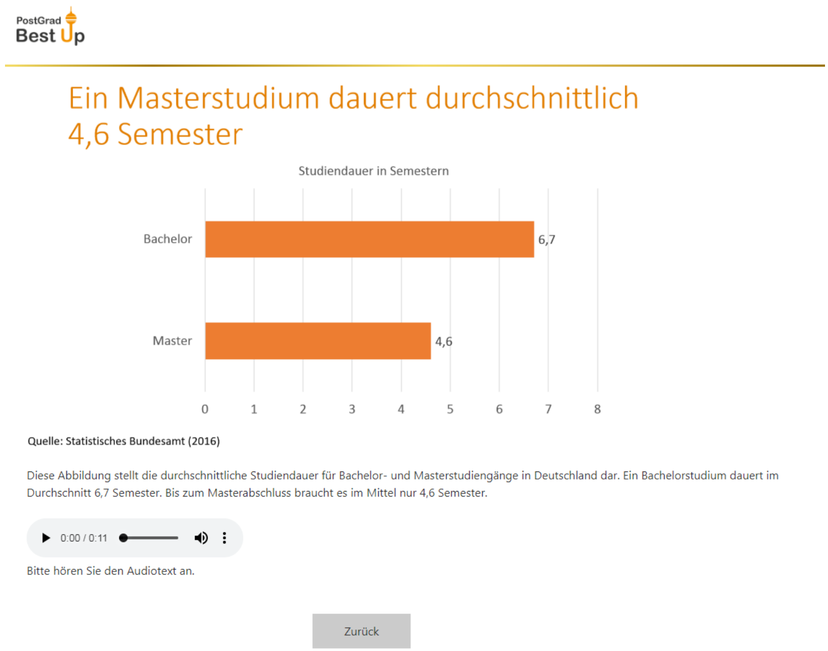
For translation of all slides see Section 2.B.5 below.

Figure 2.7: Slide 2 of the information shown to the treatment group



Text below figure: These two figures show the share of bachelor graduates that directly enroll into a master's program and the share of those who do something else after graduation. *Audio Slide 2:*Like the bachelor's program, the master's program is usually tuition-free.

Figure 2.8: Slide 3 of the information shown to the treatment group



Text below figure: This figure depicts the average time to degree for bachelor's and master's programs in Germany. A bachelor's program lasts on average 6.7 semesters. A master's degree takes only 4.6 semesters on average.
Audio Slide 3: On average, a master's degree is shorter than a bachelor's degree and, as scientific studies show, the majority of students study within the fixed number of semesters needed to complete the degree.

Figure 2.9: 1^{st} comprehension question about slide 4 of the information shown to the treatment group

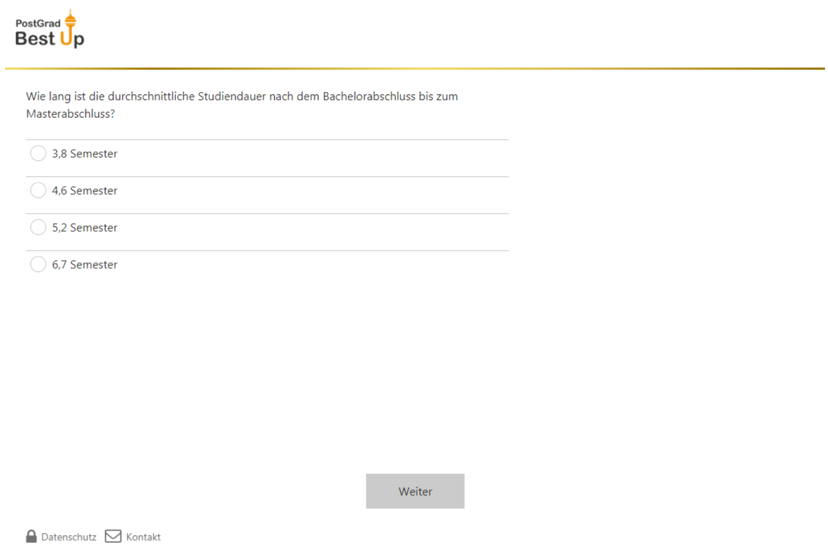
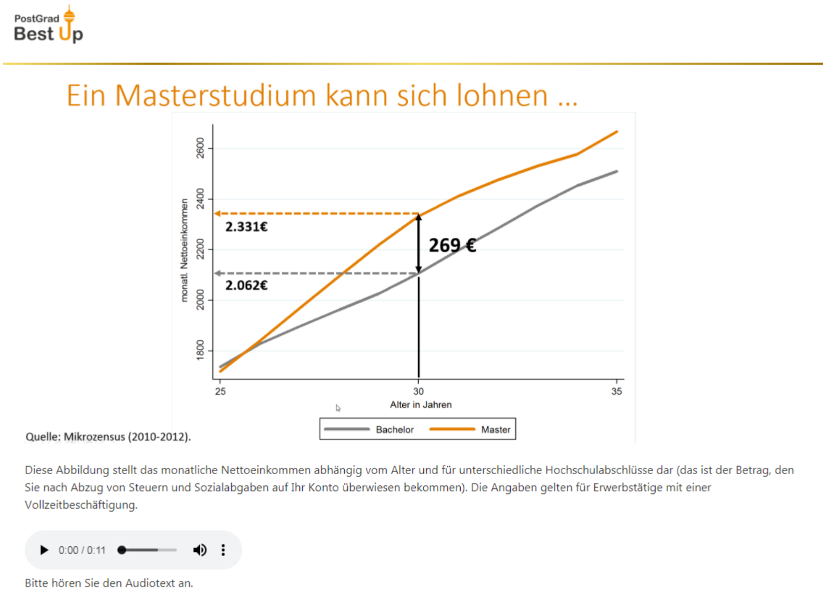


Figure 2.10: Slide 5 of the information shown to the treatment group



Text below figure: This figure shows the net monthly income (that is the amount you get transferred to your account after the deduction of taxes and social contributions) by degree type. The data applies to full-time employees.
Audio Slide 5: For instance, the average net monthly income with a master's degree at the age of 30 is 269 Euro higher compared to a bachelor's degree.

Figure 2.11: 2^{nd} comprehension question about Slide 5 of the information shown to the treatment group

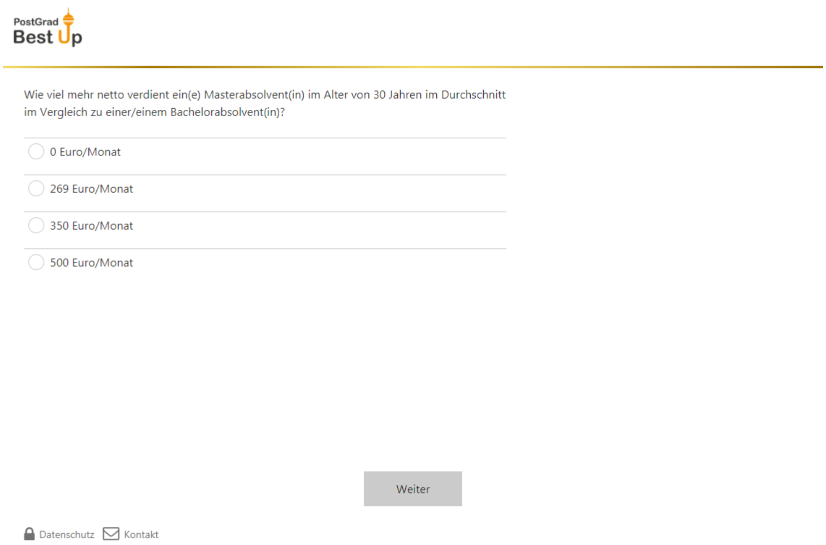
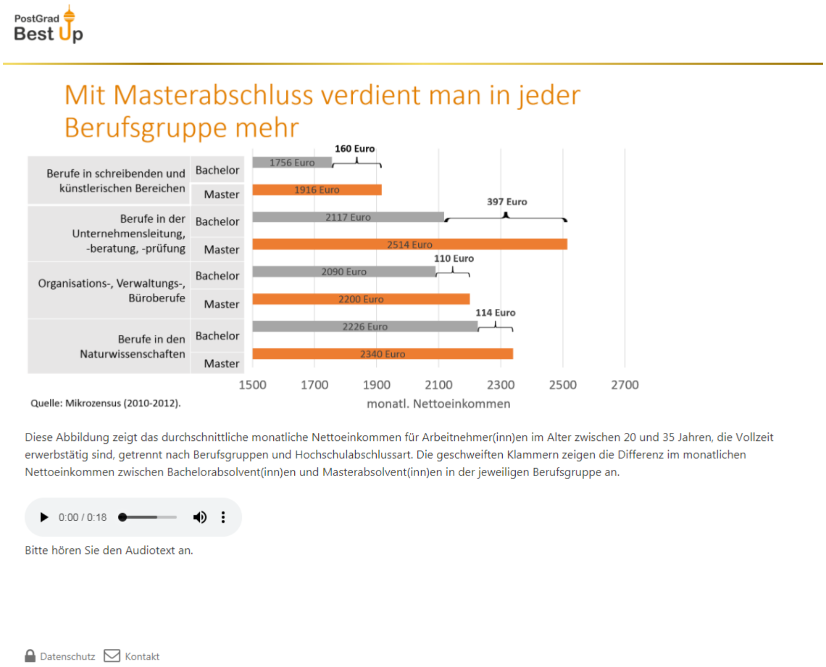


Figure 2.12: Slide 7 of the information shown to the treatment group



Text below figure: This figure shows the average net monthly earnings of full-time employees between 20 and 30 years of age by occupational group and degree type. The curly brackets indicate the difference in net monthly earnings between bachelor and master graduates by occupational group.
Audio Slide 7: In each occupational group master graduates earn more than bachelor graduates. For instance, a full-time employee with a master's degree in a consulting firm does on average earn almost 400 euro per month more than a full-time employee with a bachelor's degree.

Figure 2.13: 3^{rd} comprehension question about slide 7 of the information shown to the treatment group

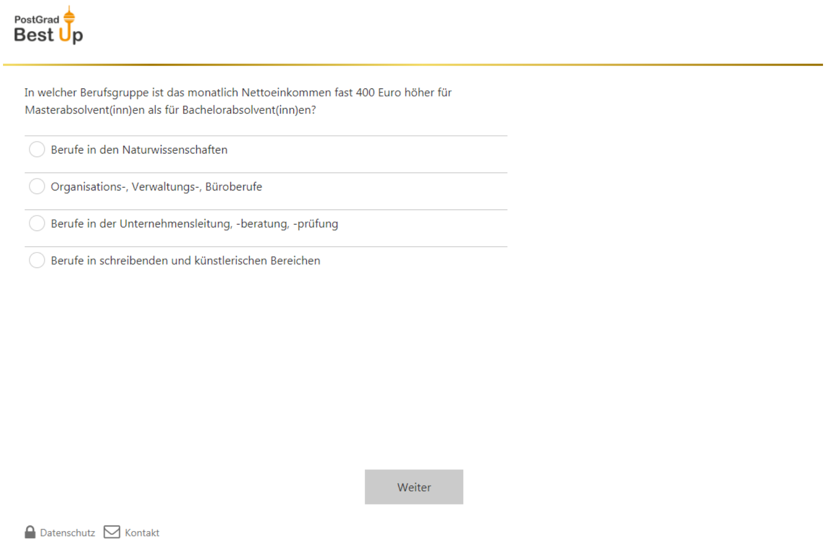
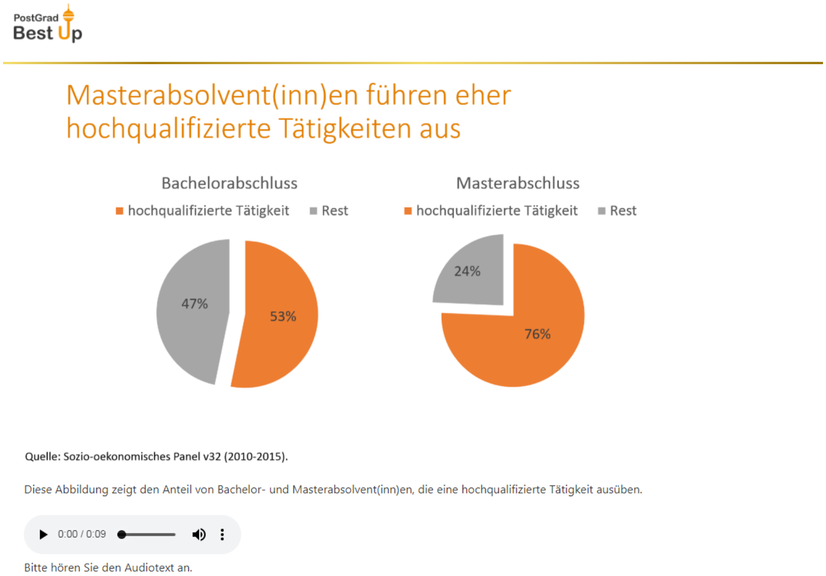


Figure 2.14: Slide 8 of the information shown to the treatment group



Text below figure: This figure depicts the share of bachelor and master graduates that work in highly skilled occupations.
Audio Slide 8: Master graduates are more likely to work autonomously and much more likely to assign tasks to other colleagues.

Figure 2.15: 4^{th} comprehension question about slide 8 of the information shown to the treatment group

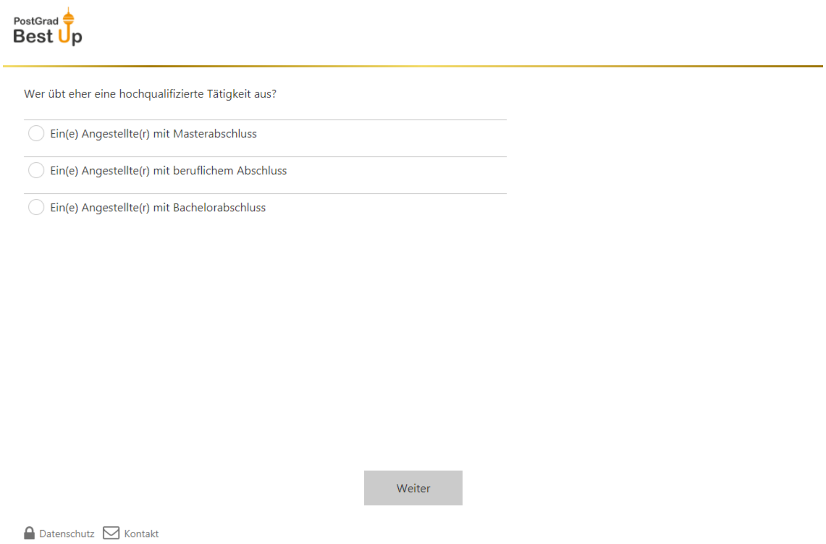
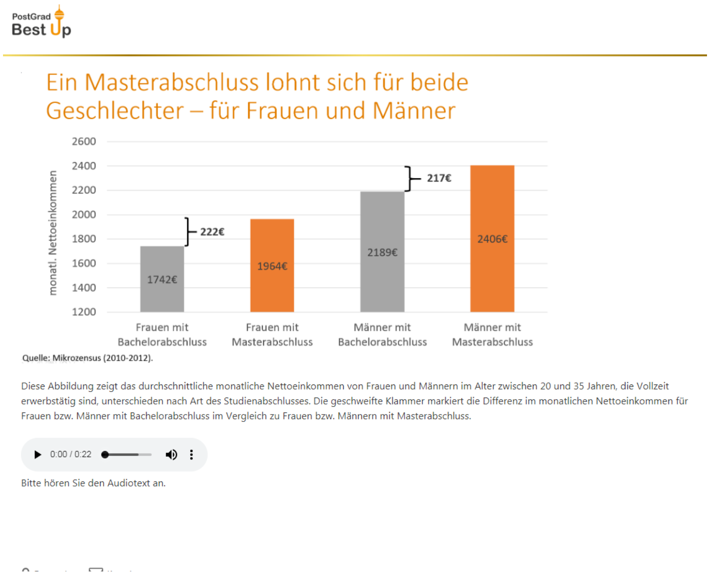
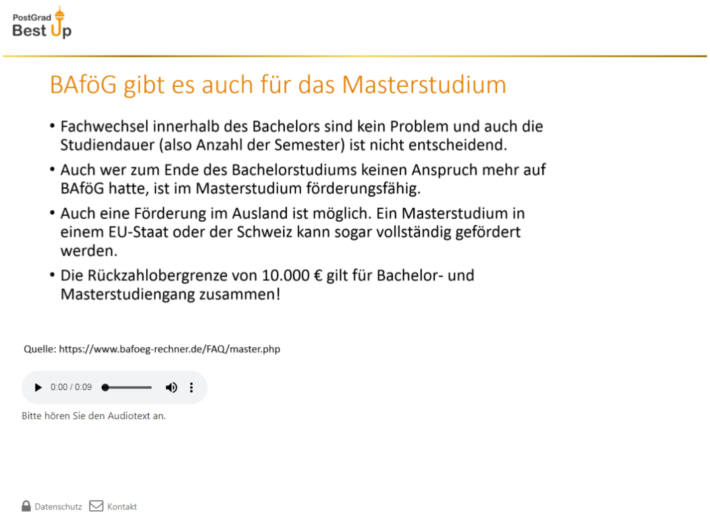


Figure 2.16: Slide 9 of the information shown to the treatment group



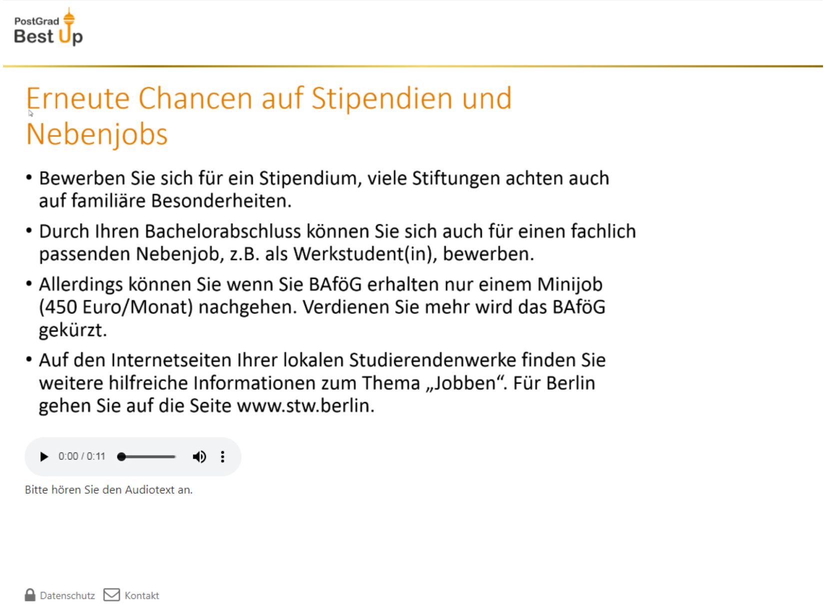
Text below figure: This figure shows the average net monthly income of full-time working women and men between the age of 20 and 35 by degree type. The curly brackets indicate the difference in net monthly income between bachelor and master graduates by gender.
Audio Slide 9: Even if the monthly income difference appears smaller, women with a master's degree also earn more than those with a bachelor's degree. Men earn more than women with both bachelor's and master's degrees. However, the difference in income between the two degrees is similar for men as it is for women.

Figure 2.17: Slide 10 of the information shown to the treatment group



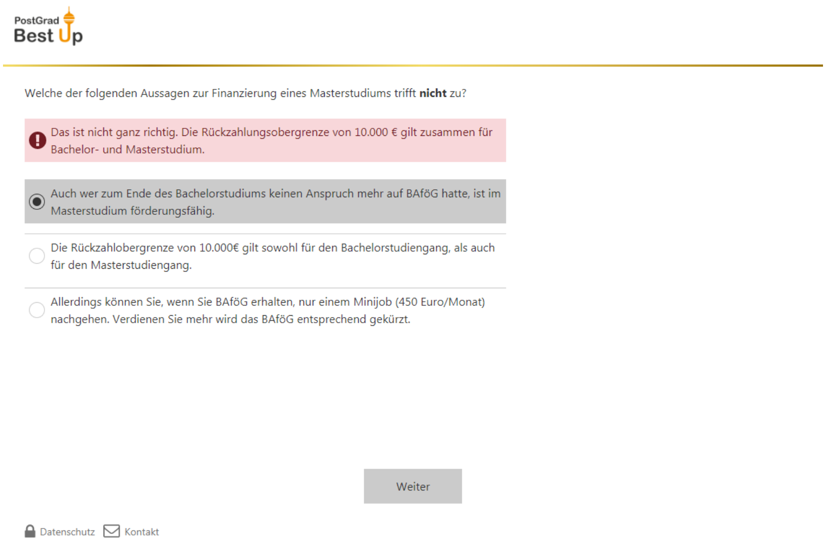
For translation of all slides see Section 2.B.5 below.
Audio Slide 10: So you can see that an application can be worthwhile!

Figure 2.18: Slide 11 of the information shown to the treatment group



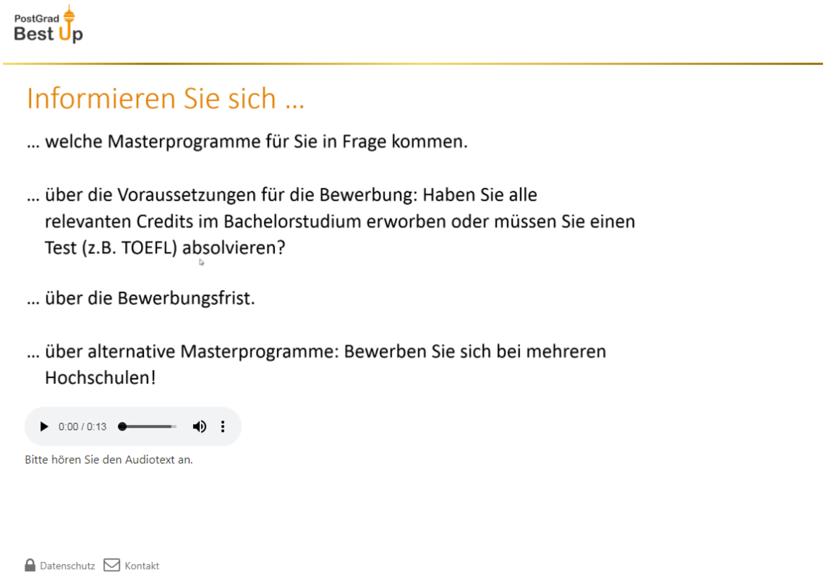
For translation of all slides see Section 2.B.5 below.
Audio Slide 11: Remember: A part-time job can be linked to your studies in terms of content, for example, if you work at the university as a student assistant.

Figure 2.19: 5^{th} comprehension question about slide 11 of the information shown to the treatment group



Note: This screenshot shows the screen if students chose the incorrect answer. Shown correction for answer option 1 or 3: “Once again, carefully read through all the points.”

Figure 2.20: Slide 12 of the information shown to the treatment group



For translation of all slides see Section 2.B.5 below.
Audio Slide 12: Contact your university's academic advising office; they can also help you with information about master's degree programs and the requirements for admission to the program.

Figure 2.21: Slide 13 of the information shown to the treatment group

For translation of all slides see Section 2.B.5 below.
Audio Slide 13: Contact your university's academic advising office; they can also help you with information about master's degree programs and the requirements for admission to the program.

2.B.4 Answers to comprehension questions

Below the comprehension questions and answer options are listed. Percentages in parentheses give the share of respondents who selected a question after the first and second attempt.

1. How long is the average time to degree in a master's program?
   - ✗ 3.8 semesters (9.68 %, 2.69 %)
   - ✓ 4.6 semesters (73.66 %, 94.09 %)
   - ✗ 5.2 semesters (2.69 %, 1.08 %)
   - ✗ 6.7 semesters (13.98 %, 2.15 %)
2. How much more net income does a master graduate earn at the age of 30 compared to a bachelor graduate?
   - ✗ 0 euro/month (0 %, 0 %)
   - ✓ 269 euro/month (97.31 %, 97.85 %)
   - ✗ 350 euro/month (1.08 %, 1.08 %)
   - ✗ 500 euro/month (1.61 %, 1.08 %)
3. In which occupational group earn master graduates almost 400 Euro more per month after taxes than bachelor graduates?
   - ✗ Jobs in natural sciences (4.84 %, 3.76 %)
   - ✗ Jobs in organization, public and office management (9.68 %, 1.61 %)
   - ✓ Jobs in corporate management, consulting and auditing (84.41 %, 94.09%)
   - ✗ Jobs in writing and artistic fields (1.08%, 0.54 %)
4. Who is more likely to work in a highly skilled occupation?
   - ✓ An employee with a master‘s degree (93.55 %, 97.31 %)
   - ✗ An employee with a vocational degree (2.15 %, 0 %)
   - ✗ An employee with a bachelor‘s degree (4.3 %, 2.69 %)
5. Which of the following statements about financing a master's degree program is not true?
   - ✗ Even those who were no longer entitled to BAföG at the end of their bachelor's degree are eligible for funding in their master's degree program. (31.18 %, 14.52 %)

- ✓ The repayment cap of 10,000 Euro applies to both the bachelor's degree program and the master's degree program. (53.76 %, 76.34 %)
- ✗ However, if you receive BAföG, you can only have a mini-job (450 Euro/month). If you earn more, the BAföG will be reduced. (15.05 %, 9.14 %)

2.B.5 English version of information treatment slides

Figure 2.22: Slide 1 of the information shown to the treatment group

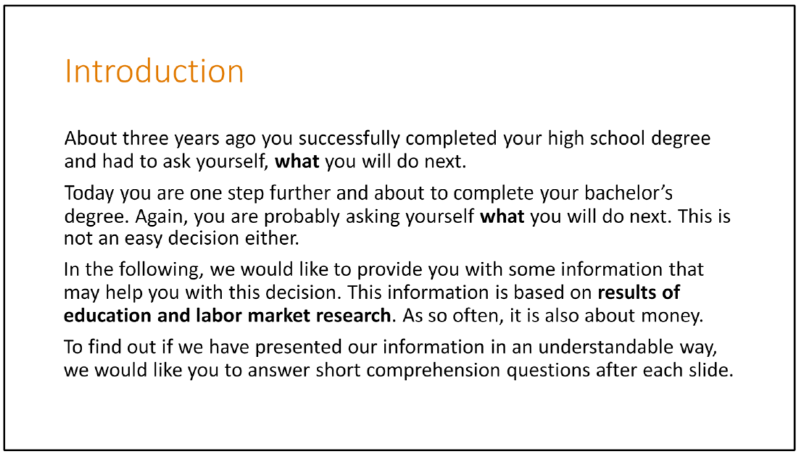
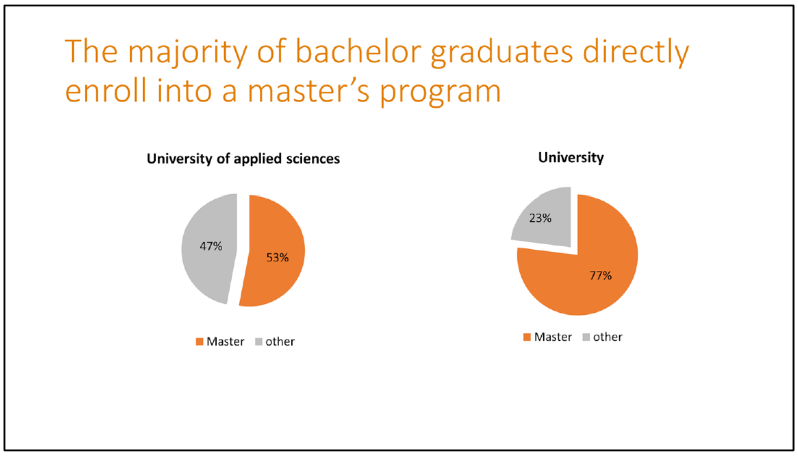


Figure 2.23: Slide 2 of the information shown to the treatment group

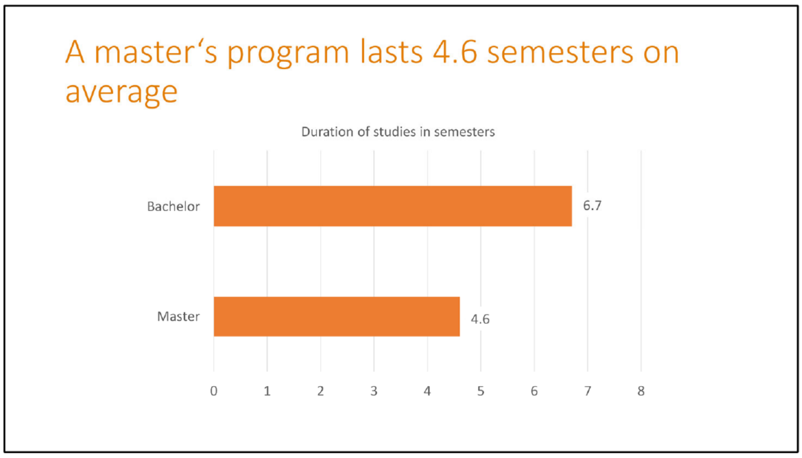


Source: Rehn et al. (2011), HIS 2011.

Text below figure: These two figures show the share of bachelor graduates that directly enroll into a master's program and the share of those who do something else after graduation.

Audiotext: Like the bachelor's program, the master's program is usually tuition-free.

Figure 2.24: Slide 3 of the information shown to the treatment group



Source: Statistisches Bundesamt (2016)

Text below figure: This figure depicts the average time to degree for bachelor's and master's programs in Germany. A bachelor's program lasts on average 6.7 semesters. A master's degree takes only 4.6 semesters on average.

Audiotext: On average, a master's degree is shorter than a bachelor's degree and, as scientific studies show, the majority of students study within the fixed number of semesters needed to complete the degree.

Figure 2.25: 1st comprehension question about slide 4 of the information shown to the treatment group

1st Comprehension Question

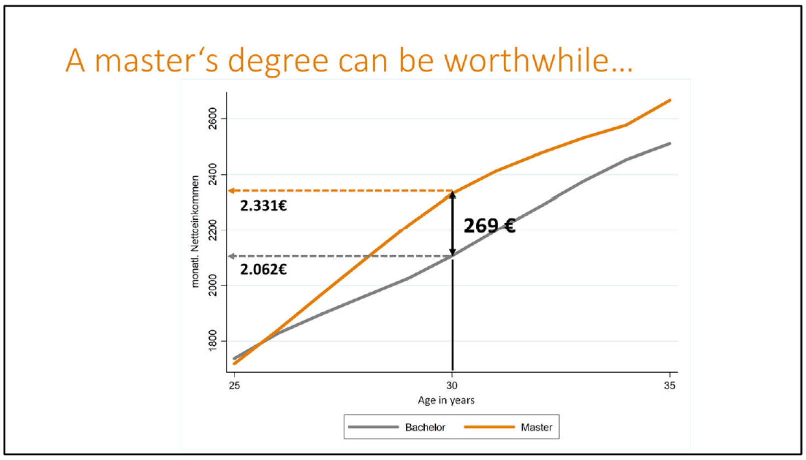
How long is the average time to degree in a master's program?

- 3.8 semesters
- 4.6 semesters
- 5.2 semesters
- 6.7 semesters

Show correction for:

3.8 semesters: almost
5.2 semesters: a bit too long
6.7 semesters: too long

Figure 2.26: Slide 5 of the information shown to the treatment group



Source: Microcensus (2010-2012).

Text below the figure: This figure shows the net monthly income (that is the amount you get transferred to your account after the deduction of taxes and social contributions) by degree type. The data applies to full-time employees.

Audiotext: For instance, the average net monthly income with a master’s degree at the age of 30 is 269€ higher compared to a bachelor’s degree.

Figure 2.27: 2^{nd} comprehension question about Slide 5 of the information shown to the treatment group

2nd Comprehension Question

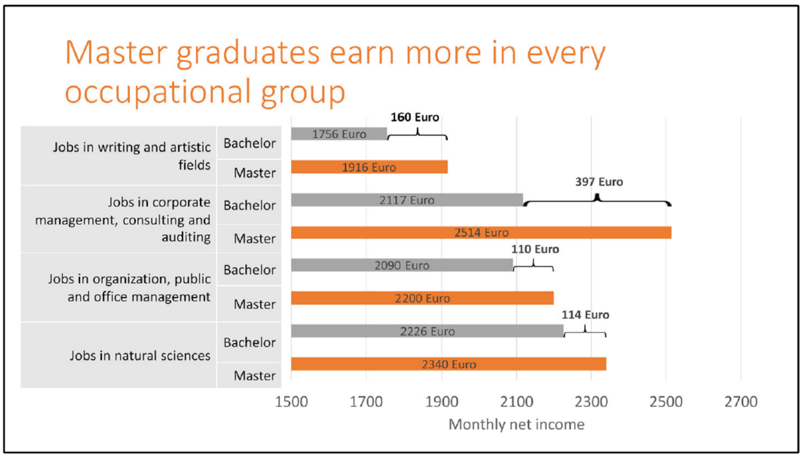
How much more net income does a master graduate earn at the age of 30 compared to a bachelor graduate?

- 0 euro/month
- 269 euro/month
- 350 euro/month
- 500 euro/month

Show correction for:

0 euro: too low
350 euro: almost
500 euro: too high

Figure 2.28: Slide 7 of the information shown to the treatment group

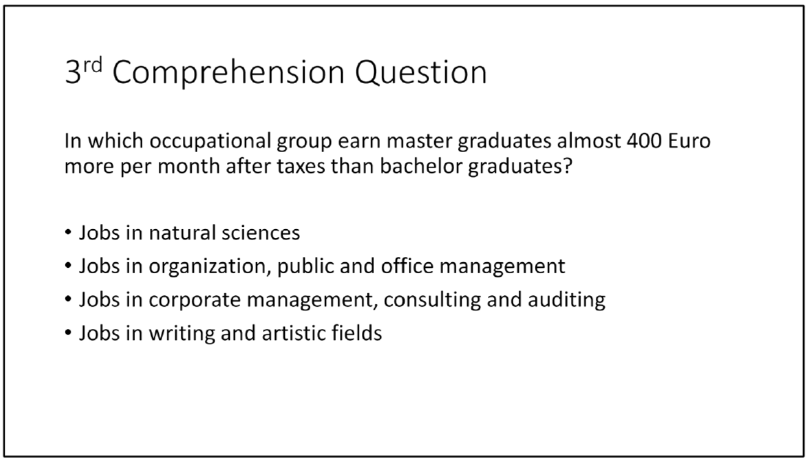


Source: Microcensus (2010-2012).

Text below figure: This figure shows the average net monthly earnings of full-time employees between 20 and 30 years of age by occupational group and degree type. The curly brackets indicate the difference in net monthly earnings between bachelor and master graduates by occupational group.

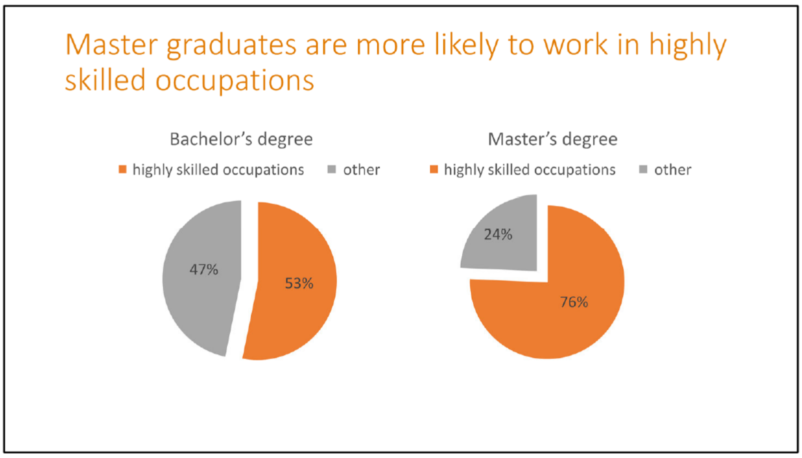
Audiotext: In each occupational group master graduates earn more than bachelor graduates. For instance, a full-time employee with a master's degree in a consulting firm does on average earn almost 400 euro per month more than a full-time employee with a bachelor's degree.

Figure 2.29: 3^{rd} comprehension question about slide 7 of the information shown to the treatment group



Correction for all wrong answers: Look again at the figure.

Figure 2.30: Slide 8 of the information shown to the treatment group



Source: Socio-economic panel v32 (2010-2015).

Text below figure: This figure depicts the share of bachelor and master graduates that work in highly skilled occupations.

Audiotext: Master graduates are more likely to work autonomously and much more likely to assign tasks to other colleagues.

Figure 2.31: 4^{th} comprehension question about slide 8 of the information shown to the treatment group

4th Comprehension Question

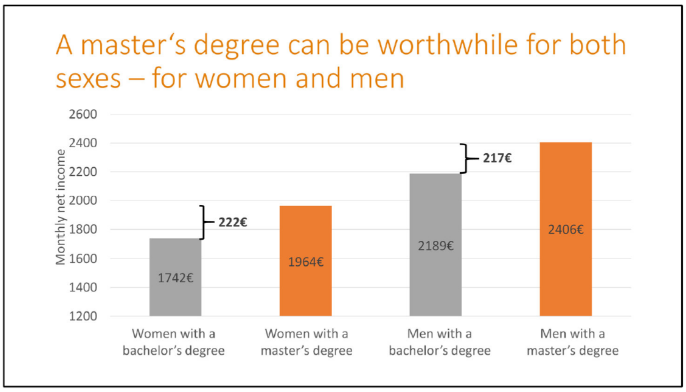
Who is more likely to work in a highly skilled occupation?

- An employee with a master's degree
- An employee with a vocational degree
- An employee with a bachelor's degree

Show correction for:

Vocational degree: Take another look at the figure.
Bachelor's degree: Take another look at the figure.

Figure 2.32: Slide 9 of the information shown to the treatment group

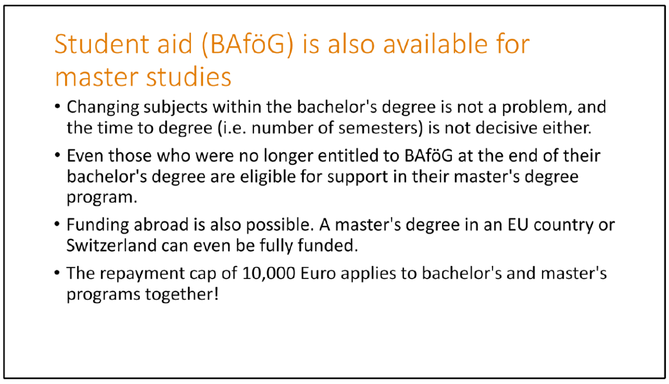


Source: Microcensus (2010-2012).

Text below figure: This figure shows the average net monthly income of full-time working women and men between the age of 20 and 35 by degree type. The curly brackets indicate the difference in net monthly income between bachelor and master graduates by gender.

Audiotext: Even if the monthly income difference appears smaller, women with a master's degree also earn more than those with a bachelor's degree. Men earn more than women with both bachelor's and master's degrees. However, the difference in income between the two degrees is similar for men as it is for women.

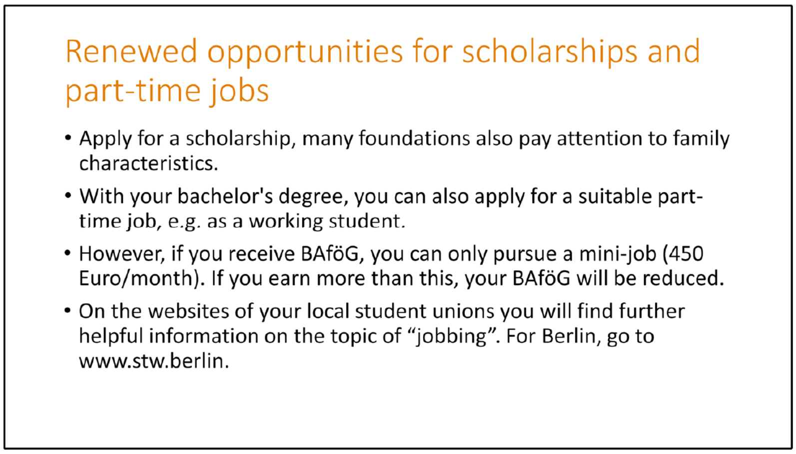
Figure 2.33: Slide 10 of the information shown to the treatment group



Source: https://www.bafoeg-rechner.de/FAQ/master.php, downloaded 7th November 2017.

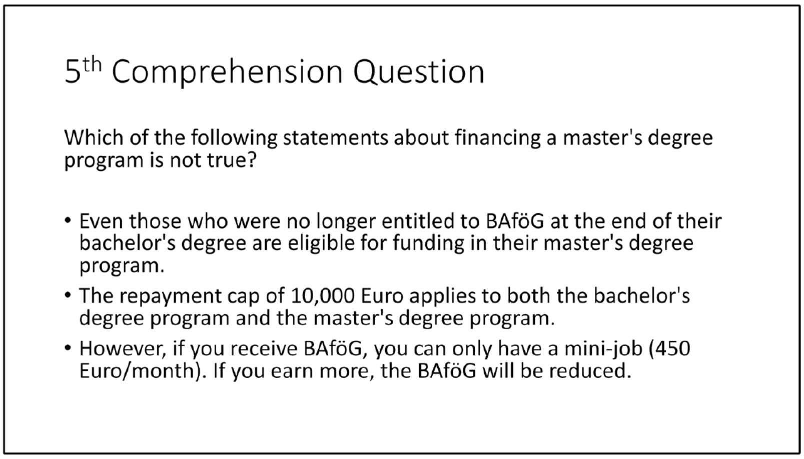
Audiotext: So you can see that an application can be worthwhile!

Figure 2.34: Slide 11 of the information shown to the treatment group



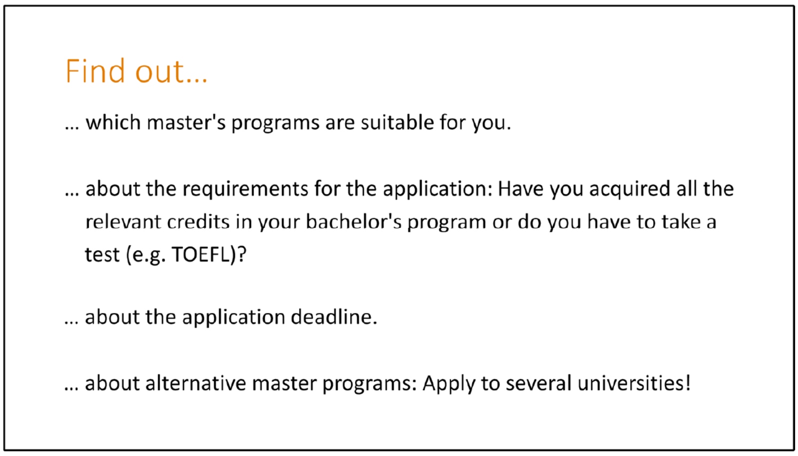
Audiotext: Remember: A part-time job can be linked to your studies in terms of content, for example, if you work at the university as a student assistant.

Figure 2.35: 5^{th} comprehension question about slide 11 of the information shown to the treatment group



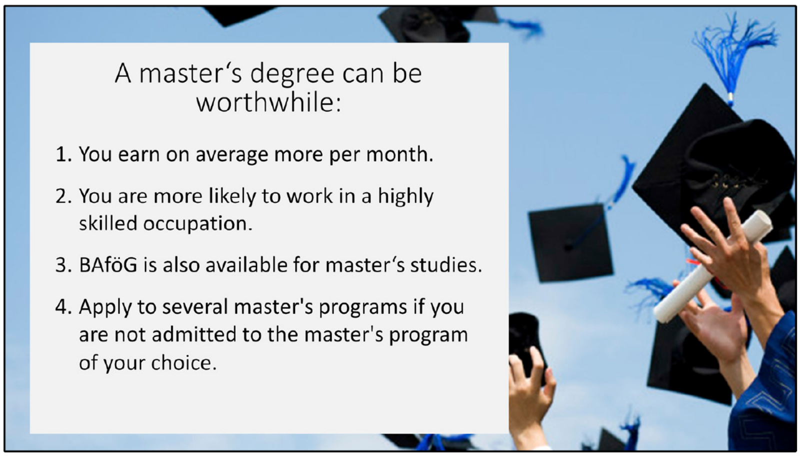
Show correction for answer option 1 or 3: Once again, carefully read through all the points.

Figure 2.36: Slide 12 of the information shown to the treatment group



Audiotext: Contact your university's academic advising office; they can also help you with information about master's degree programs and the requirements for admission to the program.

Figure 2.37: Slide 13 of the information shown to the treatment group



Audiotext: Remember: Even with a bachelor's degree you have already achieved a lot! Your prospects are very good in any case! We wish you all the best for your future path!

2.B.6 Survey question on students' beliefs about pecuniary and non-pecuniary returns (first follow-up)

Exact phrasing of the question in German: "Bitte versetzen Sie sich in die **Zeit, wenn Sie 30-35 Jahre alt** sein werden. Nehmen Sie an, dass Sie zu dieser Zeit einer **beruflichen Tätigkeit in Vollzeit** nachgehen. Einige Aspekte Ihres Lebens könnten davon abhängen, ob Sie **ausschließlich ein Bachelorstudium** oder noch **zusätzlich ein Masterstudium** abgeschlossen haben. Füer wie wahrscheinlich halten Sie es, dass ..."

Table 2.16: Question on expectations in working life from the first follow-up

	1 Sehr unwahrscheinlich	2	3	4	5	6	7 Sehr wahrscheinlich
Mit Bachelor: ... Sie einer intellektuell fordernden Tätigkeit nachgehen?	○	○	○	○	○	○	○
Mit Master: ... Sie einer intellektuell fordernden Tätigkeit nachgehen?	○	○	○	○	○	○	○
Mit Bachelor: ... Sie Familie und Beruf gut in Einklang bekommen?	○	○	○	○	○	○	○
Mit Master: ... Sie Familie und Beruf gut in Einklang bekommen?	○	○	○	○	○	○	○
Mit Bachelor: ... Sie ein überdurchschnittliches Einkommen verdienen?	○	○	○	○	○	○	○
Mit Master: ... Sie ein überdurchschnittliches Einkommen verdienen?	○	○	○	○	○	○	○
Mit Bachelor: ... Sie einer hochqualifizierten Tätigkeit oder Tätigkeit mit Führungsverantwortung nachgehen?	○	○	○	○	○	○	○
Mit Master: ... Sie einer hochqualifizierten Tätigkeit oder Tätigkeit mit Führungsverantwortung nachgehen?	○	○	○	○	○	○	○
Mit Bachelor: ... Ihre Eltern mit Ihrem Beruf zufrieden sind?	○	○	○	○	○	○	○
Mit Master: ... Ihre Eltern mit Ihrem Beruf zufrieden sind?	○	○	○	○	○	○	○

Note: The question is taken from the first follow-up survey and included the following notice: *Bitte stufen Sie Ihre Antwort zwischen 1 "Sehr unwahrscheinlich" und 7 "Sehr wahrscheinlich" ein.*

For students, who had already started a master's degree in the first follow-up (May 2018) the phrasing of the question was slightly different adding a sentence in front to the question: "Bitte gehen Sie für diese Frage davon aus, dass Sie sich momentan in einem Bachelorstudium in Deutschland befinden." ("For this question, please assume that you are currently enrolled in a bachelor's degree program in Germany.")

English-translation of the question: “Please put yourself in the **time when you will be 30-35 years old**. Assume that at that time you are **working full-time**. Some aspects of your life may depend on whether you have completed a **bachelor’s degree exclusively** or a **master’s degree on top of that**. How do you rate the probability ...”

Table 2.17: Question on expectations in working life from the first follow-up

	1 Very unlikely	2	3	4	5	6	7 Very likely
With a bachelor’s degree: ... to do intellectually challenging work?	O	O	O	O	O	O	O
With a master’s degree: ... to do intellectually challenging work?	O	O	O	O	O	O	O
With a bachelor’s degree: ... to be able to combine work and family life?	O	O	O	O	O	O	O
With a master’s degree: ... to be able to combine work and family life?	O	O	O	O	O	O	O
With a bachelor’s degree: ... to earn an above-average income?	O	O	O	O	O	O	O
With a master’s degree: ... to earn an above-average income?	O	O	O	O	O	O	O
With a bachelor’s degree: ... to work in a highly-skilled job or with managerial responsibility?	O	O	O	O	O	O	O
With a master’s degree: ... to work in a highly-skilled job or with managerial responsibility?	O	O	O	O	O	O	O
With a bachelor’s degree: ... that parents will be satisfied with your job?	O	O	O	O	O	O	O
With a master’s degree: ... that parents will be satisfied with your job?	O	O	O	O	O	O	O

Note: The question is taken from the first follow-up survey and included the following notice: *Please rate your answer between 1 “Very unlikely”and 7 “Very likely”.*

CHAPTER 3

Gender Inequality and the Relation between Market and Household Work[1]

3.1 Introduction

The topic of gender (in)equality in market and non-market work is continuing to attract attention in- and outside of academia. In the historical perspective, the rising female labour force participation has been one of the central and most-studied phenomena in labour economics (Altonji and Blank, 1999; Bertrand, 2018; Goldin, 2014; Olivetti and Petrongolo, 2016). Own income freed women from financial dependence on their spouses. While substantial progress was made in the last decades, important differences in labour supply remain. At the OECD-level, about 80% of part-time employment is done by females, and gender-gaps in income remain even conditional on labour supply: women working full-time earn about 13% lower wages than men.[2] At least since Gary Becker's *Treatise on the Family* (Becker, 1981), such differences in market work are seen as inherently linked to non-market work. In the standard non-cooperative household model, two individuals maximise utility taking into account their own and their partner's labour market returns (Lundberg and Pollak, 1996; Vermeulen, 2002). A typical implication of these models is that labour market inequality directly affects inequality also in non-market work. In short, if women cannot earn (the same amount of) independent income, they will also have to do more of the unpleasant household tasks, and have less power to make household-level decisions generally.[3]

[1]This chapter is joint work with Jonas Jessen (DIW Berlin, Freie Universität Berlin) and Felix Weinhardt (DIW Berlin, European University Viadrina, CESifo, IZA, CEP/LSE). We are grateful to Ludovica Gambaro and C. Katharina Spiess, as well as seminar participants at DIW Berlin and ifo Munich for helpful comments.

[2]Data from `https://data.oecd.org/earnwage/gender-wage-gap.htm`, accessed March 12, 2021.

[3]Individuals may experience different disutility from paid work and housework (e.g. due to the stress or reward associated with the work). Combined with positive, increasing returns to specialisation, corner solutions in household may exist where one partner is responsible for paid work and the other

Using multinational time-use data, Figure 3.4 shows the relation between the female share in market and housework across 73,214 households, for 16 countries, covering a time-span from 1974 to 2014. Female time-shares of market and household-work are pronounced inversely related. For example, we see the more egalitarian Scandinavian countries towards the lower right end, where women spend a higher share of their time in the labour market and equally a lower share doing household tasks.[4] Similarly, this relationship also holds for countries over time where the distribution in both domains becomes more equal over time (e.g., West Germany). It is this inverse relation between shares in market- and non-market work that continues to motivate research and policy regarding gender equality to this day. This descriptive evidence is in support of the (theoretical) view that gender equality in non-market work can hardly be achieved without a levelling of the playing field for market work.

This paper takes this hypothesis to a test: does the amount of individual housework contribution depend on their spouses labour supply? And if not, what are the underlying mechanisms that give rise to the observed inverse relation between share of market and share of non-market work?

Providing a causal interpretation of the relation shown by Figure 3.4 is not trivial. These descriptive statistics reflect a complicated process of joint decision making, with additional constraints or preference parameters that might be imposed by society or specific institutional settings. These might interact with third (endogenous) factors, such as the presence of children. Moreover, the partner choice itself is potentially endogenous to the later-observed household-level realisation of time-allocation (after accounting for selection into employment).

To circumvent these challenges, we study household-level time allocation before and after German reunification, making use of five time use surveys from 1985 to 2013. These include newly sourced time-use data from the GDR, which to the best of our knowledge have to date not been analysed by economists before. As we will set out below, the combination of data availability and the institutional setting of the GDR and West Germany is of particular interest. In short, this is because institutions between the two German states differed so that women in the GDR spent much more time in the labour market, compared to their Western counterparts. We argue that at least part of these institutional differences shifted female labour supply out in the GDR in ways not related to the intra-household decision process regarding time spent doing household work. Or to put it bluntly: is it true that households in the GDR were also more equal

one for housework, a "separate spheres equilibrium" (Lundberg and Pollak, 1993). However, this does not explain why the common distribution, both unconditional and conditional on working hours, is that women tend to do a large share of housework while the opposite is observed much less often.

[4]Households tasks exclude childcare and are defined in section 3.3.

in terms of their within-household allocation of housework tasks? Notably, the two German states represent data points on the opposite end of Figure 3.4. We therefore believe our results also allow drawing general implications beyond the German context.

Based on descriptive and regression analysis, this paper proceeds in four steps. We first provide descriptive evidence on the time allocated to market and non-market work in the GDR, as well as West and East Germany in 1990/91, i.e. shortly after reunification. This analysis confirms that it is true that women in the GDR worked more hours in the labour market compared to women in the West, and also spent fewer hours on household tasks. Moreover, the within-household gaps in each activity are smaller in the GDR: the within-household female share in market work is larger, and the female share in household work is smaller in the GDR and East Germany, compared to the West. This might suggest more equality across the board.[5]

In the second step, we decompose the relation between female shares in market and housework by household type. To do this, we classify households into three types, i.e. male breadwinner, dual earner, and female breadwinner, based on the time spent on market work. While the male breadwinner is the norm in the West, the dual earner is the norm in the GDR/East Germany. Interestingly, accounting for these underlying types, the female shares in housework look remarkably similar across countries/regions, and in fact show a negative relation comparable to the cross-country evidence presented in Figure 3.4. One possible explanation for these patterns is that there are no male reactions to different levels in female market work regarding their household work, and vice versa, implying that the greater equality documented in the first step of the analysis is not because of behavioural differences but mechanical in nature.

In the third step, we use regression analyses to better understand the factors that give rise to the differences in female housework shares across West and East Germany. The key finding from this analysis is that once individual labour supply (for market work) is accounted for, the gap between East and West Germany becomes statistically insignificant at conventional levels. This complements recent findings that males do not adjust their own labour supply when their wives start working (Knowles, 2013). We find that time allocation in household-tasks reacts neither.

We present a set of robustness checks to support our findings. We test sensitivity against different definitions of tasks that enter housework or the role of children, with the different availability of childcare infrastructure in the East and West. The finding of the smaller female housework shares in the GDR/East holds independent of the presence of children – in fact household without children provide the cleaner analysis since potential complimentary between child-rearing and housework can be excluded.

[5]We return to this question when discussing the incidence of the "second shift" in Section 3.8.

Our headline finding is therefore that once (institutional) differences in the labour market are accounted for, households look remarkably similar across the two German states.

This result has consequences, which we examine in the fourth and final step. Here, we examine the relation between market work and leisure. We document that spending more time in the labour market for women comes at the expense of less leisure. Accordingly, we document that working women in Germany, in the GDR as well as in both parts of reunified Germany, face the "second shift", a term initially introduced by Hochschild and Machung (1989). The two German states were always united in women having a second shift of housework after coming home from (market) work.

Taken together, these findings have important implications. First of all, institutional changes that increase women's employment (or reduce wage differences) are unlikely to affect the relation of time-allocation between market work and housework. Such policy should be seen as, at best, reducing gaps in the labour market but not be expected to shift women's time spent on housework to a fully gender-equal level or cause spouses to relieve their wives from their domestic burden. At least the extreme labour market policies and propaganda imposed by the Soviet rule and GDR over East Germans, for a duration of four decades, did not make partners' time allocation to housework more responsive to their spouses labour supply, and thus lead to gender inequality in the domestic domain. Second, our results put into perspective the narrative about higher gender equality in the GDR. While it is beyond any doubt that women in the GDR had a higher labour market attachment, compared to the West, we do not find evidence for behavioural differences in the allocation of housework. Third, our findings shed light onto the nature of gender norms. We use data from the World Value Survey to support the hypothesis of (partly) separate norms for market work and housework; East-West differences for norms on sharing household chores are much smaller than norms on wives contributing to household income / being restricted to the role of a housewife.

Our study is related to several strands of literature. First of all, we contribute to the literature that argues that female employment is the "strongest and most consistent" predictor of absolute and relative levels of women's housework and that men's employment and working hours are much weaker predictors (Coltrane, 2000). It is documented that at the household level, across countries, in recent decades women have decreased their housework contributions and men have increased theirs to a much smaller degree (Coltrane, 2000; Sayer, 2005). We contribute to this literature by studying the relation between time spent for market and housework in a compelling setting: divided and reunified Germany, with large differences in labour market policy and gender norms regarding female employment.

Second, we contribute to the literature examining how East and West Germans differ in gender-related attitudes due to exposure to different policies during the division. Using SOEP survey data, Lippmann et al. (2020) show that West German wives—but not East German ones—are more likely to increase their housework contribution and to withdraw from the labour market after outearning their husbands in order to conform with traditional gender roles. Only in West German couples is the risk of divorce increased when the wife is earning more.[6] Using the same data, Campa and Serafinelli (2019) provide evidence that East German women place more importance on career success (see also Beblo and Görges, 2018), and that both men and women hold more gender egalitarian views in East Germany. Zoch (2021) analyses East-West differences in attitudes towards maternal employment and housework and finds pronounced disparities in attitudes, but those have become smaller for younger cohorts.[7] Our findings imply that such differences—including those on housework norms—are likely to be primarily driven by higher (state-imposed) female labour force participation in the GDR and that norms on housework were not affected independently of this.

Moreover, our empirical results confirm that gender norms regarding market work need not be closely related to gender norms regarding housework (Grunow et al., 2018). We demonstrate that if preferences and social norms about housework evolve isolated from other domains of gender equality such as market work (Hakim, 2000), policies can have unintended, detrimental consequences for women. One explanation for the weak link between equality in the labour market and in domestic spheres is that female labour force participation is publicly visible which allows for local learning to take place gradually over time (Fogli and Veldkamp, 2011). In contrast, housework is rarely observed beyond families which might slow down learning processes.[8]

In the next section, we describe the institutional setting. Section 3.3 introduces the data. Section 3.4 provides the empirical analysis and results, in the three steps outlined above. Section 3.5 discusses robustness and the role of children. Section 3.6 documents the second shift for working women in both Germanies and Section 3.7 provides evidence on separate norms for market and non-market work, before we conclude in Section 3.8.

[6]Using annual rather than monthly earnings and employment data, Sprengholz et al. (2020) are, in contrast, unable to find evidence that East or West German women are more likely to leave the labour market after being the main earner.

[7]See also Bauernschuster and Rainer (2012) who document large differences in attitudes towards maternal employment and the role of wives in the family immediately following reunification.

[8]While norms about housework can change over time they seem to do so only slowly. Hwang (2016) shows that US immigrants who grew up in societies with a higher share of housewives also show larger gender gaps in the division of housework after migrating, yet gaps decrease over generations. Vargas (2016) also finds persistently higher housework contributions by Mexican immigrants in the US than for non-Hispanic whites.

3.2 Institutional setting

In 1871, after the Franco-Prussian war, a German national state was established for the first time. After the end of the atrocities of World War II, Germany was divided into four occupation zones by the victors in 1945. In 1949, the GDR was formally established in the Soviet occupation zone and the FRG consisting of the three western zones. The GDR was a socialist, one-party state under strong influence of the Soviet Union. In contrast, in the FRG a market-based democracy was established. Between 1949 and 1961 more than 3 million left the GDR to West Germany (about 20% of the population), and about 500,000 moved in the opposite direction. In August 1961, the Berlin Wall was constructed and effectively ended the migration movements (Heidemeyer, 1994).

During 41 years of formal separation the two German states diverged in many regards, including female labour force participation (Trappe, 1996). The GDR enforced a high participation rates through several policies. E.g. child care was strongly expanded[9] and not working was considered to be anti-social behaviour (Beblo and Görges, 2018). As a result, female labour force participation increased strongly and in 1989 with a rate of 89% it was among the highest in the world. Additionally, most women worked full-time and differences by marital status and children was small (Rosenfeld et al., 2004).

Gender policies in the FRG were conservative in comparison. Limited child care availability and afternoon care as well as joint taxation for married couples favoured (main) male breadwinner households (Boelmann et al., 2020). Female labour force participation was a third lower than men's and part-time work was prevalent, especially so for young mothers.[10] A popular children's song in the GDR was "Wenn Mutti früh zur Arbeit geht" (*"When mommy goes to work in the morning"*), whereas in the FRG wives by law only had "[...] the right to be employed as far as this is compatible with her marriage and family duties" until 1977 (Lippmann et al., 2020). Despite these differences, gender earning gaps were about similar with 25% for full-time workers (Krueger and Pischke, 1995), but as outlined above, selection into employment was much lower for women in the GDR. Similarly, wage gaps differed relatively little with 15% in the GDR and 18% in the FRG (Sørensen and Trappe, 1995).

The two German states were reunified in October 1990 following the fall of the Berlin Wall one year before. East Germany fully adapted the policies of the FRG,

[9] In 1989, 98% of children aged 3-6 attended child care facilities and more than 80% of children below 3 (Schmitz and Weinhardt, 2019). In contrast, in the FRG child care for under threes was basically non-existent and for older children almost all spots were part-time only.

[10] In the GDR it was mostly older women working reduced hours.

with arguably the most notable difference remaining being the much higher provision of child care spots in East Germany.

3.3 Data

GDR. For our analysis we obtained access to the 1985 and 1990 waves of the GDR time budget study (Zeitbudgeterhebung) at the German Federal Archives (BArch). To the best of our knowledge, the data has not been used by economists before. The study was conducted by the statistical office of the GDR to obtain data for the planning of demand for goods and services, to demonstrate the effectiveness of economic and social policies on the use of time outside of work and to design new reforms that foster efficient time-use (Fiebiger, 1991).

The 1985 wave documents time-use in the GDR years before the fall of the wall in 1989. The 1990 wave was collected before Germany was officially reunified into a monetary, economic and social union. Data collection of the GDR time budget study only took place among worker and employee households and retiree households.[11] Each household was supposed to fill out the survey on a pre-determined day of the week. Main tasks were documented for 24 hours, starting at midnight.

Post Reunification. We use three waves of German time use data from after the reunification. The German Time-Use Survey is a repeated cross-section of around 5,000 households taken in 1991/92 (so briefly after reunification), 2001/02 and 2012/13 (Maier, 2014). Each adult household member records his or her activities in ten minutes slots over three survey days (five minutes over two days in 1991/92). The activities are categorised in detailed three-digit activities. Besides the diary data for the survey days, the data also contain other household and individual information, such as age, marital status, schooling degree, occupation, children in the household, and household income (in categories). Throughout the analysis we treat East and West Germany separately.

Multinational time use survey. We complement the German time surveys with international time use surveys. The Multinational Time Use Study (MTUS) from the Centre for Time Use Research offers data from more than 70 national time use surveys and standardises them (Gershuny and Fisher, 2013). Table 3.4 shows the country and surveys used in the analysis. A fundamental requirement for the within-household analysis is a household-level sampling design and the existence of household identifiers in the data. We further set the restriction that we look at different-sex (married)

[11] Priller (1993) confirms that the data is representative for worker and employee households by districts but that one-person households and young male respondents are slightly underrepresented. Since we primarily study couple households this is of less concern for our study. Households with self-employees and cooperative members were not included but also represented only a very small share of the GDR population.

couples that we observe on the same day, so that we can directly observe the shares in each activity without any further assumptions. Finally, to reduce the impact of outliers we, require to have at least 400 couple-day-level observations. These conditions lead to a sample of 23 time use surveys from 13 countries.

Harmonisation of different data sets and sample selection. In a final step we harmonise the various time use studies (GDR, reunified Germany, MTUS). To make the studies comparable across years and countries, we define consistent categories of activities. The broader categories are i) total work time ii) housework iii) care for others (mostly children and elderly) iv) leisure v) personal hygiene vi) eating vii) sleeping, but we are also able to distinguish on a finer level.[12] As most of our analysis is on a household within-couple level, we impose some sample restrictions; we look at (married) couples and due to the focus on gender differences we restrict this to different-sex couples. As we look at the interplay of time invested in paid work and housework, we further restrict the couples to be of working age, i.e. 18-65. In many analyses we focus on weekdays which mostly are regular working days with a positive number of working hours, but we don't generally restrict the sample to this as adjustment effects could occur on weekend days.

3.4 Results

To examine household time allocation and the nature of the inverse relation between household-shares in time spent in market and household work, we now focus on the German case. The German case is of particular interest because of the different policies regarding female labour force participation described in section 3.2. We proceed by first examining descriptive patters for the GDR, as well as East and West Germany separately.

3.4.1 Gender differences in paid and housework work during reunification

We show a summary of the harmonised GDR time budget study and German time-use data in Table 3.1. In the GDR, the gender gap in minutes of paid work was 127 minutes per day. Women worked 102 minutes more in the household and spend 32 minutes more in care activities than men. Further, women had about 36 minutes less leisure per day. By comparing East and West German data at the earliest possible time after reunification we capture the differences between the GDR and FRG as good as possible in a single data set. For comparison and empirical analysis which do not rely on an East-West comparison, we use the GDR data. Despite a drop in female

[12]A detailed list of activities contributing to the broader categories are presented in Table 3.5.

employment between 1985/90 and 1991/92 from 92 to 68 percent, total work time of East German women was still more than twice as high as that of West German women in 1991/92. We present summary statistics by children and for the later survey waves of the German time-use survey in Tables 3.6 and 3.7.

Figure 3.1 provides further details about the distribution of minutes spent on paid work and housework. The upper panel shows that in the GDR only about 20% of women and less than 10% of men did not work on the workday recorded. While almost all women do at least some housework, about 10% of men did not spend any time on housework. The lower panel of Figure 3.1 shows that despite the lower average in paid work, the distributional gender differences in the GDR in 1985/90 and East Germany 1991/92 were relatively similar. This can be seen by comparing the solid lines across the upper and the lower panels: the overall pattern is similar, in particular when compared to the West Germany case, represented by the dashed curves in the lower panel. In West Germany, we see a much more gender traditional distribution of paid and housework work. Both the gender paid work and housework gaps are substantially larger in West Germany than East Germany, again indicating that the GDR was the more gender equitable country.

3.4.2 Female share of housework and heterogeneity in household types

Comparisons between households of different countries or regions, such as in Section 3.4.1, are not *ceteris paribus*. Conclusions that one region might be more gender equitable could therefore be problematic. To better understand the documented differences in gender gaps we now analyse time spent in paid and household work considering heterogeneity within country/region.

3.4.2.1 Household types in the GDR

Figure 3.4 showed a strong inverse relationship between the female housework and paid work share using multinational cross-country data. We begin the analysis of household heterogeneity by replicating this figure at the within-country household-level using the data from the GDR. For illustrative purposes, we define three types of households distinguished by the female share of market work; 1 $[0, 0.35)$, 2 $[0.35 - 0.65)$ and 3 $[0.65, 1]$. Type 1 is a (main) male breadwinner household, whereas the second type is a dual-earner type / equal work household type. As laid out in section 3.2, the typical household in the GDR was of the second type. Type 3 households, (main) female breadwinner, remain the exception. We restrict this analyses to households where at least one partner is working full-time ($\geq$ 7 hours) which are the most relevant cases in our context.[13]

Figure 3.5 illustrates that the cross-country negative relation between household and market work shares also replicates across households of the GDR. Second, the lower solid black bars (right-hand Y-axis) shows the prevalence of the dual-earner household. 74 percent of households are the dual-earner type, only 20 percent male breadwinner, and a residual of five percent female breadwinner.

We next split the shares in housework into minutes contributed by females and males, across our three types of households. The results of this is shown in Figure 3.6. The panels show the total minutes of housework of females/males depending on (time-)shares in paid employment, respectively. Consider the case of the male-breadwinner household. Here, females spend almost four times as much time doing housework compared to the working males: about 380 vs. 100 minutes. These figures underlay the share of almost 0.8 for the male-breadwinner household shown in Figure 3.5.

Now, consider what happens when we move to the dual-earner household: Notably, females reduce their time spent doing housework substantially, to about 200 minutes. This effectively halves their time spend in this activity. On the other hand, males in dual earner household still contribute about 100 minutes to housework. In other words, from the perspective of the male, it makes little difference if the female is working full-time, or few hours or not at all. But this is only one side of the coin: In the (rare) case of the female breadwinner household, males spend over 200 minutes on housework, whereas the full-time working females spend about 160 minutes. Again, comparing this to the dual-earner household, the adjustment only takes place by the partner who is changing their own labour supply: neither men nor women strongly adjust their housework contributions when paid work of their partner is larger or smaller.

Taken together, the gender-split of the inverse relation between household shares in time spent on the labour market and time spend doing housework seems to suggest that the overall negative relation is entirely driven by individual time-constraints: if one is working, there is less time for housework. Changes in the partner's work arrangement hardly affect own decisions, at least in this cross-sectional comparisons across three stylised types of households.

[13]This excludes cases where both partners work very little hours and a shift between different types of households can occur by small changes in one's working time. As such comparisons are of limited insight in this context, we do not consider those.

3.4.2.2 Household heterogeneity in East and West Germany

In this context, it is interesting to compare and contrast these patterns to West Germany, where female employment was substantially lower.[14] To do this, we show results based on the 1991/92 time-use survey for reunified Germany. As discussed above, at this point in time, only very few West Germans had migrated to the East. Figure 3.2 replicates Figure 3.5 for East and West Germany. At first sight, the two regions from Germany look remarkably similar with a comparable downward relationship between shares in market and household work. This may be counterintuitive at first as East Germany is commonly portrayed to be more gender egalitarian in many respects, and this was also shown in Figure 3.1.

However, the black bars indicating the distribution of household types reveal the underlying reason for this unexpected similarity in the graphs; the East German distribution resembles the one from the GDR strongly, whereas in West Germany the (main) male breadwinner type is most common. 72 percent of households belong to type 1 and only 25 percent are dual earner households. Once this differential selection into employment is taken into account within this simplistic framework, relative housework contributions differ little between East and West Germany. Figure 3.10 shows male and female contributions in minutes.

Taken together, the lack of reaction of individuals to their partner's changes in labour supply to market work, and the different incidence of household types, suggests that while large gaps in household work existed between the GDR (or East Germany following reunification) and the West, these need not necessarily reflect different behaviour at the level of the household, but are solely due to different selection into market work.

3.4.3 Decomposition of the gender gap in housework

To analyse East-West differences in a more structured way and including all household types, we now turn to a decomposition of the housework gap between East and West Germany. Again we use the German time-use survey from 1991/92, i.e. very shortly after reunification. This has the advantage that while norms are arguably still strongly influenced by the differing environments individuals were exposed to during the German division, the survey is conducted in a uniform fashion, alleviating concerns about different survey designs or sample selections (which may be an issue when comparing the GDR data to West Germany). The descriptive evidence in section 3.4.1 showed that the gender housework gap varies strongly between East and West Ger-

[14]Note that this implies that the group of women working full-time in Germany is more selective than in the GDR where full-time employment was essentially a governmental policy.

many. While women perform 209 minutes more of housework than men on weekdays in West Germany, the difference is 'only' 102 minutes in East Germany.

To elicit to what degree the East-West difference of 107 minutes is due to observed factors, we estimate regressions controlling for important individual characteristics and, most importantly, male and female market work. If after controlling for these factors a large residual East-West difference in the housework gap remains, then this is strongly suggestive of different norms regarding gender roles in the household, which play an important role independent of norms regarding those concerning the labour market.

We use the conditional decomposition developed by Gelbach (2016) to avoid the complication of sequence sensitivity when covariates are added. In a first step the housework gap is regressed on an East dummy. In a second step, the full model using all explanatory factors is estimated. Taking into account both the correlation between the dependent variables and the outcome variable as well as the correlation between the regions (East or West) and the dependent variables, the conditional decomposition from Gelbach provides consistent estimates on the role of each covariate in moving the East dummy from the baseline to the full model.

Results are presented in Table 3.2. Column 1 shows the raw difference, columns 2-5 contain the full model where we control for female and male market work in different ways. The East dummy is strongly reduced from 107 minutes to 11-16 minutes depending on the specification, i.e., a large share of the East-West gap can be explained by the covariates. Looking at the contributions of the different groups of explanatory variables, it is apparent that basic individual and household-level controls as well as household income have only a small impact. Depending on the specification, female work is responsible for 82-85 percent of the reduction of the East dummy. In contrast, male work, if anything, has only a minor effect. This supports the notion, in line with the visual evidence in Figures 3.2 and 3.10, that it is not relative contributions to paid work that determines housework, but to a large extent only women's own contribution. East-West differences in the housework gap are thus mostly due to different selection into employment of women. The descriptively more gender egalitarian division of household tasks within the GDR that we documented in Section 3.4.1 should not be mistaken for evidence for more gender egalitarian behaviour at the level of the household.[15]

This finding further suggests that unobservable differences between East and West Germany, such as differences in child care arrangements, domestic technology, gender

[15]In Table 3.8 the same table is shown, but with the dependent variable specified as the female housework share. Reassuringly, the results are very similar.

norms about housework or preferences might play a negligible role in explaining the housework gap. Some of these factors are analysed in more detail in the next section.

3.5 The role of children and robustness

3.5.1 Definition of household tasks

Table 3.9 contains the same estimates as Table 3.2, but restricts the housework to consist of the subcategories *cooking, cleaning, and shopping.* These "classic" domains of housework are the most time-consuming and are "less optional and less able to be postponed" (Coltrane, 2000). While the overall gap in this narrower housework definition is larger (219 minutes in West Germany), East West differences are smaller (83 minutes). Overall patterns, however, are very similar, and not driven by specific choices regarding activities that contribute to household work.

3.5.2 The role of children

The institutional differences in child care availability between the GDR and the FRG, particularly for children under the age of 3, which persisted after reunification, are a potential explanation for East-West differences in housework gender gaps. Having young children in child care instead of home care might reduce the amount of housework needed, which then causes West German women to work more in the household than East German women, even conditional on employment.

We study the role of children by repeating the analysis of the preceding section separately for couples with children under the age of 10 and couples with only older or no children. As results from different specifications of female and male work in Table 3.2 were in the same ballpark, we proceed using the specification with 5 categories of paid work as dummy variables. Results with the distinction by children are presented in Table 3.3. Columns 1-4 contain estimates for the female-male housework gap. First of all it is worth nothing that the gender housework gap differs little by the presence of children. In families with no young children (columns 1-2), controlling for the covariates leads the initial large East dummy to be small and statistically insignificant. The coefficient movement of the East dummy is almost entirely driven by female work. As institutional differences in child care provision (and afternoon care for primary school children) between East and West are irrelevant for this group, this is arguably the most suitable comparison for housework norms conditional on observed covariates.

Columns 3-4 contain the estimates for households with children. While the East dummy is strongly reduced by the covariates, the larger remaining gap of 36 minutes (just below 30%) highlights the relieving effect of the East German child care infrastructure for mothers.

In an additional step (columns 5-6), we change the dependent variable to the female-male *domestic* gap, which in addition to housework also includes time spent on child care activities. As in both regions women are predominantly responsible for child care the female-male gap increases to 307 minutes in West Germany. As in West German families, mother contribute relatively more child care and housework (Jessen, 2021), the gap is 163 minutes smaller than in West Germany. The covariates, again especially selection into employment for women, strongly reduce the East dummy, but it remains at a similar magnitude than for the housework gap (column 4). All in all, the distinction by children stresses that overall East-West differences in the gender-housework gap are small once employment is considered, but they remain stronger when children are part of the equation.

3.5.3 Extension into 2001/02 and 2012/13

In the first two decades after reunification, female labour force participation in East Germany has slightly dropped below the 1989 level (Rosenfeld et al., 2004). At the same time, participation in West Germany has increased steadily from the 1990s onward, reaching almost similar levels as in the East today.[16] While child care facilities for children under the age of three remain much more common in the East than in the West until today, their availability has increased in West Germany in the past three decades.

Above we used the 1991/92 wave for our primary analysis as it likely captures time allocations that are very similar to the situation in the GDR and FRG shortly before reunification. To investigate if the difference in housework gaps between East and West Germany remains stable over time, we repeat our estimation with later waves (2001/2002 & 2012/13) of the German time use survey. Results are shown in Table 3.10. The average gap reduces strongly over time—from 209 minutes in West Germany in 1991/92 to 126 minutes in 2012/13—and so do unconditional East-West differences. Once female's work is controlled for the differences are small and no significant differences between the waves can be observed. The result suggests that later spill-overs of differences between the GDR and FRG into the domain of housework are unlikely. Moreover, they support the conclusion that once individual work is accounted for, dif-

[16]Labour force participation rates have also converged for mothers, but differences persist at the intensive margin, as East German mothers are substantially more likely to be working full-time (Barth et al., 2020).

ferences between regions are insignificant even over the period of German reunification, where many institutions changed.

3.6 Paid work and leisure

The analyses above showed a strong negative relationship between paid work and housework. Importantly, if housework is not reduced to a similar degree, an increase in paid work must also come at the expense of an activity that many arguably value the most; leisure.[17] The literature has argued vividly (Hochschild and Machung, 1989, 2012) that an increase in working hours of women may lead them to do a *second shift*, i.e. paid work and housework, which comes at the expense of their leisure.

Figure 3.8 shows the relationship between paid work and leisure in 60 minute bins of paid work (scatters with less than 20 observation in the bin are dropped) using the 1991/92 German time-use survey. Several aspects are worth noting; first, there is a clear downward relationship between leisure and paid work. If one moves from not working to working full-time (450 minutes per day) leisure is reduced by 114 minutes on average. Second, women are consistently found to have less leisure conditional on time in paid work. This holds both for individuals not working and for those in full time work.[18] Women have around one hour less leisure conditional on working and this holds for women in East and West Germany. Finally, both men and women in West Germany tend to have more leisure than their East German counterparts. I.e., East German women have by far the least leisure of all groups.[19]

We decompose East-West differences of the female-male leisure gap in Appendix Table 3.11 for all survey waves. The gap in West Germany is in fact positive in 1991/92 (8.6 minutes),[20] as women work much fewer hours than men, and is very close to zero in later waves. In East Germany, the raw gap is initially almost half an hour lower, so women have on average about 20 minutes less leisure compare to men, suggesting that may indeed be working a second shift. Similar to the decomposition of the housework gap by survey wave East-West differences become smaller over time. Female working time is also for the leisure gap the factor impacting the East Germany dummy by far the strongest. After female working time is factored in, the East dummy becomes positive and in the pooled estimation is significantly different from zero.

[17]In the 1991/92 survey, paid work, housework, care, leisure, and sleep account for 89% of a day. The largest residual category are eating, and personal care and hygiene.

[18]Beblo (2001) points out that leisure gaps are largest for spouses that are both working full-time.

[19]Appendix Figure 3.12 shows that this relationship also held in the GDR.

[20]Looking at Figure 3.8 this may seem counterintuitive look, as women have every substantially less leisure conditional on working time. However, as can be seen in Table 3.1, the average West German woman is in the 120-180 minutes paid work bin, whereas the average West German man is in the 420-480 bin. Due to those large differences, women's leisure slightly exceed that of men.

3.7 Norms

Several determinants for women's second shift exist. As shown above, the better child care infrastructure in the GDR and East Germany seems to influence how much total housework is required and women benefit disproportionately from it. While the availability of household technology can play a certain yet small role over time (Knowles, 2013), it is likely to play only a negligible role in a comparison of East and West Germany.[21]

A potentially crucial determinant for the housework gender gap are gender-role attitudes. As support for joint family responsibilities may not necessarily coincide with support for joint earning and vice versa (Grunow et al., 2018), women in the GDR might have been particularly prone to suffering from the second shift. While female employment and the use of child care was heavily propagated under the GDR regime, relieving women of their double burden by encouraging men to take over more housework tasks was not central to the socialist ideology.

We use the World Value Survey data (Inglehart et al., 2014) from reunified Germany in 1990 to study if gender-role attitudes in East Germany favour the existence of a second shift. Figure 3.9 shows held beliefs about the role of women in the household and the labour market for East and West Germany. East German woman are 20 percentage points (pp) more likely to agree that both husband and wife should contribute to income and to disagree that being a housewife is just as fulfilling as a working mother (-22 pp). Yet, East-West differences in the importance of sharing household chores for a happy marriage are much smaller (7 pp). This is a set of beliefs that makes a second shift for women more likely. The slow evolution of norms about housework and its independence of norms in other domains should be taken into account when implementing policies promoting employment of women. Under absence of policies promoting gender equality in the household, the number of woman working a second shift has been large in the GDR.

3.8 Conclusion

This paper uses time-use data from the GDR as well as reunified Germany to document that women in the GDR worked more in the labour market and less in the household. However, we show that the reduction in housework is a mechanical product of individual time constraints alone: women who spend more time on the labour market have less time remaining for household work. What we do not find is that the partner's labour

[21] Results in Table 3.2 are very similar when we additionally control for whether households own a washing machine, dishwasher, microwave and a fridge.

supply matters. As a result, households in the GDR/East and West Germany only differ in their household-level time allocation because of the different amounts of time women spend on the labour market, and not because of different, i.e. more egalitarian, behaviour in the household.

This puts into perspective the narrative of gender equality of the GDR: women were working more but this only resulted in a second shift. We discuss implications for current times, where the lack of reaction functions still implies that once women increase hours in the labour market the only way to counter-act reductions in leisure time is to reduce housework even more. The finding implies that labour market policies that strengthen women's economic position by increasing their hours of market work run the risk of putting an unintended burden on women if their spouse does not provide additional support with housework.

More generally, our findings imply that gender norms in labour markets and regarding the within-household allocation of non-market tasks are unrelated. Time allocation in these domains is governed by a separate set of norms. We provide support for this hypothesis using data from the World Value Survey. Theories of household allocation should take this into account, in particular to explain the missing partner reactions.

3.A Tables

Table 3.1: Summary statistics of time-use data

	GDR (85 and 90)		East Germany (91/92)		West Germany (91/92)	
	Women	Men	Women	Men	Women	Men
Characteristics						
Age	38.94	41.22	40.88	43.28	42.49	45.57
Employed	0.92	0.98	0.68	0.80	0.62	0.91
High vocational degree	0.34	0.32	.	.	.	.
Upper secondary school	.	.	0.25	0.30	0.19	0.30
Children under 10 years in household	0.44	0.44	0.40	0.40	0.39	0.39
Weekday (Mo-Fr)	0.72	0.72	0.78	0.78	0.75	0.75
Time use in minutes (weekday)						
Paid work (total)	429.27	556.55	325.87	462.42	157.64	467.21
Housework	223.78	121.89	250.55	148.86	325.55	116.53
Care for others	45.14	13.15	60.58	25.48	77.63	22.99
Leisure	157.98	194.18	184.12	203.60	229.91	221.28
Observations	3237	3237	2154	2154	6309	6309

Notes: Table shows summary statistics of the time-use survey of the GDR and of the 1991/92 wave of the German time-use survey, separately for East and West Germany. Table 3.6 shows summary statistics for samples with and without children under 10 years, Table 3.7 presents the later waves of the German time-use survey.

Table 3.2: Decomposition of the housework gap - East and West Germany (1991/92)

Dependent variable	Female-male housework gap (West mean: 209 minutes)				
	(1)	(2)	(3)	(4)	(5)
East dummy	-107.337***	-12.291**	-11.231*	-15.700***	-12.572**
	(5.435)	(4.611)	(4.551)	(4.461)	(4.478)
Covariates:					
Basic controls		-7.084***	-7.145***	-7.049***	-7.347***
		(1.117)	(1.127)	(1.111)	(1.133)
Household income (5 categories)		-3.922	-4.160*	-4.675*	-4.278*
		(2.115)	(2.094)	(2.088)	(2.083)
Female paid work		-77.829***	-80.003***	-77.963***	-78.589***
		(3.578)	(3.647)	(3.587)	(3.601)
Male paid work		-6.211*	-4.798	-1.951	-4.551
		(3.041)	(3.098)	(3.037)	(3.103)
Specification of female / male work		5 FEs	10 FEs	linear	lin. & sq.
Share coef. movement due to female work		.819	.832	.851	.829
Observations		6,380	6,380	6,380	6,380

Notes: Table shows a decomposition of the female-male housework gap following Gelbach (2016). Lower rows show the contribution of the groups of explanatory variables in moving the East dummy. Basic controls: Children under 10 in household (0/1), household size, age, partner's age, education dummies. Robust standard errors in parentheses. Source: German Time-Use Survey

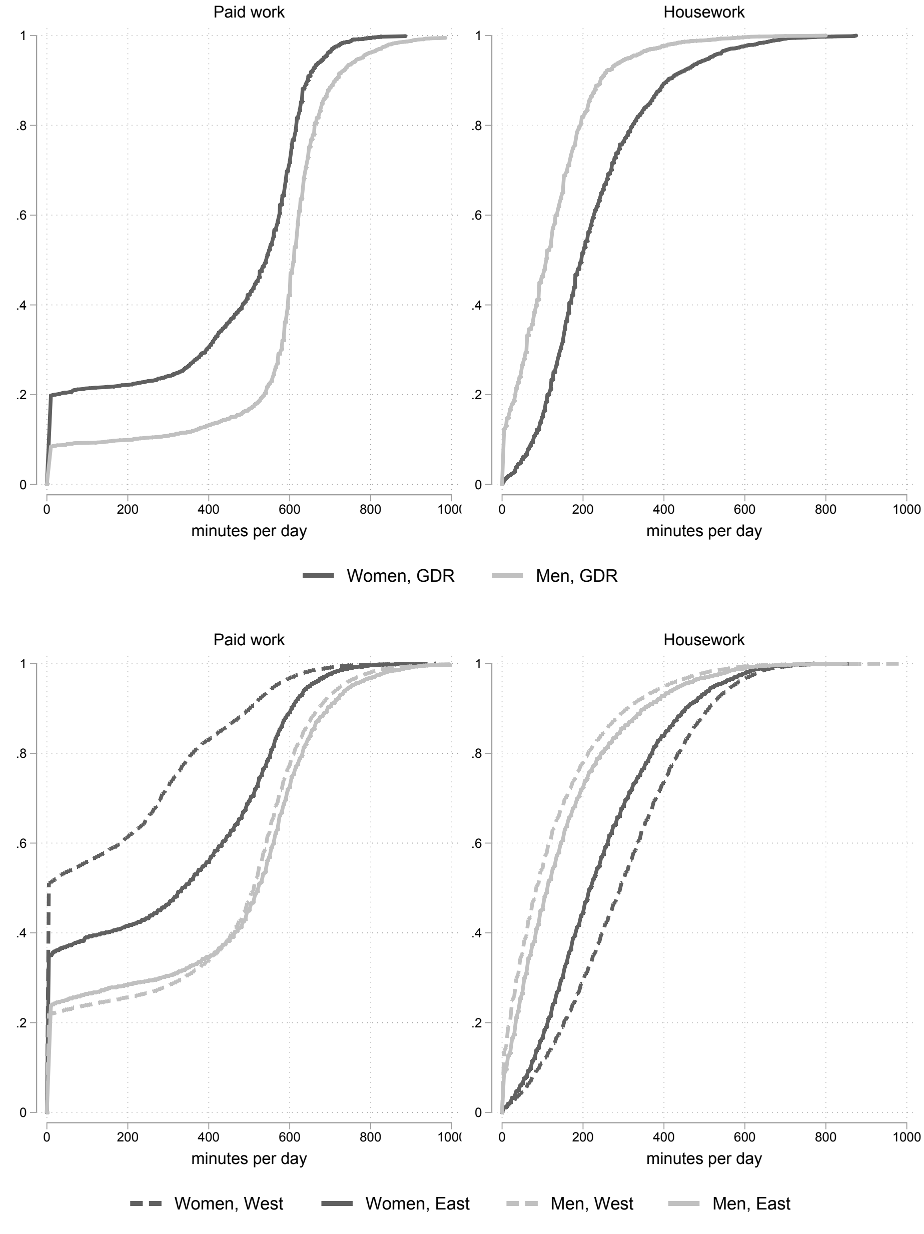
Table 3.3: Decomposition of the housework gap by children - East and West Germany (1991/92)

Dependent variable	Female-male housework gap				Female-male domestic gap	
West mean:	206 minutes		213 minutes		307 minutes	
Sample:	No children u10		Children u10			
	(1)	(2)	(3)	(4)	(5)	(6)
East dummy	-90.829***	-5.008	-131.369***	-37.557***	-163.433***	-40.006***
	(7.548)	(5.889)	(7.495)	(7.394)	(9.520)	(8.338)
Covariates:						
Basic controls		-5.183***		-8.181***		-0.436
		(1.130)		(1.956)		(1.808)
Household income (5 categories)		-5.448		0.919		2.771
		(2.944)		(2.898)		(3.332)
Female work		-69.168***		-79.720***		-117.347***
		(5.063)		(4.824)		(6.657)
Male work		-6.022		-6.829		-8.415
		(4.376)		(3.857)		(5.071)
Specification of female / male work		5 FEs		5 FEs		5 FEs
Share coef. movement due to female work		.806		.85		.951
Observations	3,868	3,868	2,512	2,512	2,512	2,512

Notes: Table shows a decomposition of the female-male housework gap and domestic work gap (housework + child care) following Gelbach (2016). Samples are split by whether children under 10 live in households. Lower rows show the contribution of the groups of explanatory variables in moving the East dummy. Basic controls: household size, age, partner's age, education dummies. Robust standard errors in parentheses. Source: German Time-Use Survey

3.B Figures

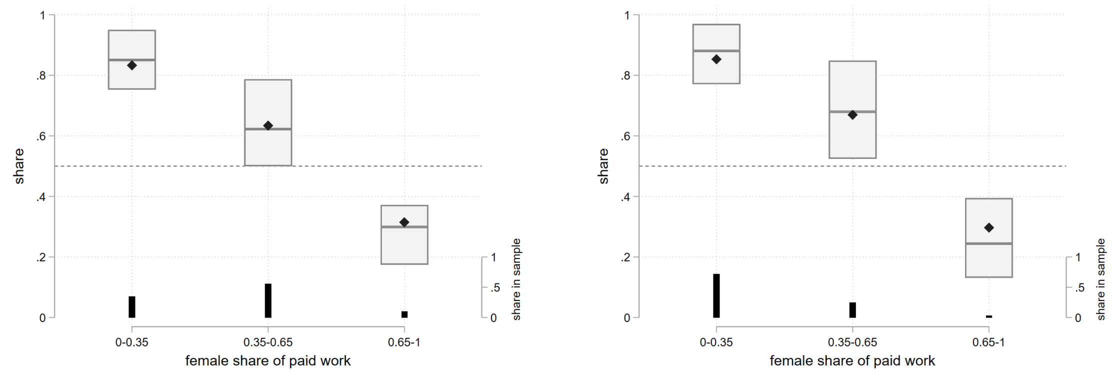
Figure 3.1: Cumulative distributions of paid work and housework



Note: The figure plots cumulative density functions of paid and housework work among women and men. The upper panel uses the GDR time budget study (1985/90), the lower panel the German time-use survey (1991/92), separately for East (solid) and West Germany (dashed).

Figure 3.2: Shares of housework by gender and share of paid work in East and West Germany (1991/92)

Panel A: East Germany Panel B: West Germany



Notes: Figure plots the males and females within household share of housework against three intervals of share of paid work. Data from the German time budget study (1991/92).

Figure 3.4: Female share of housework and paid work

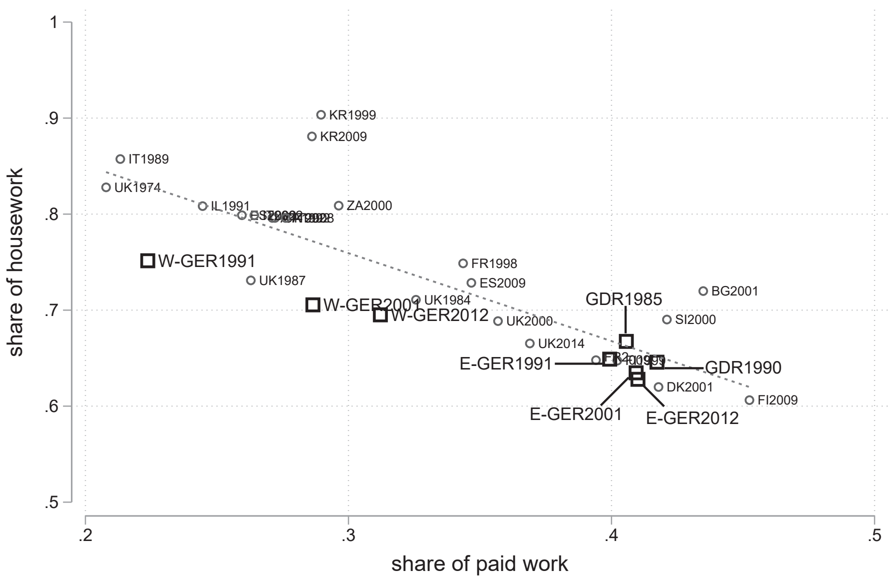
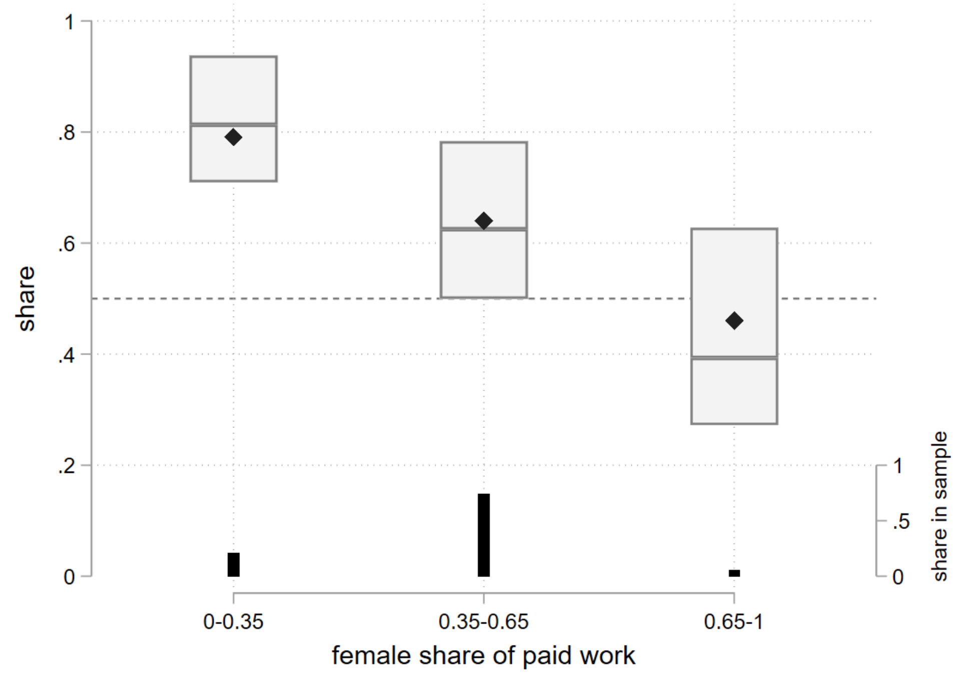


Figure plots the female share of housework and paid work in a cross-country comparison. Data from GDR time budget study (GDR), German time use data of East Germany (E-GER) and West Germany (W-GER) and the Multinational Time Use Study (other countries). Sample is restricted to different-sex couples and surveys with at least 400 couple-level observations.

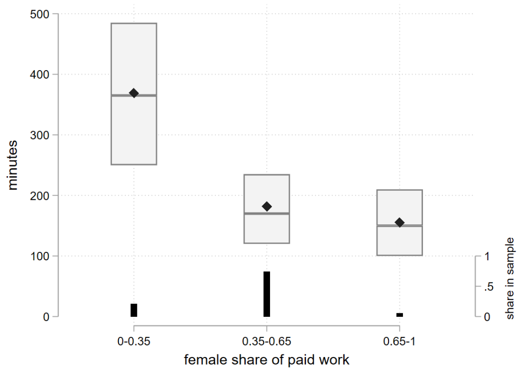
Figure 3.5: Female share of housework and paid work in the GDR



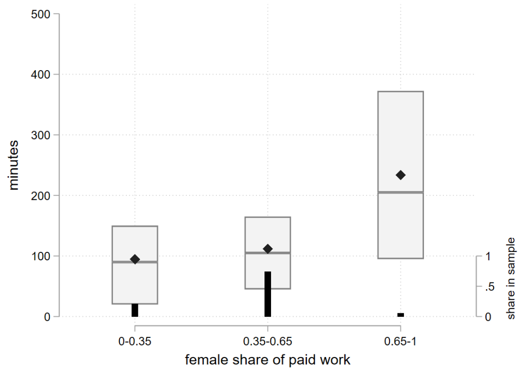
Notes: Figure plots the females within-household share of housework against three intervals of share of paid work. Diamonds indicate the mean values, range plots show 25th, 50th and 75th percentiles of the distribution. See text for additional details. Source: GDR time budget study (1985/90).

Figure 3.6: Minutes of housework by gender and share of paid work in the GDR (1985/90)

Panel A: Female housework

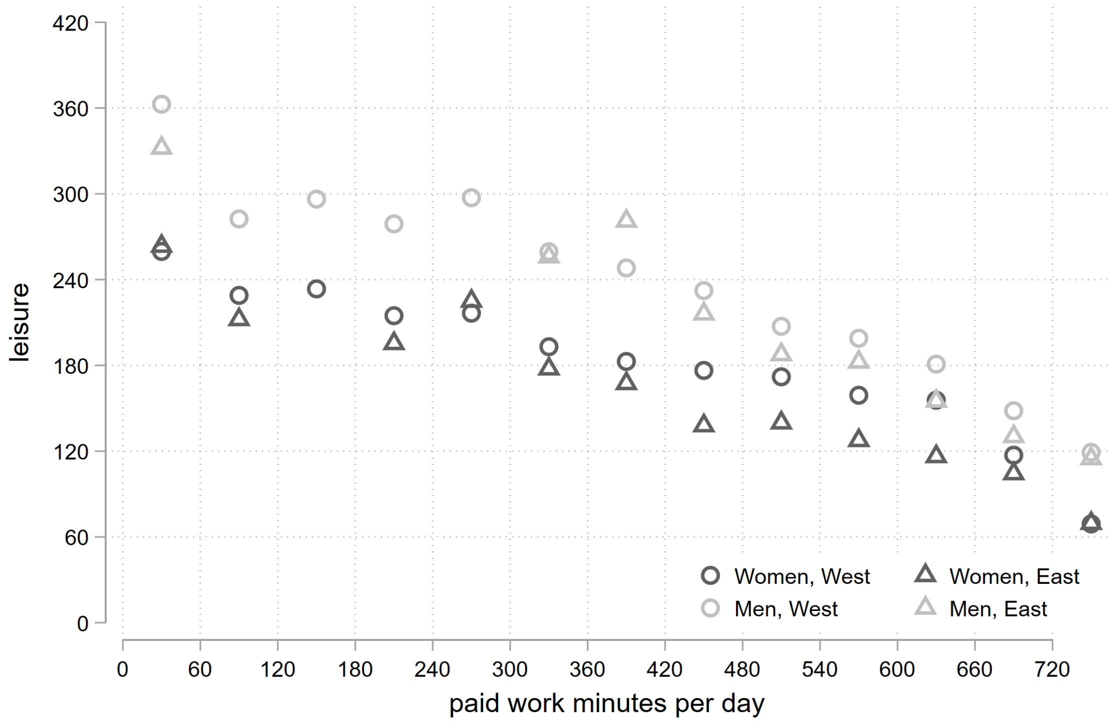


Panel B: Male housework



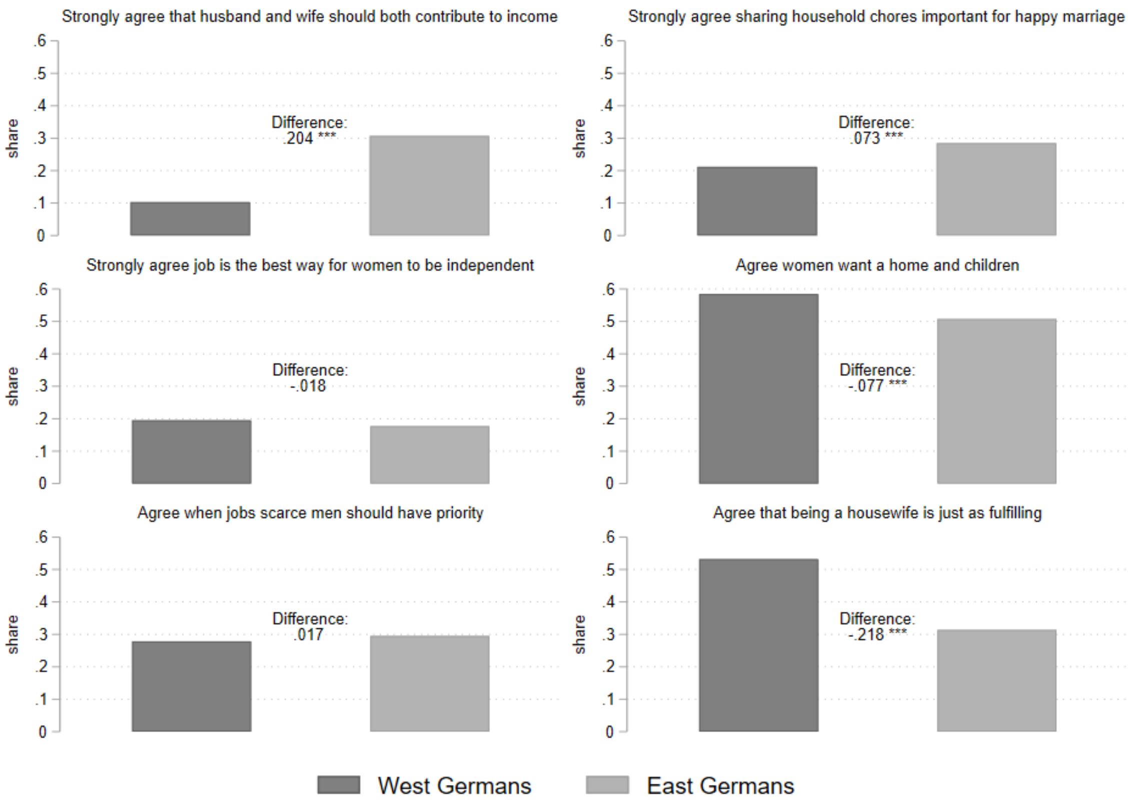
Notes: Figure plots the females and male housework in minutes against three intervals of share of paid work. Diamonds indicate the mean values, range plots show 25th, 50th and 75th percentiles of the distribution. Source: GDR time budget study (1985/90).

Figure 3.8: Paid work and leisure - East and West Germany (1991/92)



Notes: Figure plots average leisure by 60 minutes bins of paid work per weekday. Scatters with less than 20 individual observations are dropped. Source: German Time-Use Survey

Figure 3.9: Values in West and East Germany (1990)

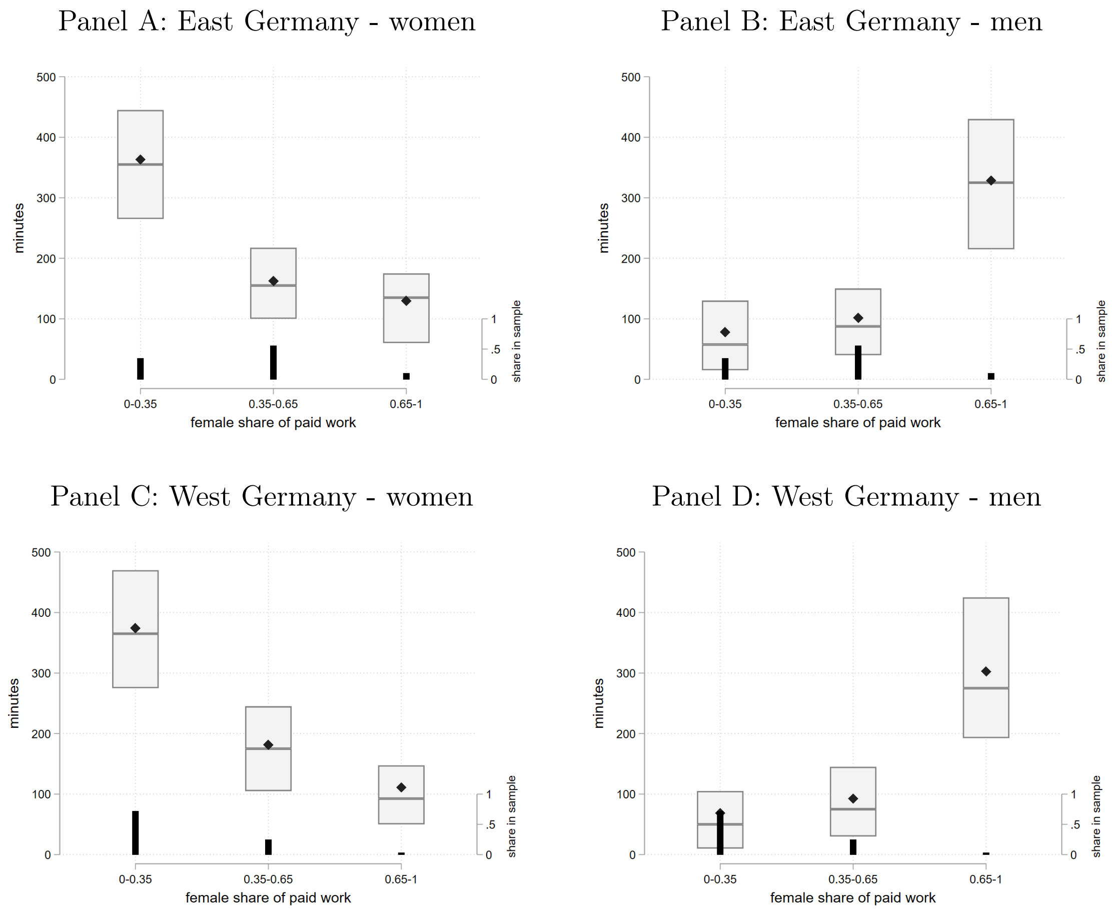


Notes: All variables are coded as binary indicators. Bars indicate the share agreeing in both regions. Source: World Value Survey

3.C Appendix

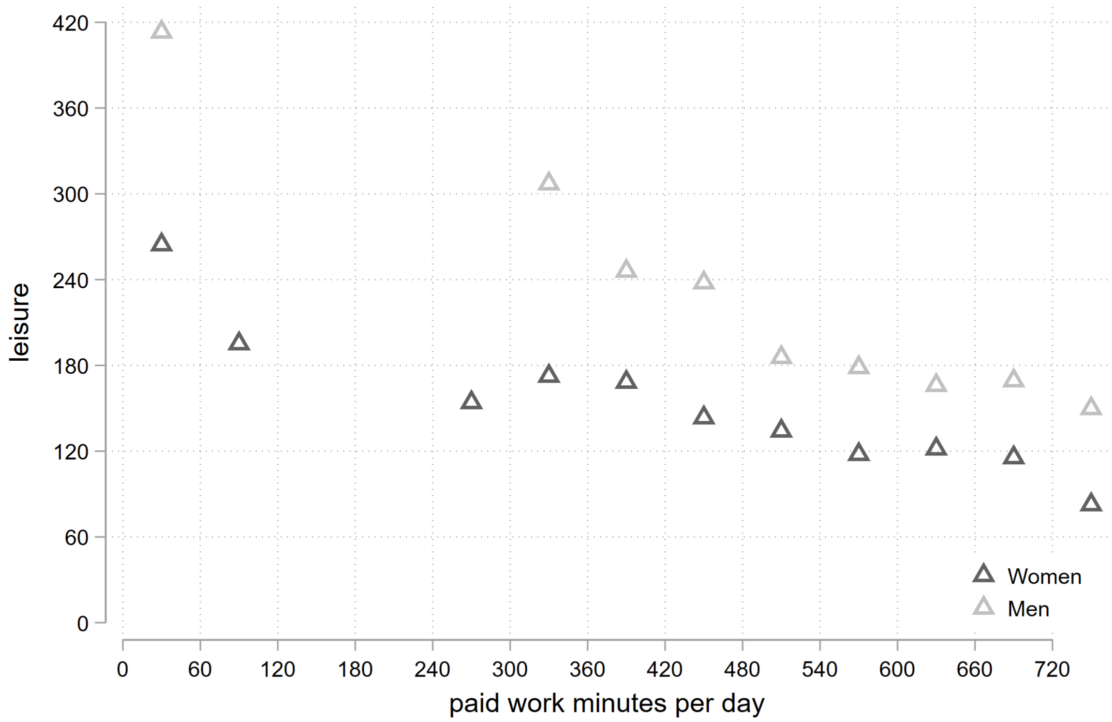
3.C.1 Additional Figures

Figure 3.10: Minutes of housework by gender and share of paid work in East and West Germany (1991/92)



Notes: Figure plots male and female minutes of of housework against three intervals of share of paid work. Data from the German time budget study (1991/92).

Figure 3.12: Paid work and leisure - GDR (1985/90)



Notes: Figure plots average leisure by 60 minutes bins of paid work per weekday. Scatters with less than 20 individual observations are dropped. Source: GDR time budget study

3.C.2 Additional Tables

Table 3.4: Countries and surveys used from the Multinational Time Use Study

Country	Survey year	Observations
Austria	1992	3,571
Bulgaria	2001	1,042
Canada	1992	3,571
Denmark	2001	1,002
Finland	1999	1,055
	2009	729
France	1998	2,275
	2009	2,602
Israel	1991	534
Italy	1989	2,582
	2002	2,716
	2008	2,292
Slovenia	2000	957
South Africa	2000	684
South Korea	1999	11,405
	2009	5,240
Spain	2002	5,010
	2009	2,043
United Kingdom	1974	2,867
	1984	461
	1987	1,971
	2000	1,864
	2014	1,138

Notes: Table shows the list of countries, surveys and number of observations that are used from the Multinational Time-Use Study. Sample is restricted to couples aged 18-65, observations are at the couple-level.

Table 3.5: Time-use data: detailed activities (minutes per weekday)

	GDR (85 and 90)		East Germany (91/92)		West Germany (91/92)	
	Women	Men	Women	Men	Women	Men
Paid work (total)	429.27	556.55	325.87	462.42	157.64	467.21
Working	364.00	471.86	266.82	396.24	131.10	403.14
Work-related (breaks, travel time etc.)	64.09	82.20	42.76	54.54	17.90	57.45
School / studies	1.19	2.49	16.29	11.64	8.64	6.63
Housework	223.78	121.89	250.55	148.86	325.55	116.53
Cooking	59.50	13.07	85.29	24.92	106.92	17.24
Cleaning	79.41	9.44	78.78	13.91	119.26	10.59
Fixing and building things	14.76	29.43	8.10	40.62	8.29	32.41
Shopping	37.76	19.40	25.39	14.63	32.16	11.99
Gardening	21.71	43.19	19.87	26.44	23.86	20.67
Other housework	10.64	7.36	33.12	28.34	35.05	23.62
Care for others	45.14	13.15	60.58	25.48	77.63	22.99
Child care	41.59	11.80	47.87	18.23	59.10	16.91
Care for adults	3.55	1.35	2.25	1.00	5.38	1.63
Leisure	157.98	194.18	184.12	203.60	229.91	221.28
Cultural activities	16.00	21.57	2.30	3.65	7.71	7.04
Sports (active and passive)	12.32	13.07	11.83	14.66	21.32	22.06
Media consumption	90.38	120.23	109.34	132.90	111.95	129.31
Social contacts	24.35	22.79	51.35	43.62	74.88	50.17
Other leisure	24.78	24.10	9.29	8.77	14.04	12.71
Observations	2328	2328	1673	1673	4707	4707

Notes: Table shows fine-grained activities that are contributing to the broader categories the analyses build on.

Table 3.6: Summary statistics of time-use data - by children

	GDR (85 and 90)		East Germany (91/92)		West Germany (91/92)	
	Women	Men	Women	Men	Women	Men
Panel A: No children under 10 years in household						
Characteristics						
Age	45.16	47.36	46.56	48.78	47.49	50.43
Employed	0.92	0.97	0.66	0.76	0.63	0.87
High vocational degree	0.29	0.35	.	.	.	.
Upper secondary school	.	.	0.21	0.29	0.13	0.26
Children under 10 years in household	0.00	0.00	0.00	0.00	0.00	0.00
Weekday (Mo-Fr)	0.72	0.72	0.77	0.77	0.75	0.75
Time use in minutes (weekday)						
Paid work (total)	446.92	548.24	318.46	439.31	182.46	444.03
Housework	231.40	127.00	278.32	162.48	331.75	125.08
Care for others	11.42	3.78	18.34	9.95	24.35	9.11
Leisure	165.54	203.15	143.78	172.31	167.69	184.72
Observations	1812	1812	1286	1286	3836	3836
Panel B: Children under 10 years in household						
Characteristics						
Age	31.04	33.41	32.47	35.14	34.74	38.02
Employed	0.92	0.99	0.72	0.87	0.60	0.97
High vocational degree	0.40	0.29	.	.	.	.
Upper secondary school	.	.	0.31	0.31	0.29	0.36
Children under 10 years in household	1.00	1.00	1.00	1.00	1.00	1.00
Weekday (Mo-Fr)	0.71	0.71	0.79	0.79	0.74	0.74
Time use in minutes (weekday)						
Paid work (total)	406.56	567.23	336.49	495.58	118.45	503.82
Housework	213.97	115.32	210.68	129.32	315.76	103.02
Care for others	88.53	25.20	121.21	47.75	161.75	44.89
Leisure	148.26	182.65	116.95	142.29	135.02	149.64
Observations	1425	1425	868	868	2473	2473

Notes: Table shows summary statistics of the time-use survey of the GDR and of the 1991/92 wave of the German time-use survey. The sample is split by children under 10 in the households.

Table 3.7: Summary statistics of time-use data - 2001/02 and 2012/13

	2001/02				2012/13			
	East Germany		West Germany		East Germany		West Germany	
	Women	Men	Women	Men	Women	Men	Women	Men
Characteristics								
Age	44.03	46.54	44.69	47.57	45.59	48.48	46.19	49.05
Employed	0.70	0.79	0.67	0.86	0.70	0.83	0.64	0.85
High vocational degree	.	.	.	.	.	.	.	.
Upper secondary school	0.27	0.28	0.31	0.37	0.33	0.36	0.41	0.45
Children under 10 years in household	0.19	0.19	0.34	0.34	0.30	0.30	0.32	0.32
Weekday (Mo-Fr)	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
Time use in minutes (weekday)								
Paid work (total)	270.60	379.48	165.95	389.50	246.75	347.15	172.74	353.70
Housework	259.59	165.72	304.21	138.72	200.84	127.78	241.21	115.45
Care for others	27.45	12.02	47.64	16.69	42.73	17.57	50.81	20.24
Leisure	237.34	254.50	263.34	265.69	235.35	243.92	249.74	251.36
Observations	1479	1479	6536	6536	1222	1222	4892	4892

Notes: Table shows summary statistics for the later German time use surveys.

Table 3.8: Decomposition of the female housework share - East and West Germany (1991/92)

Dependent variable	Female-male housework gap (West mean: 0.751)				
	(1)	(2)	(3)	(4)	(5)
East dummy	-0.102***	-0.013*	-0.013*	-0.019**	-0.014*
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
Covariates:					
Basic controls		-0.005***	-0.006***	-0.006***	-0.006***
		(0.001)	(0.001)	(0.001)	(0.001)
Household income (5 categories)		-0.010***	-0.010***	-0.010***	-0.010***
		(0.003)	(0.003)	(0.003)	(0.003)
Female work		-0.071***	-0.074***	-0.065***	-0.073***
		(0.004)	(0.004)	(0.003)	(0.004)
Male work		-0.003	-0.000	-0.003	0.000
		(0.004)	(0.004)	(0.004)	(0.004)
Specification of female / male work		5 FEs	10 FEs	linear	lin. & sq.
Share coef. movement due to female work		.794	.826	.783	.826
Observations		6,372	6,372	6,372	6,372

Notes: Table shows a decomposition of the (log) female-male housework gap following Gelbach (2016). Lower rows show the contribution of the groups of explanatory variables in moving the East dummy. Basic controls: Children under 10 in household (0/1), household size, age, partner's age, education dummies. Robust standard errors in parentheses. Source: German Time-Use Survey

Table 3.9: Decomposition of the housework gap (cooking, cleaning, shopping) - East and West Germany (1991/92)

Dependent variable	Female-male housework gap (West mean: 219 minutes)				
	(1)	(2)	(3)	(4)	(5)
East dummy	-82.526***	-12.993***	-11.800**	-14.425***	-12.384***
	(3.968)	(3.836)	(3.809)	(3.740)	(3.741)
Covariates:					
Basic controls		-7.100***	-7.065***	-7.011***	-7.202***
		(1.019)	(1.022)	(1.016)	(1.030)
Household income		0.656	0.450	0.102	0.355
		(1.766)	(1.752)	(1.749)	(1.746)
Female paid work		-60.064***	-61.610***	-60.355***	-60.821***
		(2.780)	(2.828)	(2.796)	(2.793)
Male paid work		-3.025*	-2.501	-0.837	-2.473
		(1.346)	(1.373)	(1.303)	(1.366)
Specification of female / male work		5 FEs	10 FEs	linear	lin. & sq.
Share coef. movement due to female work		.864	.871	.886	.867
Observations		6,380	6,380	6,380	6,380

Notes: Table shows a decomposition of the female-male housework gap following Gelbach (2016). Lower rows show the contribution of the groups of explanatory variables in moving the East dummy. Basic controls: Children under 10 in household (0/1), household size, age, partner's age, education dummies. The housework definition in this table is more narrow than the one shown in Table 3.2. See Table 3.5 for an overview of subcategories of household. Robust standard errors in parentheses. Source: German Time-Use Survey

Table 3.10: Decomposition of the housework gap - East and West Germany (2001/02, 2012/13 and pooled)

Dependent variable	Female-male housework gap					
West mean:	165 minutes		126 minutes		172 minutes	
Survey wave:	2001/02		2012/13		Pooled	
	(1)	(2)	(3)	(4)	(5)	(6)
East dummy	-71.622***	-9.184	-52.700***	-11.783*	-83.523***	-13.565***
	(6.857)	(5.107)	(6.550)	(4.903)	(3.611)	(2.744)
Covariates:						
Basic controls		-6.054***		-5.201***		-6.382***
		(1.453)		(1.409)		(0.719)
Household income (5 categories)		1.082		0.324		0.139
		(0.977)		(0.958)		(0.723)
Female paid work		-50.926***		-33.370***		-59.131***
		(4.381)		(3.910)		(2.293)
Male paid work		-6.539		-2.670		-1.998
		(4.541)		(3.965)		(2.189)
Survey wave						2.009***
						(0.394)
Specification of female / male work		5 FEs		5 FEs		5 FEs
Share coef. movement due to female work		.816		.816		.905
Observations	5,313	5,313	4,039	4,039	15,732	15,732

Notes: Table shows a decomposition of the female-male housework gap following Gelbach (2016) by survey wave. Lower rows show the contribution of the groups of explanatory variables in moving the East dummy. Basic controls: household size, age, partner's age, education dummies. Robust standard errors in parentheses. Source: German Time-Use Survey

Table 3.11: Decomposition of the leisure gap - East and West Germany (1991/92)

Dependent variable	Female-male leisure gap							
West mean:	8.6 minutes		-2.35 minutes		-1.62 minutes		2.05 minutes	
Survey wave:	1991/92		2001/02		2012/13		Pooled	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
East dummy	-28.110***	7.475	-14.817**	2.063	-6.944	10.341*	-19.047***	7.010**
	(3.886)	(4.453)	(5.613)	(5.066)	(5.912)	(5.155)	(2.837)	(2.701)
Covariates:								
Basic controls		2.708***		5.933***		3.309**		3.535***
		(0.690)		(1.276)		(1.076)		(0.526)
Household income (5 categories)		-3.205		-0.332		-1.416		-1.630*
		(2.050)		(0.920)		(1.005)		(0.696)
Female work		-32.895***		-20.581***		-17.159***		-25.646***
		(1.950)		(2.154)		(2.232)		(1.200)
Male work		-2.193		-1.901		-2.019		-0.153
		(1.732)		(2.710)		(2.873)		(1.339)
Survey wave								-1.448***
								(0.288)
Share coef. movement due to female work		.924		1.219		.993		1.012
Observations		6,380		5,313		4,039		15,732

Notes: Table shows a decomposition of the female-male leisure gap following Gelbach (2016). Lower rows show the contribution of the groups of explanatory variables in moving the East dummy. Basic controls: Children under 10 in household (0/1), household size, age, partner's age, education dummies. Robust standard errors in parentheses. Source: German Time-Use Survey

CHAPTER 4

Improving Preschool Provision and Encouraging Demand: Heterogeneous Impacts of a Large-Scale Program[1]

4.1 Introduction

Robust early childhood development lays the foundation for greater human capital development in childhood and beyond (Heckman, 2006; Cunha and Heckman, 2007; Almond et al., 2018; World Bank, 2018). Enrollment rates in pre-primary education are growing worldwide (UNICEF, 2019)—a momentum supported by strong evidence that early child nurture and stimulation can lead to positive impacts on later outcomes (Shonkoff et al., 2016; Black et al., 2017; Britto et al., 2017). While preschool programs can be highly effective, much of this evidence is based on relatively small targeted programs (Gertler et al., 2014; Elango et al., 2015). Whether positive effects on early development can be sustainably engendered through large-scale programs is less clear,

[1]This chapter is joint work with Adrien Bouguen (University of Santa Clara), Deon Filmer (World Bank) and Tsuyoshi Fukao (World Bank). The project is a joint effort between the Cambodian government, the World Bank, a team of field researchers (Angkor Research), and a team of academic researchers. We are particularly grateful to the World Bank for its constant support: special thanks to Simeth Beng. This research was funded by SIEF. Special thanks to Alaka Holla for her useful comments throughout the program's duration. The project could not have been conducted without the collaboration of the Ministry of Education, Youth and Sport in Cambodia. We are particularly grateful to Sok Sokhom and Lynn Dudley. Data were collected in the field by a team of very dedicated field researchers from Angkor Research: special thanks to John Nicewinter, Ian Ramage, Benjamin Lamberet, Kimhorth Keo, and Ratanaksophea Saing. Finally, many researchers also contributed to this research through their very useful comments: Craig McIntosh, Patrick Kline, Christopher Walters, Karen Macours, Diego Vera, Markus Frölich, Paul Gertler, Supreet Kaur, Edward Miguel, Katja Kaufmann, Elisabeth Sadoulet, Analia Schlosser, Alain de Janvry, Clément de Chaisemartin, Antoine Camous, Harald Fadinger, Luc Behaghel, Clement Imbert, and Marc Gurgand. The findings, interpretations, and conclusions expressed in this paper are those of the authors and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments they represent. The study was preregistered at AEA's Social Science Registry (AEARCTR-0001045).

particularly when the quality of services is low (Engle et al., 2011; Britto et al., 2011; Ichino et al., 2019; Andrew et al., 2019) or services are not accessed by children who might benefit from them (Cornelissen et al., 2018). The issue of scalability concerns high-income as well as low- and middle-income countries.

In this study, we evaluate the impacts of a large-scale preschool construction program on school participation and child development in a low-income country setting, Cambodia, using a randomized controlled trial. The main objective of the construction program is to provide a quality preschool experience to every child in the treatment villages who meet the age criteria (3-5 years old). We assess whether the construction of new preschools is sufficient to induce behavioral responses, and whether additional awareness campaigns about preschool services matter to stimulate enrollment, parental involvement and, ultimately, child cognitive and socio-emotional development. Our evaluation has three treatment arms: the construction of formal community preschools (CPS) and associated activities within villages (which we refer to as T1); the addition of a home visit by a village chief plus a door-to-door campaign to promote awareness about CPS and the value of education (T2); and the further addition of a home-based program consisting of trained "core parents", who provided more intensive caregiver training sessions that focused on good parenting, the value of nutrition, and the importance of preschool (T3). The 305 study villages were randomly assigned either to a control group (C) or one of the treatment arms.

The number of experimental studies on the effectiveness of large-scale early childhood education and care programs in low-income countries is limited and results tend to be mixed. While positive short-term effects on child development are found for Mozambique (Martinez et al., 2017b), other studies find insignificant or even negative effects (Brinkman et al., 2017; Bouguen et al., 2018a; Blimpo et al., 2019; Bernal et al., 2019).[2] Differences in socio-economic conditions, along with the nature and quality of preschools offer and its counterfactual (quality of other preschool arrangements or quality of parenting), make it difficult to draw broad conclusions, particularly when these differences are unobservable or undocumented. Specifically, the literature often lacks detailed information on both the quality of the preschool supply and a clear understanding of the preschool demand mechanisms. Similar issues plague the literature in higher income countries where large and long-lasting results (Belfield et al., 2006; Carneiro and Ginja, 2014) coexists with disappointing (Puma et al., 2012) and even negative results (Baker et al., 2008, 2019).

[2]Evidence on preschool programs in low- and middle-income countries based on non- and quasi-experimental study designs generally points to positive effects of preschool attendance (Berlinski et al., 2008, 2009; Engle et al., 2011; Rao et al., 2014).

Our contribution to the literature is threefold. First, we provide evidence on a large, well-implemented public preschool construction program in rural Cambodia, with detailed information on both children and on the quality of the new preschool services provided. The key outcomes of interest are preschool and subsequent school enrollment, along with child cognitive and socio-emotional development. We collected three rounds of data on about 7000 children to analyze impacts: a baseline survey at the start of the program, a midline survey one year later, and an endline survey two years later. Information on households, villages and preschools (including in-class observation) allows us to investigate the extent to which impacts vary by background characteristics, including family socioeconomic status.

Second, using T2 and T3 treatment branches, our study analyzes two different strategies to increase preschool demand. Previous research has shown, that the provision of information about the return to education can affect education choices (Nguyen, 2008; Jensen, 2010). Therefore, preschool construction was complemented with two different demand-side interventions, the door-to-door program and the home-based program, that aimed to increase preschool enrollment by increasing caregiver's awareness about the availability and importance of preschool education, and to strengthen demand through community mobilization. In addition, the goal was to use demand-side interventions to shift additional children into preschool in order to enable inferences to be made about the impact of preschool on marginal beneficiaries (Carneiro et al., 2011, 2017; Nybom, 2017; Cornelissen et al., 2018).

Third, using our detailed classroom observations, we precisely document the quality of the new preschools in comparison with the available alternatives. While most of the experimental or quasi-experimental studies on preschool programs provide some indicators of "structural" quality (such as class size, child-teacher ratio or teacher qualifications), "process" quality (such as the pedagogical practices and the quality of teacher-child interaction) is less often observed and therefore cannot be compared to other available preschool services (or between studies). This is a limitation since process quality seems particularly predictive of child development (Araujo et al., 2016, 2019). In a large experiment in Colombian public preschools, Andrew et al. (2019) show that pedagogical training can improve the learning environment and children's cognitive development while improvements in structural quality alone do not. At the same time, however, other evidence on teacher training programs show mixed effects with positive impacts often fading out quickly (Yoshikawa et al., 2015; Ozler et al., 2018; Wolf et al., 2019). Through detailed classroom observations, we are able to explore the degree to which the establishment of the new preschools improved structural and process quality of CPS compared to its alternatives in Cambodia and how this mediates the effectiveness of the intervention.

In this paper, we study reduced form effects for each treatment group using midline and endline data, with a focus on subgroup effects and the quality of the new CPS and other available preschools. Yet, in some villages, other preschools were already present. Specifically, in many villages, the new CPS replaced an already existing informal preschool (IPS) which was provided at the teacher's home, at the local Pagoda, or outdoors, and whose quality was very limited.[3] In more rare cases, villages had access to a formal state preschool (SPS)—co-located in a primary school and of arguably better quality than both IPS and CPS. These SPS remained available to preschool-age children after the new CPS opened while the IPS were shut down.[4] In this context, reduced form effects are driven by children who would not have benefited from any preschool absent the program as well as children switching from another preschool arrangement to the CPS. These reduced form effects are empirically valid and policy relevant, as they reflect the program's impacts in a typical Cambodian preschool environment. A companion paper, Berkes and Bouguen (2019) explicitly studies the issue of substitution and effects by unobserved counterfactuals based on midline data only.[5]

We have four main findings. First, the construction of CPS—the implementation of which closely followed the program and experimental design—increased preschool participation. Children in villages with a newly built CPS were about 11 percentage points (pp) more likely to have ever attended a preschool by the time they were between four and six years old (compared to 59% in the control). Put differently, they were enrolled, on average, about one month more than children in the control group (who had been enrolled for about 5.4 months). Using detailed information on the quality of preschools, we find that the CPS construction lead to a substantial increase in the quality of the infrastructure and material but did not improve the quality of educational processes.

Second, the average intent-to-treat impacts of the intervention on child development outcomes (i.e., the simple comparison of children in treatment versus control villages)

[3]In a previous paper based on an earlier version of this program, Bouguen et al. (2018a) finds that IPS had no impacts on child development

[4]Note that we use the terms CPS and IPS for expository clarity here; this is not the formal or usual nomenclature in Cambodia where both types of preschool are considered "informal".

[5]Using a constant subLATE framework (Kline and Walters, 2016; Hull, 2018), Berkes and Bouguen (2019) find that the modest average reduced form effects are consistent with substantial effects on children who would have stayed at home absent the program, while the effect on children who would have gone to existing preschools (IPS and SPS) is small and insignificant. A similar pattern is found by Dean and Jayachandran (2020) who study effects of attending kindergarten in India and find that short-run effects on cognition are driven mostly by children who would have not attended kindergarten absent the program.

were small one year after the program started and had disappeared two years afterwards. More specifically, children in treatment villages scored about 0.04 standard deviation (SD) higher on an index of cognitive development (made up of measures of early literacy and numeracy along with executive function), and about 0.09 SD lower on a measure of socio-emotional problems (the Strengths and Difficulties Questionnaire) after one year of implementation.[6] After two years, these impacts were smaller and no longer statistically significant.[7] As discussed below, the lack of persistence of early impacts is consistent with results found in other contexts and is not necessarily inconsistent with long-lasting impacts.

Third, these overall results mask considerable variation. Strikingly, the impacts on cognitive development for children from the wealthiest or most educated quartile of households in the sample were large and statistically significant: 0.09 SD after one year of the program and 0.13 SD after two years (while impacts on enrollment levels were similar across wealth quartiles). While preschool is sometimes suspected to be a substitute of parental involvement in high income countries (Baker et al., 2008), our results suggest that less disadvantaged families were better able to take advantage of the CPS in this case. We show that, when treated, more educated and wealthy parents are more likely to enroll children in primary school and we suggest that they offer a home environment that complements the schooling they receive. We also have evidence that wealthier and more educated parents increase their parental involvement when their children enroll at school.

Fourth, we find that enrollment increases were not different across the treatment arms. Children in villages where the demand-side interventions were deployed were not systematically more likely to enroll in CPS (or any type of preschool). This was despite the fact that caregivers in those villages are more likely to recall having received a leaflet promoting preschool enrollment (+8 pp) and participating in home-based program sessions (+10 pp). Our findings therefore suggest that the demand-side intervention did not modify the effects of the supply-side component of the program: it did not significantly increase enrollment and did not modify the performance of the CPS. While substantial effort went into designing and implementing the demand-side component, we cannot rule out the possibility that it did not have the intensity required to make a large enough difference and that a more intensive program (e.g., more frequent household visits, more intensive home-based programs) might have had bigger impacts.

[6] These ITT impacts are consistent with substantially large LATE impacts. Yet, the complex substitution patterns make the interpretation of the LATE complex. This issue is covered in our companion paper (Berkes and Bouguen, 2019)

[7] While the time difference between the midline and endline is only one year, the two-year follow-up captures a very different enrollment scenario than the one-year follow-up. At the two-year follow-up, 26% of the sample is enrolled at primary school, in stark contrast to the one-year follow-up (4%).

At the same time, the construction of the CPS itself had a demand-side "information" component, in the sense that it was likely an important village event that households would have known about. It is possible that, in the context of small and rural villages in Cambodia, additional demand-side programs did not convey any information over and above the construction itself.

There are various potential reasons why we find relatively modest average impacts. First, much of the enrollment change stemmed from enrollment shifts across preschool types rather than increases in overall enrollment in preschool. Indeed, while some CPS enrollments came from households that would not have enrolled absent a newly built CPS, the construction of CPS induced a large reduction in informal preschool (IPS) enrollment and a more modest reduction in state preschool (SPS) enrollment. Since the construction of CPS aimed to upgrade informal arrangements and ensure a higher and uniform quality standard, the substitution between CPS and IPS was expected. Less expected, however, was the slight reduction (-8 pp) in enrollments at SPS. Reduced-form effects reflect both the enrollment of children who would not have enrolled in school in the absence of the program and the substitution between CPS and preschools that were already available to parents, IPS and SPS.

Second, by design, establishing the new CPS led to substantial increases in infrastructure and material availability either in comparison to the IPS (which included no formal structure and mainly self-produced materials) or to locations with no preschool. But this improvement in so-called "structural" quality was not accompanied by a concomitant improvement in the quality of the education content ("process" quality). Classroom observations show that CPS were substantially better than IPS in structural quality (except teacher characteristics), and also different to SPS (teacher characteristics were worse, classroom setting was similar, classroom equipment was better in CPS than SPS); they also show that curriculum content and quality of pedagogy, as well as the frequency and quality of teacher-child interactions, were only slightly better in CPS than IPS—while those in SPS were substantially better. Our results are therefore consistent with the literature showing that pedagogical practices and the quality of teacher-child interactions are particularly important for child development (Araujo et al., 2016, 2019; Andrew et al., 2019).

Third, our finding of small short-term impacts followed by a later (medium-term) fade-out, for example once children are in primary school, is not unprecedented in the literature. Recent studies in Colombia and Malawi have documented similar patterns (Andrew et al., 2018; Ozler et al., 2018). Evidence from the United States suggests that medium-term fade-out might nevertheless be consistent with long-term improved outcomes (Heckman et al., 2013) perhaps because other—non-measured—aspects of child development improved in the short-run (Duncan and Magnuson, 2013). In addition, other research suggests that reaping the full medium- and long-term benefits of preschool might require complementary investments at later stages of human capital development. For example, it is possible that high-quality primary schools, which may be in short supply in Cambodia, are required to sustain impacts.[8] If wealthier or more educated households are more able to invest in complementary and high-quality inputs during preschool and then into the primary school years (including by having access to higher quality primary schools), this might explain the sustained impacts on cognitive development for children from those households. Further research that tracks individuals over longer periods of time would be necessary to investigate these various hypotheses.

The paper is structured as follows. Section 4.2 provides some context and background, and then details the program; Section 4.3 describes our evaluation design and the data we use in our analysis (including information on covariate balance across treatment arms and on attrition); Sections 4.4 and 4.5 respectively outline our empirical framework and discuss the findings. A final brief section concludes.

4.2 Background and Program Design

4.2.1 Early Childhood Development programs in Cambodia

Despite two decades of robust economic growth, Cambodia remains one of the least developed countries in Southeast Asia, with a GDP per capita estimated at $1,160 in 2015 ($3,300 in PPP terms). The country also faces multiple challenges in the education sector. With a preschool enrollment rate in 2009 of 40% among five-year-olds (MoEYS, 2014), Cambodia fares poorly in comparison to neighboring Thailand and Vietnam.[9] To increase the capacities and quality of its education system, the Cambodian government received a first grant from the Global Partnership for Education (GPE I) of $57 million for the period 2008—2012. The government, in cooperation with the World Bank, used part of the resources to invest in the expansion of the national early education system, which is composed of formal preschools, informal preschools, and parenting programs. Over the period of this grant, preschool enrollment of five-year-olds increased from 40% in 2009 to 56% in 2012 and 66% in 2016, while enrollment of three- and four-year-olds remained at a low 20% and 37% in 2016, respectively (MoEYS, 2014, 2017). Bouguen et al. (2013) have evaluated the impact of the early

[8]Johnson and Jackson (2018) document such dynamic complementarities between the Head Start program and investments in primary and secondary schools in the United States.
[9]Source: Data from UNESCO Institute for Statistics.

childhood education and development (ECED) components of GPE I on child development outcomes. The remaining \$18 million were spent to improving the education system as a whole (teacher training, scholarship, national assessment). They find no impacts on outcomes overall, although the study period was marked by implementation issues including low individual take-up and delays in program implementation.

To improve primary school readiness and be on track with the Sustainable Development Goals,[10] the Government of Cambodia, with the support of the World Bank, launched another education expansion program for the period 2014—2018. The plan was financially supported by a second GPE grant (GPE II) of \$38 million, again administered by the World Bank, with the objective of strengthening the existing foundation of Cambodia's education system. Of this amount, about \$20 million were allocated to a component focused on ECED programs. Our study focuses on the sub-component that includes the construction of formal community preschools.

4.2.2 Supply-side intervention: Formal community preschools

Before GPE II, two distinct types of public preschools existed in Cambodia: state preschools (SPS) and community preschools. Since these community preschools were unstructured and lacked uniform quality standards, we refer to them here as informal (community) preschools (IPS).[11,12] GPE II introduced a new type of community preschool with a structured age-appropriate educational program and uniform quality standard, which we refer to as (formal) community preschools (CPS).

SPS are financed by the Ministry of Education, Youth and Sport (MoEYS). SPS teachers benefit from two years of formal training in a MoEYS teacher training center in Phnom Penh. They receive a monthly salary of roughly \$250 to teach for three hours a day, five days a week. Almost all SPS are attached to a public primary school and have access to classrooms equipped with teaching and play materials, along with better overall infrastructure (including sanitation facilities).[13]

In contrast, IPS are typically not attached to a primary school. The local community establishes the IPS and covers operational costs. This includes the IPS teacher stipend, which is at the discretion of the local commune council. Constrained by council budgets,

[10] The SDG 4.2 states that all children should benefit from at least one year of pre-primary education by 2030.

[11] According to government data (MoEYS, 2017), of the 7,241 preschool facilities in Cambodia in 2016, 55% were SPS, 39% were IPS, and 6% were private preschools. However, these preschools are not evenly distributed across the country, and 38% of the 1,646 communes in Cambodia had no preschool facility.

[12] See Bouguen et al. (2013) for an impact evaluation of each type of preschool developed in the context of the GPE I.

[13] The program designers and research team did not anticipate the presence of SPS in the experimental villages. Villages with a preexisting SPS were not supposed to be eligible for the program.

it varied from $30 to $50 per month at the time of our baseline; most IPS teachers rely on other income sources for their livelihood. IPS teachers are trained for about 35 days by provincial education departments before they begin work. Teachers provide a two-hour preschool class five days a week. The quality of IPS can differ substantially across villages as communes have to establish IPS with their own funds. Consequently, IPS classes are often held in a teacher's home, a community hall, or a pagoda. In most cases, IPS lack even the most rudimentary equipment and materials.

To increase preschool access and to improve the unsatisfactory quality of IPS, the Cambodian government used the GPE II grant to establish 500 new CPS. Most of these replaced previously existing informal arrangements; some were established in villages that previously had no preschool or were too large to be served by one preschool alone. In contrast to poorly resourced and low-quality IPS, CPS benefit from uniform quality standards such as a standardized building directly financed by the GPE II. CPS have a capacity of 25 children and are fully equipped with tables, chairs, and a blackboard.[14] MoEYS is responsible for the curriculum, teacher recruitment and training, and ongoing monitoring of the facility, including regular payment of teacher salaries. CPS teacher training lasts 35 days and includes structured lessons in pedagogical strategies, curriculum content, testing, and how to operate a CPS. Teachers are also trained in the basics of child development, child rights and parental education. All teachers participate in a written examination before and after the training. Further, teachers are provided with a package of teaching materials tailored to the CPS curriculum. The CPS teacher is usually a community member who, after completion of the training, gives a two-hour class each day, five days a week, to children aged three to five years. It is important to recognize that while the CPS "replaced" IPS, the improvement was substantial in terms of facility infrastructure and availability of materials, as well as in terms of the quality of teacher training.

Irregular payments of IPS teacher salaries and the lack of a specialized building made the IPS model unsustainable, and many schools planned under GPE I either did not open or operated only for a short time (Bouguen et al., 2013).[15] Given these problems with the IPS model and the high costs of SPS, the Cambodian government considered CPS a promising, less costly alternative that—due to a better curriculum and uniform quality standard—could prove similarly effective to SPS. CPS require fewer resources for building construction and teacher salaries than SPS, since teachers are relatively less educated and are recruited locally. A similarly quick and large scale-up of SPS would have been significantly more costly (and likely unfeasible). It also

[14] See the pictures in Figure 4.1 for an example of the standardized CPS building.

[15] These villages were often described as being IPS villages, even though the IPS had ceased to function regularly.

would have proven difficult due to the more intensive teacher training program and the lack of sufficient personnel to train many teachers at once. The Cambodian government therefore decided to scale up CPS as a feasible intermediate model between IPS and SPS. The primary conceptual differences between CPS and IPS are the standardized building, equipment, and monitoring of regular teacher salary payments. CPS teachers do not receive the same salary, training, and support as SPS teachers.[16]

4.2.3 Demand-side intervention: Door-to-door awareness campaign and parenting program

The door-to-door (D2D) program implemented as a part of this intervention was a demand-side intervention aimed at stimulating demand for ECED programs by speaking directly to individual caregivers. The goal was to sensitize them to the value of preschool education and guide them through the enrollment of their children at CPS. An additional component was to provide information about returns to education. Such information has been shown to effectively increase attendance and change the social composition of students in other lower-income contexts (Nguyen, 2008; Jensen, 2010). The local village head and the field staff responsible for the study's data collection performed the D2D activities. At baseline, caregivers were informed about the new preschools and key details, such that a preschool has been constructed and that there is no school fee. In addition, caregivers received a printed leaflet that had more information about the newly established CPS. It noted that the teacher had been trained and it suggested how preschool in general could help children improve primary school readiness and, potentially, their overall educational attainment. In addition, the leaflet provided information about average income in Cambodia by educational background, visualized in a graph and using data from the Cambodian Socioeconomic Survey 2009. Caregivers received another informational leaflet about one year later, at midline.[17]

The home-based program (HBP) formerly operated in Cambodia as an independent early childhood education service to support parents of children aged zero to five years. However, it was redesigned as supplementary to CPS, aiming to enhance the effect of CPS enrollment. The program is implemented by local "core parents" who receive initial and ongoing training from MoEYS. The 35 day long training covers a wide range subjects such a child rights, pre- and postnatal care of mothers, hygiene, nutrition,

[16]Based on initial findings from this impact evaluation in 2018, after our endline survey, the government provided additional funding to communes to significantly increase salaries of 500 CPS teachers. For these teachers, CPS class duration increased from two to three hours and responsibility for the HBP was assigned to CPS teachers.

[17]The content of the leaflets was developed in cooperation with MoEYS. After baseline, we received feedback from village heads that the leaflet contained too much information (see Figure 4.2). Therefore, the midline leaflet was simplified to focus only on the advent of a new CPS (see Figure 4.3).

disease prevention, developmentally appropriate activities for children, school readiness, disabilities, health services, and child protection. Similarly to CPS teachers, core mothers participate in a written examination before and after the training. They are responsible for promoting enrollment of children aged three to five years into CPS and for leading monthly informational meetings with parents of children aged zero to six years. “Core parents” are volunteers who only receive stipends while in training. The HBP was supposed to take place regularly; it was designed as a more intensive demand-side intervention than the D2D.

4.2.4 Costs of CPS and other preschools

Before turning to the evaluation design and findings, it is useful to briefly review the costs of the program.[18] Total annual costs for running 500 CPS, of which 250 were initially planned to be constructed within the study sample, were estimated to be $3,136,743. The average annual costs per school were $6,273. Using the average class size of 24.5 from the endline survey, this implies annual costs of $256 per child, which compares to roughly 22% of the GDP per capita ($1,160 in 2015). The estimated annual SPS costs are much higher at $12,602 per school and $426 per child.[19]

4.3 Evaluation Design and Data

4.3.1 Randomization and sampling

This evaluation of the CPS program is based on a randomized controlled trial. All sample villages are situated in the south and northeast parts of Cambodia (see map in Figure 4.4). Eligibility criteria for villages to participate in the study were: expressed demand for a CPS, a high poverty rate, and a large number of children aged zero to five years.

The study sample is composed of 305 villages. Before baseline, villages were randomly assigned to the control group or one of the three treatment groups: CPS (T1); CPS and D2D (T2); or CPS, D2D, and HBP (T3). Randomization was stratified to

[18] A full discussion of cost estimates can be found in the SIEF impact evaluation report (Berkes et al., 2019a). The cost estimates are based on the ingredient method which aims to cost every resource required to make an intervention happen. This includes construction costs as well as support costs, such as management, administrative and overhead costs which support the intervention above the level of direct community level implementation. Therefore, our cost measures are not directly comparable to studies that base their cost-effectiveness analyses on programmatic costs only (e.g. Brinkman et al. (2017)).

[19] The cost estimates reported here are based on a model using mid-range estimates for unobserved costs. Under all models, costs per child at SPS were 50%—126% higher than at CPS. Due to the lacking uniform quality standards no cost estimates for IPS were calculated.

obtain a sample for which treatment is balanced within each of the 13 provinces of our sample.[20] The design is summarized in Table 4.1.

Table 4.2 gives an overview of data collection activities and timing of the preschool construction. Our analysis is based on three main waves of data collection: a baseline data collection in May—July 2016, a midline survey in April—June 2017 , and an endline survey in May—July 2018. In addition, a brief monitoring survey was conducted in late 2016 to confirm that CPS construction was proceeding as scheduled. With 91% of CPS constructed before follow-up, Table 4.2 confirms the construction rolled out as planned. Nevertheless, and despite our effort to ensure that the baseline survey preceded CPS construction, a completed CPS building was already available at baseline in 17% of the treatment group villages. It is challenging to conduct an experiment like this with school construction. On one hand, fielding the baseline too early (well before any construction) would have increased the risk, in case of construction delay, that our baseline sampled children would have been too old to attend the newly built preschools.[21] On the other hand, fielding the baseline too late would have resulted in baseline measures that were arguably already affected by the program. We discuss the implications of the slight overlap between baseline survey and construction below.

During the baseline data-collection exercise in 2016, the survey firm sampled up to 26 eligible households per village, using an adapted version of the EPI walk.[22] This method guarantees the baseline sample is representative for households with children of preschool age within the village.[23] Eligible households include at least one child between 24 and 59 months old at baseline. Children were therefore between three and five years old at midline and between four and six years old at endline.

For the baseline data collection, 7,053 eligible households were identified. Extra household sampling was conducted at midline in villages where the number of eligible

[20]The randomization was performed with a list of 310 eligible villages provided by the government. Of these, 60 were assigned to the control group, 123 to T1, 63 to T2, and 64 to T3. Unfortunately, the randomization list contained erroneous village names and five of them were duplicated or could not be identified after the randomization even after substantial effort by MoEYS and the data collection firm. Therefore, the total number of villages decreased to 305. We treated this dropout as random and did not replace the villages. The randomization list also contained villages for which a CPS teacher was no longer available or for which no land could be secured for CPS construction. Therefore, only 91% of treatment group villages received a CPS. Since these factors are potentially endogenous, we do not treat these as random. To maintain ex ante expected balance between control and treatment villages, these villages were therefore not removed from our sample.

[21]As described in Bouguen et al. (2013), construction delays occurred in a previously evaluated program in Cambodia. This considerably reduced take-up, exposure time, and statistical precision.

[22]EPI refers to the Expanded Programme on Immunization of the World Health Organization; see e.g., Henderson and Sundaresan (1982).

[23]The sample is not nationally representative. Villages were selected based on criteria such as expressed interest in the program, poverty rate, lack of a functioning preschool, and teacher availability. Sample households are exclusively from rural areas in southern and eastern Cambodia. This is because western provinces received CPS under a previous preschool construction program.

households at baseline was below 10 (using the same sampling method again) with an additional 53 households added to the sample at midline in this step.

4.3.2 Data

4.3.2.1 Survey and instruments

For each household, a household survey, caregiver survey, and one assessment per eligible child were conducted.[24] At the village level, interviews were conducted with village heads and preschool teachers. In addition, the endline data collection was complemented by a classroom observation survey conducted at all preschools within the sample villages.

The household survey includes information about family structure, household wealth, and other socioeconomic background characteristics. We construct a dwelling quality index and household assets index using Multiple Correspondence analysis (MCA). Each variable was coded using the sign of the coordinate of the first MCA dimension. We then used the standardized version of each coded variable, and averaged them to create a wealth index from which we derive wealth quartiles. We also construct an education background variable which is defined as the maximum years of schooling by any household member or caregiver and derive quartiles from this variable. Belonging to the upper quartile of this variable is equivalent to having a household member with at least 10 years of schooling. Belonging to the lower quartile of this variable implies no household member has more than 4 years of schooling.[25]

The caregiver survey includes questions regarding the child's preschool enrollment and socio-emotional development, as well as 25 questions about parenting practices. The latter measure three key dimensions of parenting: "cognitive parenting", "emotional parenting", "and negative parenting". More cognitive parenting means parents are more likely to engage in activities that contribute to the development of their child's cognitive competencies (e.g. by playing games, reading books, or playing with toys). More socioemotional parenting means parents interact more in ways that provide emotional support and responsiveness (e.g. comforting, encouraging or complimenting the child). More negative parenting means parents are more likely to use harsh or punitive approaches to discipline their child. Last, the survey includes information about home

[24]The caregiver is defined as the direct relative (parent, grandparent, aunt/uncle, or adult sibling) who takes care of the child most of the time. In most cases, the caregiver is a biological parent (60.4% at baseline, 58.7% at midline). In the provinces of Kampong Speu, Kandal, Prey Veng, Svay Rieng and Takeo, the caregiver is often a grandparent of the child. These are provinces with relatively high levels of manufacturing industry; mothers who work in factories only occasionally get to spend time with their child during the work day.

[25]The wealth index and education background measure have a correlation coefficient of 0.43.

visits, D2D activities, and participation in the HBP, as well as the caregiver's perceived returns to education, and a short test of the respondent's nonverbal reasoning ability (based on the Raven's Progressive Matrices Test).

The approximately 45-minute child assessment includes a comprehensive battery of cognitive tests (executive function, language, early numeracy, fine motor development and, at baseline and midline only, gross motor development) as well as anthropometric measures (height and weight). Most of the child tests are based on the Measuring Early Learning Quality and Outcomes (MELQO) toolkit. (See UNESCO (2017) for a description of the measures-of-development process and Raikes et al. (2019) for evidence of validity).[26]

Additional child tests were added to the MELQO items to increase the sensitivity and breadth of the child assessment. The additional tests included the following: the Dimensional Change Card Sort (Zelazo, 2006)[27]; a receptive vocabulary test based on picture recognition derived from the Test de Vocabulario en Imagenes Peabody (TVIP), a version of the Peabody Picture Vocabulary Test adapted for Spanish-speaking populations that was normed to low-income populations in Mexico and Puerto Rico (Dunn et al., 1986; Dunn and Dunn, 2007); a test for knowledge of reading concepts (based on a monitoring tool used by the Cambodian Ministry of Education's Early Child Development Department); and a sustained attention test. Children's socio-emotional development was measured using the caregiver-reported Strengths and Difficulties Questionnaire. Since the overall purpose was to generate a well-functioning test for Cambodia (as opposed to maintaining consistency for other purposes, such as international comparisons), adaptations were prioritized to ensure adequate fit to the local context. While most tests in the early literacy, numeracy, and fine-motor skill domains relate directly to the CPS curriculum, executive function tests are only indirectly related.

Before constructing the composite scores of child test domains, individual tests were first scored and standardized, thus ensuring that all the tests contributed equal variance to the composite score. For almost all the tests, scoring was done by assigning one point for each correct response and summing these points to create an individual score for each test. When a child was unable to complete the practice trial of a test, a score of zero was assigned for this test, as long as the child participated in the other tests. Standardization of each test score was done by subtracting its sample mean and dividing it by its sample standard deviation. All standardized test scores of one domain (e.g., executive function) were then summed into a domain score and standardized again

[26] An in-depth discussion of midline child tests, scoring methods, cultural adaptations, pretesting procedures, and questions about parenting practices can be found in Berkes et al. (2019).

[27] No cultural adaptation of words or pictures was conducted since the test was working well in the field and changes were not deemed necessary by the local staff.

by subtracting its sample mean and dividing by the sample standard deviation of the domain score for better interpretability. The individual tests, their distributions, and the scoring methods are summarized in Appendix 4.B.3.

The village and teacher surveys included questions about the ECED services available in the sample village. This allowed us to monitor implementation of program interventions and alternative services. Parts of the teacher survey and the classroom observation tool administered at endline are based on the Measuring Early Learning Environments (MELE) module of MELQO. MELE includes key domains for quality in early learning environments and the sample items used to measure them. During pretests preceding the endline data collection, constructs to be measured were selected and specific items were adapted to take into account culture-specific views on what defines a high-quality learning environment. We divided the final module into five domains: teacher characteristics, equipment, classroom setting, curriculum content and pedagogy, and teacher-child interactions.

4.3.2.2 Balance at baseline

The main baseline characteristics of villages, households, children, and their caregivers are summarized in Table 4.3.[28] To test for statistically significant differences, we regress each variable on binary indicators for the treatment groups and a set of province dummies to account for stratified randomization. Overall, the sample is balanced in child, household, and caregiver characteristics. The characteristics for T1 and T3 have only minor differences with the control group. However, the characteristics for T2 have larger differences, in particular for variables that one might expect to be associated with child outcomes as well. We address this issue by controlling for these characteristics in our main regressions.

Table 4.3 also highlights some significant baseline differences in preschool enrollment and take-up of demand-side interventions. These differences are due to the aforementioned early roll-out of some preschool interventions. Children in program villages are about 8 pp more likely to have enrolled in preschool at baseline (driven by community preschools which, by definition, were newly constructed). Similarly, the HBP intervention started in some treatment villages before baseline. These baseline imbalances at the individual level are confirmed by imbalance in the availability of preschools and HBP at the village level, shown in the baseline panel of Table 4.6. At baseline, some villages in the treatment group had access to CPS and many treatment groups' villages had access to HBP.

[28]Balance checks for additional variables are presented in Tables 4.15 and 4.16.

There are reasons to believe that these early interventions are unlikely to affect our estimation. As seen, baseline results do not pick up any significant cognitive differences between the treatment groups and the control, indicating that early enrollment in preschool intervention had not given a significant benefit to the treatment group. Relatedly, expressed in number of days at school, early enrollment in preschool programs is only marginally superior in the treatment groups. At baseline, treatment children had approximately only 10 extra days of preschool compared with the control. It further means that the roll-out of our baseline survey was in average 10 days too late in comparison with the roll-out the CPS construction. The case is less favorable for HBP, which has been rolled out before baseline.[29] Since HBP is considered as a demand-side intervention with no expected short-term effects on cognitive performances, early participation of parents in the HBP is unlikely to have any detrimental impacts to our experimental design.

A potentially greater cause for concern is the fact that SPS are reported, at baseline, to be less available in the treatment groups compared to the control group (Table 4.6). While the point estimates are insignificant for T1 and T3, and again insignificant when the treatment groups are considered jointly (- 7 pp), the T2 villages are significantly less likely to have access to a SPS (-11 pp). Since SPS have much better quality than CPS (as we discuss in section 4.3.2.4 below), this small imbalance, which we attribute to chance, is likely to provide a baseline relative cognitive advantage to the control group, therefore driving down our estimates of CPS impacts on cognitive development.[30]

4.3.2.3 Attrition

To minimize attrition associated with seasonal migration related to the agricultural cycle, all survey waves were completed before the beginning of August of each year. Families often move to different villages or are too busy for interviews due to agricultural work during the school breaks in August and September. Attrition and the random assignment are not related for both the midline and endline samples (Column 1 Table 4.17). Moreover, attrition is not significantly related to a set of variables strongly

[29] This is probably due to the fact that setting up an HBP takes much less time than building a CPS.

[30] Another interpretation of the imbalance in the SPS availability is as a SUTVA violation. If, for instance, the experiment caused SPS to be opened in priority in control villages or, inversely, if the presence of a CPS caused an SPS to shut down in the treatment village, this would be a violation of the SUTVA. We do not, however, believe this to be a likely explanation. If the SPS were indeed scaled up in the control group (and/or scaled down in the treatment groups), we would expect to see an increasing number of SPS being open in the control (and/or being shut in the treatment groups) as time goes on. Table 4.6 confirms that the gap in SPS availability is present at baseline but stay stable over time. Moreover, setting up SPS would take time, and the imbalance we observe is at baseline—it would be very unlikely that these changes would have occurred so quickly—i.e. before the deployment of CPS.

associated with child development (Column 2). There is also no evidence of differential attrition with respect to baseline controls except for T2 at the endline follow-up, where children with a high height-for-age z-score were more likely to attrit from the sample. The addition of households to the sample at midline causes differential attrition to become significant for T1 at midline since fewer households were added to this group (Column 3). The overall level of child attrition is around 10% at both midline and endline, and mostly due to seasonal migration.[31] While not unreasonably high compared to similar studies in other countries, and while not likely to differentially affect the various treatment arms of the study (and therefore not affect the internal validity of our analysis), we recognize that the external validity of our results is limited to households that didn't relocate at midline or endline.

4.3.2.4 Quality of CPS and other preschools

We construct measures of preschool quality based on data from two sources, classroom observations and parent reports.

Preschool Quality Measures

We use the teacher survey and classroom observation tools to assess how new CPS compare to SPS and IPS in terms of structural and process quality in our sample.[32] While recruitment and training procedures are similar for teachers of IPS and CPS, only CPS benefit from a standardized building. In contrast, SPS teachers are recruited through a different process and teach in classrooms within the premises of a primary school. Like any evaluation, the impacts we identify are context specific and depend on not only the quality of the intervention itself but also its alternatives. We therefore compare the different preschools in five dimensions of structural quality (teacher characteristics, equipment, classroom setting) and process quality (curriculum and pedagogy, teacher-child interactions) to give the shed light on factors that might be driving the results.

Our results from the preschool quality measures are summarized using aggregate measures for each of the five dimensions of structural and process preschool quality, calculated using the first principal component of individual variables (Table 4.4). CPS

[31] Reasons for attrition of 597 households at midline: child does not currently live in the village (542), caregiver/child temporarily not at home (34), reason unknown (10), household unknown (5), refusal (3), child passed away (3).

[32] Note that this is not representative of IPS and SPS in Cambodia. In particular, it does not represent the sample of IPS after these were initially established, i.e., what was evaluated in Bouguen et al. (2013). Only IPS still operating at the time of endline are represented in this sample. These IPS are positively selected in terms of sustainability and therefore also most likely positively selected in terms of quality.

and IPS teachers are similar according to the teacher characteristics score, while SPS teachers are significantly better according to this score. However, CPS outperforms IPS on the classroom setting and equipment scores. In these two dimensions, CPS are equivalent and maybe even superior (for equipment) to SPS, confirming that structural quality has been significantly improved by the construction program. Our results are strikingly different for the process quality scores. CPS and IPS are not significantly different in terms of curriculum and pedagogy or teacher-child interaction, whereas on both aggregate measures, SPS outperform IPS and CPS.

Unpacking these comparisons into the individual items that make up the aggregate scores show that SPS teachers are more experienced, perform better in the Raven's Progressive Matrices Test and have more training than IPS and CPS teachers (Tables 4.18 and 4.19). They are also better paid and receive their salary more regularly. In terms of equipment, CPS significantly dominate IPS and even in some dimension SPS, except for access to electricity, water source and toilet facilities. Compared with IPS, CPS classrooms have less breaks, are more likely to follow a curriculum, have more attendance, more enrollment, have more storybook activities (Table 4.19). Yet, the quality of the classroom setting, pedagogy and teacher-child interactions remain far from the quality observed in SPS. Crucially, the teacher-interaction individual variables are very similar in CPS and IPS, while SPS engage more with the children and provide significantly more encouragements.

Overall, these results confirm that CPS are of better quality than IPS but maybe not on the dimensions that are the most important for children's performance. The upgrading of IPS to CPS mostly affected the learning environment by providing better equipment, while teacher quality was only affected marginally in the measured dimensions. In contrast, SPS teachers are significantly more educated and perform better in terms of instructional quality. In other words, while the switch from IPS to CPS is likely to affect children positively, a switch from SPS to CPS could affect them negatively.

Parental assessments

Last, we use questions from the caregiver survey to compare differences in missing days, perceived teacher quality, financial contributions, and travel time between the types of preschools (Table 4.5). Ideally, we would like to compare parents' perceived quality differences between preschools to validate whether observed quality differences are also known to parents. However, data on perceived quality are limited: they are only available for enrolled children and about the type of preschool the child attends and hence, the information provided in Table 4.5 should be interpreted with care.

At both midline and endline that CPS classes get cancelled significantly more often than SPS classes. Though differences tend to be relatively small, parents of children enrolled at CPS assess the kindness of CPS teachers as higher than that of their SPS counterparts. This could be explained by the greater likelihood that CPS teachers are from the same community as the parents than are SPS teachers. Yet at endline, parents of children enrolled at SPS assess the reliability of SPS teachers as higher than that of CPS teachers. While financial contributions for teacher salary, construction, and renovation tend to be almost zero for all types of schools, contributions for school materials are substantial.[33] On average, parents of children enrolled at CPS have paid \$35 in the current school year at midline and \$47 at endline; contributions at SPS were \$49 at midline and \$68 at endline.[34] CPS seem to be the cheapest option for parents: enrollment in CPS is slightly more expensive than IPS in terms of teacher salary contribution (+0.8 USD) but is significantly cheaper in terms of school material contribution (-6.6USD). Notably, while SPS tend to be of higher quality than CPS for most dimensions of our classroom observation survey, sending a child to an SPS is likely to be associated with higher costs than sending her to a CPS or IPS.

4.4 Empirical Framework

By virtue of randomization, we can straightforwardly estimate intent-to-treat impacts (ITT) of the intervention. Specifically, we estimate ITT effects by first pooling all treatment arms and then allowing for different impacts for each of the different treatment groups by using the specification:

$$Y_{ip} = \alpha + \beta_g Z^g_{ip} + \mathbf{X}'_{ip}\delta + \mu_p + u_{ip}, \tag{4.1}$$

where Y_{ip} is an outcome (e.g., test score) measured post-treatment for child i in province p. Z^g_{ip} is a binary variable for any treatment (in the pooled treatment specification) or a vector with three binary indicators for treatment groups $g \in \{T1, T2, T3\}$ (in the disaggregated specification), with the control group as the reference category.

[33] The exact question asked is "How much money have you spent on school material (paper, board, chalk, cloth, water, food) for your child since the beginning of the school year?" Hence, it does not distinguish between money for food that is used for "pedagogical" reasons and money that is used to buy snacks for consumption. While we do not have data on this, we cannot rule out that teachers sell food as an indirect way of increasing their salaries.

[34] At most CPS and IPS, the school year starts in late October or early November. About half of CPS and IPS end their school year at the end of July, while the other half end the school year at the end of August. Most SPS have their school year from late October or early November until the end of August. Since the endline data collection was conducted about one month later than the midline data collection, some caution is warranted with comparisons between the waves.

$\mathbf{X}_{ip}$ is a vector of control variables which, in the basic specification, includes the baseline test score, child age and child age squared. To improve statistical power and to control for potential imbalances, we also estimate a model with additional control variables including child gender, baseline height-for-age z-score, household size, household wealth quintile dummy variables, and caregiver baseline age, gender, score on the Raven's Progressive Matrices Test assessment, and parenting scores. We note that the precision of the estimates improves only slightly using these additional control variables, and point estimates remain very stable. The specification includes μ_p, which is a set of province fixed effects to account for stratified randomization at the province level (Bruhn and McKenzie, 2009). u_{ip} is an individual error term where within-village correlation will be taken into account for estimation of the variance. The ITT estimates of the impact of each treatment on outcomes are given by the β_gs.

In our context, the ITT effects should be interpreted with the understanding that they stem from children who would not have benefited from any preschool absent the program, but also from children switching from an existing preschool arrangement to the CPS. These specific substitution patterns tend to drive down the magnitude of the ITT estimates and may mask the potentially larger impacts that CPS may have on just the children who would have stayed home absent the CPS. Modeling the counterfactuals explicitly, Berkes and Bouguen (2019) shows that this is indeed the case. Nevertheless, the ITT estimate is empirically valid and policy relevant, as it reflects the program's impacts in a typical Cambodian preschool environment.

4.5 Results

4.5.1 CPS construction

Before discussing the impact of the intervention on outcomes, we review the CPS construction status and overall availability of preschools at the time of the baseine, midline, and endline surveys (Table 4.6). As discussed above, there are imbalances in SPS availability at baseline that remain observable at midline and endline. We attribute these imbalances to chance and note that, if anything, such imbalance should drive our estimates of impacts on child development downward. In addition, since the imbalance is larger in T2, the downward bias should also be larger in T2, a hypothesis that we return to in Section 4.5.2.

Importantly for interpreting our results, 81% of control group villages have some sort of preschool at midline; by endline, this is almost 85%. Nevertheless, the number is significantly higher in the treatment group villages, where availability of any preschool is already at 93%, 100%, and 98% for T1, T2, and T3, respectively, by midline. The time patterns reflect how the construction of new CPS has affected school infrastructure in the villages. Most formal CPS have replaced preexisting IPS arrangements. Some IPS nevertheless remain open in treatment group villages, either due to failure to implement a formal CPS or because they are run independently by a local pagoda or an NGO. Notably, adherence to the CPS construction allocation was high: there were no formal CPS in control villages at midline and only one at endline. Overall, CPS constructions and simultaneous IPS closures have improved the classroom setting and equipment quality of available preschools. Column 7 in Table 4.4 shows that for these dimensions, the treatment group preschools are significantly better than preschools in the control group. As expected, due to the small differences in process quality and teacher characteristics between CPS and IPS, no treatment effects are observed for these dimensions.

4.5.2 Enrollment outcomes

We first analyze child enrollment into preschool—a key government objective for the program. As mentioned above, follow-up data collection took place from April—June 2017 (midline) and May—July 2018 (endline); the school year for all types of preschools begins in late October or early November. Table 4.7 summarizes the impact on enrollment into any type of preschool, into each type of preschool, and into primary school, by treatment status. The measure of enrollment in this table is the child's status on the day of the midline and endline survey visits. At midline, the interventions increased enrollment into any type of preschool by 10.7 percentage points (10.2, 9.4, and 12.8 in T1, T2, and T3, respectively), compared to a control group enrollment rate of 39.6%. In other words, the intervention increased preschool enrollment by over 25%.[35]

The overall increase in preschool enrollment (observed on the day of the visit) is entirely driven by an increase in CPS enrollment of 41% at midline and 32% at endline (Column 2 of Table 4.7—the decline at endline is caused by the transition of six year-olds to primary school.[36] Differences in enrollment across treatment types are not statistically significant, with the exception of T2 villages, which have slightly higher CPS enrollment than T1 and T3 at endline (this may simply reflect the lower baseline availability of SPS in T2).

The increase in CPS enrollment is fueled in large part by a large substitution from IPS to CPS (IPS enrollments were 25 pp lower at midline and 20 pp lower at endline)

[35] To interpret the results in a straightforward way, we do not include any control variables in this regression. Including them does not alter the results.

[36] At endline, enrollment into formal CPS is not exactly 0 in the control group. A small number of children in a control group village in the province Kratie reported attending a formal CPS in an adjacent T3 village.

which was by design as the new CPS replaced the IPS (Table 4.7). There was also a more moderate reduction in SPS enrollment (SPS enrollments were almost 6 pp lower at midline and 5 pp lower at endline). This reduction could be driven by two mechanisms: first, parents actually switching as a response to the program or second, the slight baseline imbalance in the availability of SPS (Table 4.6). Although we cannot empirically distinguish between these mechanisms, the consistency in the gaps in baseline SPS availability and enrollment (Table 4.6 and Table 4.7) suggest that it is the imbalance that explains the result. In addition, the fact that SPS have higher quality than CPS makes it less likely that parents would switch. Both mechanisms would have the effect of driving down the ITT estimates of the impact of the program on child development.[37]

In addition to enrollment on the day of the survey visit, we analyze two other indicators of enrollment: ever enrolled in preschool and duration of enrollment, both measured by the time of the endline survey (Table 4.8). The impacts on ever enrolled are consistent with those for the day of the visit, although the percentage increase in enrollment is smaller (a roughly 11 percentage point increase over a control group mean of about 59%, so an increase of about 19%). The cross-preschool-type patterns suggest that much of this increase comes from a large reduction in the probability that children are ever enrolled in an IPS and a more modest reduction in SPS enrollment.

These results suggest that about 50% of the eligible children did not attend preschool at any point in time even when a preschool was available in the village. When asked why they did not send their child to preschool, parents indicate two main reasons. First, many parents expressed difficulties bringing and picking up the child at preschool (51%). Parents also often declared being afraid to let the kids go to school by themselves (62%) or that the school was too far away (35%). Since CPS sessions were only given in the morning, those parents who had to work outside of the village (in fields far from the village or in a garment factory) and/or were living too far from the preschool could not send their child to school. Second, many parents who did not enroll their child to preschool declare that the kid was too "afraid", not "ready" ("does not speak enough") or not "mature" enough to go to school. Only a handful of parents declared that that the preschool turned them down (8%) or that they could not afford to go school (0.78%). Only 1.7% of the parents declared that the school was too crowded or at capacity. These results suggest that supply-side limitations have little bearing in

[37] An earlier evaluation of GPE I (Bouguen et al., 2018a) showed that enrollment in newly established SPS was in large part driven by substitution from underage enrollment in primary school. The results here do not suggest that there is substitution between CPS and primary school. This is despite the fact that at endline, about a third of the children were at least six years old and eligible for primary school.

explaining low enrollment. Similarly, very few parents invoked information constraints: only 3% declare that they "did not think about it". Information does not seem to be the most likely constraints for not sending the child to school either. Finally, in results not presented here, we found that T2 and T3 had no effect on the reasons given by parents for not sending their child to school.[38] This may in part explain why T2 and T3 program did not significantly improve demand for schooling in our context. We will come back to this issue in sub-section below when covering the impact of demand-side interventions 4.5.5.

The intervention also significantly affected the months enrolled at preschool. By endline, the intervention had increased the average months enrolled at (any) preschool by 0.8 to 1.2 months from a counterfactual of 5.4 months in the control group (Table 4.8). This is driven by an increase in the duration of CPS enrollment of 4.1 to 5.1 months (alongside a decrease in IPS and SPS).[39]

Finally, we look at heterogeneous impacts on total months enrolled by estimating the model for different groups of the population (Table 4.9). When looking at months enrolled at any preschool, only the middle wealth quartiles are significantly affected, with wealth quartile 4 showing similar increases in months enrolled at CPS but stronger substitution with SPS. A similar pattern appears when using the baseline education background of the households. Wealth quartile 4 also shows an increase in primary school enrollment which is significant at the 10% level. This is particularly interesting since this effect is not driven by the oldest children: it suggests that there might be positive complementarities between parental involvement and CPS enrollment which in turn leads to early enrollment into primary school. We look below at the consequences of these different enrollment behaviors on child development outcomes.

4.5.3 Child development outcomes

In the short term (by midline), the program improved child cognitive development by 0.04 SD on average (Table 4.10). Effects are slightly larger for T1 (0.048 SD) and T3 (0.045 SD), and small and insignificant for T2 (0.009 SD). While generally statistically significant, these impacts are nevertheless small in magnitude. The program also significantly reduced the occurrence of socio-emotional problems in all three treatment groups at midline (0.09 SD). But at endline, all these impacts are both smaller in magnitude and not statistically different from zero. Point estimates for the cog-

[38]Since the treatments affected enrollment, these analyses are not causal but simply indicative of possible effects.

[39]We also show how enrollment patterns depend on child age in Figures 4.5 and 4.6. In all three treatment groups, the probability of CPS enrollment is increasing with child age until children begin to enroll into primary school at the beginning of age six.

nitive development index are approximately reduced by half in T1 and T3 and have completely vanished for the socio-emotional problems measure.

We next estimate effects for the individual cognitive domains (executive function, language, and early numeracy) as well as those for gross- and fine-motor development (Table 4.11). Impacts are most pronounced for early numeracy, followed by language development; they are insignificant for executive function. In the pooled model, only the impacts on early numeracy are statistically significantly different from zero. At endline, language development was significantly positively affected (T1) while fine motor development and executive function were negatively affected (T2). In the pooled model, none of the impacts is statistically significant, and all are close to zero.

To further investigate which set of skills drives the impacts, we look at effects on the individual test-item scores (Table 4.20). While estimates vary between treatment groups and waves, impacts on verbal counting skills and number and letter knowledge are above 0.03 SD for the joint treatment group (any T) estimate for both follow-ups. The finding is striking since these tests are directly linked to the curriculum. Similarly, the initial letter identification test is curriculum based and is positively affected (+0.05 SD) at the endline. Yet, other curriculum-based tests, such as the name writing, reading words, and fine motor tests do not show these consistent impacts.

We estimate the model on population subgroups to investigate heterogeneous impacts (Table 4.12). At midline, the results suggest that it is children who were age 4 at baseline who are driving the findings—consistent with the fact that it is children of this age who would have been affected by CPS construction. Furthermore, midline effects on child development are primarily driven by non-stunted children, suggesting that the CPS were unable to foster development of the most-disadvantaged children and therefore contributed to increasing socioeconomic gaps in cognitive development. This finding is reinforced by the impacts on cognitive development by wealth and education background quartiles, which show that only children from the wealthiest or more educated households benefit from CPS—with the wealthiest children having larger impacts at endline than at midline.[40] Our findings therefore suggest that only these better off children were able to fully take advantage of the CPS.

The heterogeneous impacts cannot be explained by differences in preschool enrollment since children of the wealthiest or most educated quartile did not enroll more in preschool, nor did they stay there longer (Table 4.9). We see three alternative mechanisms to account for the effect on the children from the wealthier and most educated households. First, the effect could be explained by the slight increase in the *primary*

[40]Similar to the discussion of Table 4.11, the impact on cognitive development of wealth and education background quartile 4 is driven by the subdomains language development and early numeracy. These results are available from the authors upon request.

school enrollment among the wealthier households (Column 5 of Table 4.9). At the same time, however, there is no differential effect on primary school enrollment for children from the most educated families, suggesting that primary school enrollment effects are not the sole driver of the heterogenous child development impacts. Second, it is likely that wealthier and more educated parents provide a favorable environment at home that complements the education received at school. In results not shown here, we could not identify any treatment heterogeneity according to parenting patterns at baseline. This suggests that it is the environment provided by wealthier or most educated parents rather than initial parenting *per se* that drives the heterogeneous impacts. Third, wealthier or more educated parents may be more likely to increase their parental investment when their children are enrolled in preschool. We test for this potential channel by looking at the impact of the CPS program on parenting patterns at midline and endline, and find that the program positively affected cognitive parenting at midline. (Table 4.21). Moreover, this effect is even stronger on the wealthier or most educated parents (Table 4.22).[41] Nevertheless, these average and heterogeneous effects on parental impacts disappear at endline, suggesting that they are unlikely to be the only mechanism since the effect on their children's development are sustained at endline.[42]

Overall, our findings suggest that only relatively privileged children—children with wealthier or more educated parents—who were able to fully take advantage of the CPS. These parents are likely to offer an environment that complements the preschool investment. They may also be able to send their children earlier to primary school or to select a school that best fits the need of their children.

4.5.4 Demand-side interventions

Self-reported exposure to the demand-side interventions is likely to be less reliable than for preschool enrollment since these were often just a one-time interaction many months in the past. Nevertheless, we analyze the impacts on indicators for these "take-up" measures in order to explore whether they may partly drive the lack of differential impacts across T1, T2, and T3 (Table 4.13).[43] We interpret ever having received a home visit to discuss child development, or ever having received an information

[41] A similar pattern is observed by Padilla (2019) who documents effects of the Head Start program in the US on parenting outcomes. She finds positive effects on parents' cognitively stimulating behaviors and no effects on parent's use of harsh punishment. The effects on cognitively stimulating behaviors are driven by parents with higher cognitive stimulation prior the intervention.

[42] Table 4.21 further documents that there is no substitution between parental involvement and preschool availability.

[43] We refer to these as "take up" recognizing that they are not driven purely by the households' choices; they alse depend on implementation as well as recall.

leaflet, as indicators of D2D activities (T2 and T3). No treatment groups report any additional home visits as a result of the program. It is possible that home visits from the village head might be difficult to capture (and for respondents to recall) due to their informal nature. This interpretation is also consistent with the fact that almost 70% of respondents in control villages report having received a home visit, so it is possible this measure is too blunt.

On the other hand, respondents in T2 and T3 were more likely to report having received an information leaflet whereas those in T1 were not (as one would expect given the content of the different treatments). This impact is apparent at both midline (+ 6 pp for T2 and +8 pp for T3) and endline (+11 pp for T2 and +8 pp for T3). While the result is somewhat reassuring—coverage is significantly different across treatment arms—the overall percentage of respondents who report receiving a leaflet is small in all groups. The highest take-up by endline is reported in T2 villages, where it only reaches 18%. Even allowing for poor respondent memory, and despite the field team's effort to ensure that 100% of sample households in T2 and T3 received a leaflet, this value is low and suggests the intensity of the intervention may have been weak.

The patterns in the likelihood of having participated in an HBP (either "ever" or "more than once") are likewise somewhat reassuring. The impact of having been assigned to T3 significantly increases these likelihoods (+19 pp at midline and +10 pp at endline). But the impact of being assigned to T1 is also statistically significant (although about half the magnitude). This is surprising since these villages had no formal HBP deployed as a part of the intervention. At the same time, and consistent with the discussion above, the overall take-up rates implied by these numbers are small. The highest participation rate reported for ever having participated in an HBP by endline is just below 30% (T3), again suggesting the intensity of the intervention was weak.

Despite these (statistically albeit small) significant differences in the intensity of the demand-side programs across groups, we find that they did not significantly increase enrollment, they did not modify the performance of the CPS, and they did not affect parenting behaviours. We cannot rule out the possibility that a more intensive set of programs (e.g., more frequent home visits, more intensive HBP) might have had bigger impacts. But it is also possible that, in the context of small rural villages, the construction of the CPS itself had a demand-side "information" component in the sense that it was likely an important village event that households would have known about. In such a context, it could have been the case that the additional demand-side programs did not convey any information over and above the construction itself.

4.5.5 Perceived return to schooling

The demand-side interventions in this program included a component designed to affect caregiver perceptions about the value of preschool and of education overall. We summarize these effects through two main sets of outcome measures: reported optimal age for starting preschool and reported return to education (Table 4.14). The results are consistent with the intervention lowering the age that caregivers think is optimal for starting preschool. A significantly higher share of caregivers reported that the optimal age to be enrolled at preschool for the first time is three or four years old. But there is no significant difference across the different treatment groups, suggesting it is likely the CPS construction itself (and any activity or information dissemination surrounding it) that is driving the result—not the specific demand-side interventions.

The magnitude of the impact is somewhat larger at midline than at endline, further suggesting that caregivers might change their views on this question as their children grow older or that there is fade-out in the impact of these perception impacts. Initial excitement associated with the CPS construction might fade or caregivers might lower their perceived value of preschool based on their experience with the new CPS.

The distribution of leaflets had virtually no impact on the perceived wage increment associated with schooling, with the exception of a small impact for T2 at midline (last two columns of Table 4.14). Reported average values for the annual increments associated with schooling are 6.8% for primary school and 8.5% for secondary school (at endline, virtually the same as at midline). These numbers are remarkably similar to those reported by Humphreys (2015) based on Cambodia's Socio-Economic Survey (CSES) for 2010. In that analysis, the return to a year of primary school was estimated to be 5.8% for men and 3.5% for women, and the corresponding rates for a year of schooling at secondary school were estimated to be 14.2% and 14.6%.[44] The implicit values included on the flyer (based on our analysis of CSES 2009) were 8% and 12%.[45] One hypothesis for the lack of impact of the demand-side interventions, therefore, is that people had a reasonably accurate perception of the average return to schooling, and new information did not affect their priors.

[44]The secondary school values reported here are averages over those in Humphreys (2015), which are estimated separately for lower and upper secondary.

[45]These are implicit because they are based on cell means reported for people who have completed those years of schooling. Note that the rate for primary school is calculated as the difference between "completed" and "not completed" primary school, which we modelled as six versus three years of schooling in the calculation of the "return."

4.6 Conclusion

Consistent with findings in other studies, our results show that it is hard to implement large-scale preschool programs that consistently result in improved child development outcomes. While our results document an increase in net new enrollments in preschool, they do not document large average impacts on short- (one-year) or medium-term (two-year) impacts on child cognitive or socio-emotional outcomes. The findings highlight that, notwithstanding the robust evidence that high-quality preschool experiences boost child development, it is difficult to engender those experiences in a program at scale in a low-income, capacity-constrained environment.

Our findings have both programmatic and research implications. From a programmatic point of view, they suggest prioritizing a review of the quality of preschools—in this case the CPS model—and ways to enhance that quality. While our results are context-specific, they nevertheless highlight the general importance of counterfactual childcare arrangements. These arrangements are common in many countries and include family care along with informal preschool arrangements. In order to substantively impact child development outcomes, formal preschool provision needs to deliver substantial value added over this counterfactual. While the intervention we study succeeded in providing better infrastructure and materials relative to informal arrangements, it did not provide a substantively better teaching and learning experience. Therefore the results suggest that in Cambodia, and likely in other similar contexts, a key focus should be on improving the process quality aspects of preschool, namely improvements in teacher pedagogical skills and ultimately teacher-child interactions. This might require better training of teachers (in Cambodia, training for CPS teachers is currently substantially shorter and less intensive than that for SPS teachers), but perhaps also approaches that build on intrinsic and potentially extrinsic motivation of teachers to ensure that they are supported in, and recognized for, putting better teaching methods into practice.

From a research point of view, our findings suggest that more work is required to understand the drivers of preschool demand and quality. The demand-side approaches tested in this evaluation were not enough to mobilize much additional enrollment over and above simply building preschools, suggesting that other factors inform the decision to send children to preschool. Direct or indirect costs may play a role, so approaches to reduce those costs—for example, cash transfers or even further reductions in travel distances—might be necessary to induce higher participation rates. In addition, if the quantity and quality of preschool services do not meet families' needs, households might have low demand for them. Therefore, increasing the quantity (time spent in preschool per day) or the quality of preschools may increase demand. Moreover, the fact that children from wealthier families did benefit from the newly constructed preschools suggests that other demand-side factors (such as complementary investments in child development) act as additional constraints. Further investigation of these factors is required to better understand and address them.

Appendix

4.A Tables

Table 4.1: Random treatment allocation to 305 villages

Group	CPS	D2D	HBP	Villages
T1	☒	☐	☐	120
T2	☒	☒	☐	64
T3	☒	☒	☒	63
Control	☐	☐	☐	58

Treatment design based on random allocation of 305 villages to three treatment groups and a control group.

Table 4.2: Timetable

Period	**Activity**	**CPS construction**
03/2016	Begin CPS construction	0% completed
05/2016 – 07/2016	Baseline data collection	17% completed
12/2016	Monitoring survey (by phone)	82% completed
04/2017 – 06/2017	Midline data collection	91% completed
05/2018 – 07/2018	Endline data collection	91% completed

Percentages refer to share of villages in the three treatment groups for which construction of a new CPS was reported as completed on the day of data collection.

Table 4.3: Balance in baseline covariates

	N (1)	C (2)	T1-C (3)	T2-C (4)	T3-C (5)	Any T-C (6)
Child characteristics						
Cognitive development index	7491	0.030	-0.001	0.006	-0.006	-0.000
		[0.977]	(0.038)	(0.045)	(0.042)	(0.036)
Socio-emotional problems	7611	-0.003	0.027	0.042	0.002	0.024
		[0.996]	(0.043)	(0.048)	(0.045)	(0.040)
Age (yrs) w. decimals	7632	3.476	0.008	0.020	-0.039	-0.001
		[0.824]	(0.029)	(0.032)	(0.031)	(0.027)
Female	7632	0.506	-0.020	-0.030*	-0.004	-0.019
		[0.500]	(0.017)	(0.018)	(0.017)	(0.015)
Stunted (lhfa<2sd)	7473	0.341	0.018	0.020	0.032	0.022
		[0.474]	(0.018)	(0.022)	(0.022)	(0.017)
Household and caregiver characteristics						
Household size	7632	5.966	-0.030	0.064	-0.004	0.001
		[2.110]	(0.103)	(0.133)	(0.113)	(0.097)
Wealth score	7631	-0.047	-0.001	0.036	0.089	0.032
		[0.971]	(0.066)	(0.076)	(0.067)	(0.059)
Caregiver female	7629	0.890	0.024**	0.029**	0.017	0.023**
		[0.313]	(0.012)	(0.013)	(0.014)	(0.011)
Caregiver age	7629	40.777	0.415	-1.188	0.173	-0.054
		[14.688]	(0.644)	(0.742)	(0.680)	(0.584)
Caregiver years of education	7621	4.681	-0.286	-0.264	0.292	-0.127
		[11.120]	(0.428)	(0.450)	(0.495)	(0.392)
Cognitive parenting	7611	-0.006	0.018	-0.012	0.073*	0.025
		[1.002]	(0.040)	(0.049)	(0.044)	(0.037)
Socioemotional parenting	7612	-0.008	-0.021	-0.112**	-0.022	-0.044
		[1.003]	(0.038)	(0.048)	(0.044)	(0.035)
Negative parenting	7613	0.002	0.052	0.092*	0.013	0.052
		[0.999]	(0.043)	(0.047)	(0.044)	(0.038)
Baseline program attendance						
Attending preschool	7612	0.153	0.073***	0.063**	0.108***	0.080***
		[0.360]	(0.023)	(0.027)	(0.026)	(0.021)
Careg. participated in a HBP	7609	0.104	0.031*	0.055***	0.184***	0.078***
		[0.306]	(0.017)	(0.020)	(0.023)	(0.017)
Caregiver received a D2D	7609	0.017	0.003	0.012*	0.021***	0.010*
		[0.130]	(0.006)	(0.007)	(0.008)	(0.006)

Differences in means are based on OLS regressions at the child level on province fixed effects and three treatment group dummies (columns 3-5) or a joint treatment group dummy (column 6). Robust standard errors clustered at village level in parentheses. Control group standard deviations in square brackets. Additional balance checks are presented in Tables 4.15 and 4.16. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.4: Structural and process quality by type of preschool and treatment status

	CPS (1)	IPS (2)	SPS (3)	IPS-CPS (4)	SPS-CPS (5)	C (6)	Any T - C (7)	N (8)
Structural quality scores								
Teacher characteristics	-0.37	-0.22	1.41	0.148	1.781***	0.057	-0.067	328
				(0.114)	(0.100)		(0.161)	
Classroom setting	0.08	-0.51	0.22	-0.592***	0.138	-0.494	0.585***	326
				(0.148)	(0.136)		(0.128)	
Equipment	0.35	-1.07	-0.18	-1.419***	-0.539***	-0.940	1.110***	327
				(0.157)	(0.125)		(0.167)	
Process quality scores								
Curriculum and pedagogy	-0.05	-0.23	0.38	-0.179	0.428***	-0.085	0.100	329
				(0.141)	(0.152)		(0.165)	
Teacher-child interactions	-0.06	-0.10	0.29	-0.044	0.347**	-0.028	0.034	321
				(0.139)	(0.136)		(0.151)	

Columns 1—3 show averages by type of preschool. Column 6 shows average for all schools in control group. Columns 4-5 show differences between types of preschools. Column 7 shows difference between treatments groups and the control group. Differences are based on a regression of summary scores on binary preschool type variables (columns 4-5) or a binary treatment group variable (column 7) using robust standard errors. Summary scores are prepared using the first principal component of the individual variables and are standardized to a mean of zero and a standard deviation of one. Observations with missing values are dropped from the summary scores. For details about individual variables of summary scores see Table 4.18 and 4.19. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.5: Perceived preschool quality and financial contributions of parents

	CPS (1)	IPS (2)	SPS (3)	IPS-CPS (4)	SPS-CPS (5)	N (6)
Midline						
Regular school days without class taking place (last 30 days)	2.48	1.51	0.92	-0.969***	-1.556***	3378
				(0.321)	(0.305)	
Days missed due to personal reasons (last 30 days)	4.32	3.51	3.42	-0.813**	-0.902***	3378
				(0.332)	(0.332)	
Perceived kindness of teacher (1-10)	8.74	8.25	8.57	-0.482***	-0.164*	3379
				(0.130)	(0.090)	
Perceived professional knowledge of teacher (1-10)	8.62	8.09	8.56	-0.523***	-0.060	3379
				(0.127)	(0.093)	
Perceived reliability of teacher (1-10)	8.38	8.02	8.43	-0.364**	0.049	3379
				(0.141)	(0.122)	
Contribution to teacher salary in current school year (USD)	0.14	0.42	0.16	0.281**	0.022	3360
				(0.139)	(0.069)	
Contribution to school material in current school year (USD)	35.15	31.70	49.22	-3.440	14.080***	3354
				(2.641)	(2.748)	
Contribution to construction/renovation since child attends (USD)	0.67	0.67	0.64	-0.000	-0.029	3345
				(0.215)	(0.135)	
Endline						
Regular school days without class taking place (last 30 days)	2.13	1.78	1.31	-0.354	-0.818***	3193
				(0.334)	(0.216)	
Days missed due to personal reasons (last 30 days)	3.47	2.82	3.15	-0.649**	-0.313	3050
				(0.280)	(0.309)	
Perceived kindness of teacher (1-10)	8.87	8.71	8.69	-0.161*	-0.183**	3193
				(0.093)	(0.074)	
Perceived professional knowledge of teacher (1-10)	8.72	8.57	8.80	-0.153*	0.073	3193
				(0.083)	(0.070)	
Perceived reliability of teacher (1-10)	8.65	8.64	8.84	-0.014	0.190**	3193
				(0.116)	(0.078)	
Contribution to teacher salary in current school year (USD)	0.18	0.96	0.12	0.776***	-0.058	3158
				(0.251)	(0.066)	
Contribution to school material in current school year (USD)	46.74	53.32	68.29	6.578**	21.549***	3165
				(3.155)	(2.468)	
Contribution to construction/renovation since child attends (USD)	0.74	0.77	0.87	0.027	0.129	3172
				(0.185)	(0.161)	

Contributions are trimmed at the 99th percentile to control for outliers. Columns 1—3 show averages by type of preschool. Columns 4—5 show differences between types of preschools. Robust standard errors clustered at village level in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.6: Preschool infrastructure

Village has...	Obs. (1)	*C* (2)	T1-C (3)	T2-C (4)	T3-C (5)	Any T-C (6)
Baseline						
...any preschool	305	*0.759*	0.008	0.003	0.085	0.027
			(0.069)	(0.078)	(0.073)	(0.062)
...CPS	305	*0.000*	0.167***	0.143***	0.203***	0.170***
			(0.034)	(0.044)	(0.051)	(0.024)
...IPS	305	*0.603*	-0.103	-0.032	-0.010	-0.061
			(0.079)	(0.090)	(0.089)	(0.072)
...SPS	305	*0.224*	-0.049	-0.113*	-0.052	-0.066
			(0.065)	(0.068)	(0.073)	(0.060)
...HBP	305	*0.414*	0.053	0.189**	0.524***	0.210***
			(0.080)	(0.090)	(0.072)	(0.072)
Midline						
...any preschool	305	*0.810*	0.123**	0.190***	0.174***	0.153***
			(0.057)	(0.052)	(0.054)	(0.053)
...CPS	305	*0.000*	0.858***	0.984***	0.938***	0.911***
			(0.032)	(0.016)	(0.030)	(0.018)
...IPS	305	*0.655*	-0.564***	-0.623***	-0.546***	-0.574***
			(0.068)	(0.067)	(0.074)	(0.065)
...SPS	305	*0.241*	-0.058	-0.130*	-0.070	-0.079
			(0.067)	(0.069)	(0.074)	(0.061)
...HBP	305	*0.362*	0.105	0.162*	0.591***	0.245***
			(0.078)	(0.090)	(0.069)	(0.071)
Endline						
...any preschool	305	*0.845*	0.097*	0.155***	0.140***	0.123**
			(0.052)	(0.048)	(0.050)	(0.049)
...CPS	305	*0.017*	0.841***	0.983***	0.905***	0.894***
			(0.036)	(0.017)	(0.038)	(0.025)
...IPS	305	*0.724*	-0.607***	-0.677***	-0.599***	-0.623***
			(0.066)	(0.065)	(0.072)	(0.062)
...SPS	305	*0.259*	-0.050	-0.132*	-0.040	-0.068
			(0.069)	(0.072)	(0.078)	(0.063)
...HBP	305	*0.259*	0.191***	0.281***	0.632***	0.328***
			(0.074)	(0.086)	(0.070)	(0.066)

Table shows preschools available in sample villages. C is the control group mean. Columns 3-6 show the difference between treatment group and control group mean based on a regression of preschool availability on a set of dummy variables for each treatment group (columns 3-5) or on a dummy variable for all treatment groups (column 6). No control variables included in regression model. Estimates correct for heteroskedasticity. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.7: Enrollment on day of survey by type of school

	Any preschool (1)	CPS (2)	IPS (3)	SPS (4)	Primary (5)
Midline (age 3-5)					
Group: T1	0.102***	0.389***	-0.247***	-0.041*	0.004
	(0.038)	(0.025)	(0.034)	(0.024)	(0.009)
Group: T2	0.094**	0.432***	-0.263***	-0.075***	0.001
	(0.043)	(0.027)	(0.033)	(0.023)	(0.010)
Group: T3	0.128***	0.426***	-0.232***	-0.066***	-0.010
	(0.041)	(0.030)	(0.037)	(0.023)	(0.009)
Group: Any T	0.107***	0.410***	-0.247***	-0.056**	-0.000
	(0.034)	(0.016)	(0.033)	(0.022)	(0.008)
Control group mean	*0.396*	*-0.000*	*0.284*	*0.112*	*0.040*
p-value: T1=T2	0.832	0.241	0.0879	0.0206	0.699
p-value: T1=T3	0.427	0.340	0.425	0.0991	0.112
p-value: T2=T3	0.377	0.884	0.0736	0.475	0.288
Observations	6992	6992	6992	6992	6992
Endline (age 4-6)					
Group: T1	0.062*	0.297***	-0.194***	-0.041	0.004
	(0.034)	(0.021)	(0.031)	(0.027)	(0.024)
Group: T2	0.079**	0.373***	-0.228***	-0.066**	-0.024
	(0.039)	(0.027)	(0.031)	(0.027)	(0.025)
Group: T3	0.057	0.307***	-0.191***	-0.058**	0.002
	(0.037)	(0.028)	(0.034)	(0.027)	(0.026)
Group: Any T	0.065**	0.319***	-0.202***	-0.052**	-0.004
	(0.031)	(0.015)	(0.030)	(0.025)	(0.022)
Control group mean	*0.402*	*0.009*	*0.258*	*0.135*	*0.257*
p-value: T1=T2	0.598	0.0226	0.0153	0.201	0.154
p-value: T1=T3	0.864	0.783	0.888	0.364	0.952
p-value: T2=T3	0.533	0.0796	0.0526	0.712	0.223
Observations	7015	7015	7015	7015	7015

No additional control variables included in regression model of outcome variable on treatment group dummy variables. Estimates correct for heteroskedasticity and within-village correlations. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.8: Total enrollment at endline by type of school

	Any preschool	CPS	IPS	SPS	Primary
	(1)	(2)	(3)	(4)	(5)
Ever enrolled by endline (age 4–6)					
Group: T1	0.094**	0.478***	-0.301***	-0.057	0.004
	(0.039)	(0.028)	(0.040)	(0.036)	(0.024)
Group: T2	0.112***	0.595***	-0.357***	-0.106***	-0.023
	(0.043)	(0.029)	(0.038)	(0.035)	(0.025)
Group: T3	0.126***	0.528***	-0.286***	-0.088**	0.002
	(0.041)	(0.036)	(0.044)	(0.035)	(0.026)
Group: Any T	0.107***	0.521***	-0.311***	-0.078**	-0.003
	(0.036)	(0.019)	(0.038)	(0.032)	(0.022)
Control group mean	*0.591*	*0.017*	*0.413*	*0.200*	*0.262*
p-value: T1=T2	0.574	0.00324	0.00118	0.0490	0.160
p-value: T1=T3	0.296	0.266	0.574	0.230	0.896
p-value: T2=T3	0.700	0.141	0.00450	0.457	0.253
Observations	7015	7015	7015	7015	7015
Total months enrolled by endline (age 4–6)					
Group: T1	0.839*	4.131***	-2.879***	-0.413	0.078
	(0.450)	(0.268)	(0.412)	(0.283)	(0.202)
Group: T2	0.934*	5.088***	-3.342***	-0.813***	-0.181
	(0.503)	(0.294)	(0.402)	(0.270)	(0.215)
Group: T3	1.165**	4.533***	-2.714***	-0.653**	0.050
	(0.481)	(0.355)	(0.439)	(0.282)	(0.218)
Group: Any T	0.949**	4.478***	-2.952***	-0.577**	0.005
	(0.414)	(0.182)	(0.403)	(0.255)	(0.185)
Control group mean	*5.370*	*0.102*	*3.774*	*1.494*	*2.053*
p-value: T1=T2	0.817	0.0155	0.00135	0.0435	0.134
p-value: T1=T3	0.393	0.362	0.467	0.260	0.873
p-value: T2=T3	0.600	0.224	0.00271	0.414	0.227
Observations	7015	7015	7015	7015	7015

No additional control variables included in regression model of outcome variable on treatment group dummy variables. Estimates correct for heteroskedasticity and within-village correlations. Total months enrolled at endline measures total months of enrollment at current school and all previous schools. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.9: Total months enrolled by type of school and subgroups

	Any preschool (1)	CPS (2)	IPS (3)	SPS (4)	Primary (5)
Total months enrolled by endline (age 4–6)					
Female	1.062**	4.756***	-3.159***	-0.535*	0.022
	(0.494)	(0.214)	(0.466)	(0.281)	(0.232)
Male	0.863**	4.219***	-2.738***	-0.618**	0.003
	(0.425)	(0.184)	(0.419)	(0.269)	(0.199)
Age 2	1.030***	3.646***	-2.426***	-0.189	-0.076
	(0.391)	(0.187)	(0.363)	(0.146)	(0.080)
Age 3	0.772	4.701***	-3.023***	-0.906**	0.131
	(0.532)	(0.228)	(0.534)	(0.380)	(0.199)
Age 4	1.025*	5.166***	-3.469***	-0.671*	-0.125
	(0.539)	(0.230)	(0.538)	(0.359)	(0.359)
Stunted	0.818	4.313***	-3.086***	-0.408	-0.253
	(0.500)	(0.211)	(0.488)	(0.250)	(0.227)
Not stunted	0.818	4.313***	-3.086***	-0.408	-0.253
	(0.500)	(0.211)	(0.488)	(0.250)	(0.227)
Wealth quartile 1	0.242	3.722***	-3.208***	-0.272	-0.006
	(0.680)	(0.244)	(0.725)	(0.302)	(0.275)
Wealth quartile 2	1.439***	4.515***	-2.765***	-0.311	0.014
	(0.526)	(0.227)	(0.473)	(0.271)	(0.281)
Wealth quartile 3	1.528**	5.068***	-3.083***	-0.457	-0.474
	(0.598)	(0.250)	(0.562)	(0.342)	(0.316)
Wealth quartile 4	0.539	4.748***	-2.812***	-1.397***	0.621*
	(0.552)	(0.256)	(0.544)	(0.418)	(0.336)
Education background quartile 1	0.810	4.349***	-3.201***	-0.339	-0.192
	(0.634)	(0.238)	(0.621)	(0.272)	(0.297)
Education background quartile 2	1.094**	4.564***	-3.123***	-0.348	-0.037
	(0.554)	(0.216)	(0.559)	(0.285)	(0.286)
Education background quartile 3	1.215***	4.669***	-2.581***	-0.873**	0.046
	(0.461)	(0.252)	(0.459)	(0.359)	(0.279)
Education background quartile 4	0.838	4.467***	-2.809***	-0.820**	0.287
	(0.582)	(0.248)	(0.474)	(0.380)	(0.311)
Cognition quartile 1	0.674	3.487***	-2.696***	-0.118	-0.085
	(0.444)	(0.202)	(0.442)	(0.170)	(0.152)
Cognition quartile 2	1.104*	4.400***	-2.799***	-0.497*	-0.201
	(0.604)	(0.230)	(0.582)	(0.282)	(0.228)
Cognition quartile 3	1.329**	4.838***	-2.886***	-0.623*	0.033
	(0.587)	(0.236)	(0.593)	(0.370)	(0.295)
Cognition quartile 4	0.714	5.404***	-3.567***	-1.123***	0.373
	(0.626)	(0.260)	(0.520)	(0.421)	(0.392)

No additional control variables included in regression model of outcome variable on joint treatment group (T1-T3) dummy variable. Estimates correct for heteroskedasticity and within-village correlations. Total months enrolled at endline measures total months of enrollment at current school and all previous schools. Education background refers to the maximum years of schooling by a household member or caregiver. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.10: ITT impacts on child development outcomes

	Cognitive development index		Socioemotional problems	
Midline (age 3–5)				
T1	0.044*	0.048**	-0.062	-0.068*
	(0.023)	(0.020)	(0.038)	(0.038)
T2	-0.000	0.009	-0.097**	-0.103**
	(0.027)	(0.024)	(0.042)	(0.042)
T3	0.050*	0.045**	-0.128***	-0.116***
	(0.025)	(0.022)	(0.045)	(0.045)
Any treatment	0.034*	0.037**	-0.088**	-0.090**
	(0.020)	(0.018)	(0.035)	(0.035)
Observations	6917	6917	6990	6990
Endline (age 4–6)				
T1	0.023	0.031	0.024	0.013
	(0.027)	(0.024)	(0.038)	(0.037)
T2	-0.026	-0.015	-0.001	-0.011
	(0.033)	(0.029)	(0.043)	(0.042)
T3	0.032	0.023	-0.016	-0.005
	(0.034)	(0.031)	(0.044)	(0.043)
Any treatment	0.013	0.017	0.007	0.002
	(0.024)	(0.021)	(0.035)	(0.035)
Observations	6966	6966	7014	7014
Additional controls	N	Y	N	Y

All regressions control for baseline value of dependent variable, child age, child age squared, and province fixed effects. Additional control variables include gender, baseline height-for-age z-score, household size, household wealth quintile dummy variables, and baseline caregiver age, gender, Raven's score, and parenting scores. Missing baseline covariates are replaced by the sample mean and interacted with a missing covariate dummy. Standard errors clustered on village level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.11: ITT impacts on subdomains of child development

	Executive function	Language	Early numeracy	Fine motor	Gross motor
Midline (age 3–5)					
T1	0.042	0.048*	0.061**	0.070**	0.014
	(0.026)	(0.028)	(0.029)	(0.032)	(0.032)
T2	0.004	0.025	0.014	-0.002	-0.009
	(0.028)	(0.036)	(0.033)	(0.038)	(0.037)
T3	0.038	0.039	0.061*	0.052	0.023
	(0.027)	(0.030)	(0.032)	(0.034)	(0.034)
Any treatment	0.031	0.040	0.049*	0.046	0.010
	(0.023)	(0.026)	(0.025)	(0.030)	(0.029)
Observations	6917	6917	6917	6917	6917
Endline (age 4–6)					
T1	-0.008	0.051*	0.007	-0.018	
	(0.025)	(0.028)	(0.026)	(0.031)	
T2	-0.046	0.006	-0.013	-0.079**	
	(0.029)	(0.036)	(0.030)	(0.033)	
T3	0.015	0.026	0.015	-0.053	
	(0.030)	(0.035)	(0.032)	(0.034)	
Any treatment	-0.012	0.033	0.004	-0.043	
	(0.021)	(0.026)	(0.023)	(0.027)	
Observations	6966	6966	6966	6966	
Additional controls	Y	Y	Y	Y	Y

All regressions control for baseline value of dependent variable, child age, child age squared, and province fixed effects. Additional control variables include gender, baseline height-for-age z-score, household size, household wealth quintile dummy variables, and baseline caregiver age, gender, Raven's score, and parenting scores. Missing baseline covariates are replaced by the sample mean and interacted with a missing covariate dummy. Standard errors clustered on village level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.12: ITT effects by subgroups

	Midline (age 3–5)		**Endline (age 4–6)**	
Baseline characteristics	Cognitive development index	Socio-emotional problems	Cognitive development index	Socio-emotional problems
Female	0.042*	-0.085*	0.017	0.003
	(0.024)	(0.044)	(0.031)	(0.046)
Male	0.033	-0.091**	0.020	0.005
	(0.024)	(0.043)	(0.024)	(0.044)
Age 2	0.024	-0.077	-0.008	-0.026
	(0.025)	(0.062)	(0.020)	(0.064)
Age 3	0.038	-0.080	0.035	0.027
	(0.031)	(0.049)	(0.034)	(0.050)
Age 4	0.050	-0.102*	0.018	0.004
	(0.038)	(0.056)	(0.044)	(0.065)
Stunted	-0.005	-0.070	0.011	-0.014
	(0.025)	(0.052)	(0.028)	(0.053)
Not stunted	0.060***	-0.093**	0.020	0.003
	(0.023)	(0.041)	(0.027)	(0.040)
Wealth quartile 1	0.035	-0.099	-0.020	-0.042
	(0.033)	(0.078)	(0.029)	(0.083)
Wealth quartile 2	0.011	-0.134**	0.013	-0.054
	(0.033)	(0.052)	(0.034)	(0.054)
Wealth quartile 3	0.013	0.009	-0.034	0.097
	(0.039)	(0.062)	(0.042)	(0.066)
Wealth quartile 4	0.085**	-0.099	0.134***	0.030
	(0.037)	(0.069)	(0.042)	(0.064)
Education background quartile 1	0.032	-0.151**	0.018	-0.092*
	(0.034)	(0.063)	(0.036)	(0.054)
Education background quartile 2	0.023	-0.042	-0.006	0.097
	(0.033)	(0.056)	(0.036)	(0.068)
Education background quartile 3	-0.008	-0.090	-0.038	-0.005
	(0.036)	(0.076)	(0.038)	(0.080)
Education background quartile 4	0.116***	-0.086	0.113**	0.016
	(0.040)	(0.063)	(0.044)	(0.061)
Cognition quartile 1	-0.021	-0.007	-0.042	0.040
	(0.033)	(0.063)	(0.031)	(0.070)
Cognition quartile 2	0.052*	-0.103	0.042	-0.029
	(0.027)	(0.069)	(0.030)	(0.056)
Cognition quartile 3	0.038	-0.075	-0.033	0.057
	(0.036)	(0.068)	(0.043)	(0.069)
Cognition quartile 4	0.085*	-0.134**	0.103*	-0.055
	(0.044)	(0.058)	(0.053)	(0.062)

Tables shows estimates from separate regressions of an outcome variables on a joint treatment group (T1-T3) dummy variable and control variables. Education background refers to the maximum years of schooling by a household member or caregiver. All regressions control for individual baseline test scores, child age, child age squared, province fixed effects, gender, baseline height-for-age z-score, household size, household wealth quintile dummy variables and baseline caregiver age, gender, Raven's score, and parenting scores. Missing baseline covariates are replaced by the sample mean and interacted with a missing covariate dummy. Standard errors clustered on village level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.13: Measuring take-up of demand-side interventions

	Obs.	C	T1-C	T2-C	T3-C
Midline:					
Ever received home visit to discuss development of child	6552	0.688	0.012	-0.018	0.027
		(0.02)	(0.02)	(0.03)	(0.03)
Ever received information leaflet about preschool	6552	0.053	0.020*	0.056***	0.079***
		(0.01)	(0.01)	(0.01)	(0.01)
Participated in HBP to discuss development of children (last 12 months)	6552	0.176	0.065***	0.045	0.190***
		(0.02)	(0.02)	(0.03)	(0.03)
··· participated more than once (last 12 months)	6552	0.118	0.040**	0.027	0.133***
		(0.02)	(0.02)	(0.02)	(0.03)
Endline:					
Ever received home visit to discuss development of child	6575	0.742	0.001	0.038	0.03
		(0.02)	(0.03)	(0.03)	(0.03)
Ever received information leaflet about preschool	6575	0.075	0.008	0.105***	0.083***
		(0.01)	(0.01)	(0.02)	(0.02)
Participated in HBP to discuss development of children (last 12 months)	6575	0.178	0.056**	0.03	0.100***
		(0.02)	(0.02)	(0.02)	(0.03)
··· participated more than once (last 12 months)	6586	0.096	0.047***	0.035*	0.075***
		(0.01)	(0.02)	(0.02)	(0.02)

C is the control group mean and constant in a regression of the outcome variable on a set of dummy variables for each treatment group. No control variables included in regression model. Estimates correct for heteroskedasticity. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.14: Perceived return to education

	Optimal preschool age ≤3	Optimal preschool age ≤4	Optimal preschool age ≤5	Return primary school	Return secondary school
Midline:					
Group: T1 (CPS)	0.062**	0.065**	0.010	0.007	0.000
	(0.024)	(0.025)	(0.011)	(0.004)	(0.003)
Group: T2 (CPS+D2D)	0.049*	0.080***	0.017	0.010**	0.004
	(0.026)	(0.027)	(0.011)	(0.005)	(0.004)
Group: T3 (CPS+D2D+HBP)	0.105***	0.106***	0.013	0.004	0.002
	(0.029)	(0.027)	(0.012)	(0.005)	(0.004)
Control group mean	*0.259*	*0.631*	*0.947*	*0.065*	*0.086*
Group: T2 or T3	0.078***	0.093***	0.015	0.007	0.003
	(0.023)	(0.024)	(0.011)	(0.004)	(0.004)
Observations	6989	6989	6989	6911	6971
Endline:					
Group: T1 (CPS)	0.031	0.073***	0.014	0.004	0.001
	(0.020)	(0.025)	(0.013)	(0.004)	(0.003)
Group: T2 (CPS+D2D)	0.040*	0.069**	0.011	0.005	0.002
	(0.023)	(0.029)	(0.014)	(0.004)	(0.003)
Group: T3 (CPS+D2D+HBP)	0.041*	0.079***	0.026*	0.001	0.002
	(0.024)	(0.027)	(0.014)	(0.004)	(0.004)
Control group mean	*0.188*	*0.501*	*0.916*	*0.068*	*0.085*
Group: T2 or T3	0.041**	0.074***	0.019	0.003	0.002
	(0.020)	(0.024)	(0.012)	(0.004)	(0.003)
Observations	7021	7021	7021	6992	7004

No control variables included in regression model. Row "Group: T2 or T3" shows parameter of regression model with a joint dummy variable for T2 and T3. Estimates correct for heteroskedasticity and within-village correlations. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

4.B Appendix

4.B.1 Additional Figures

Source: Own pictures.

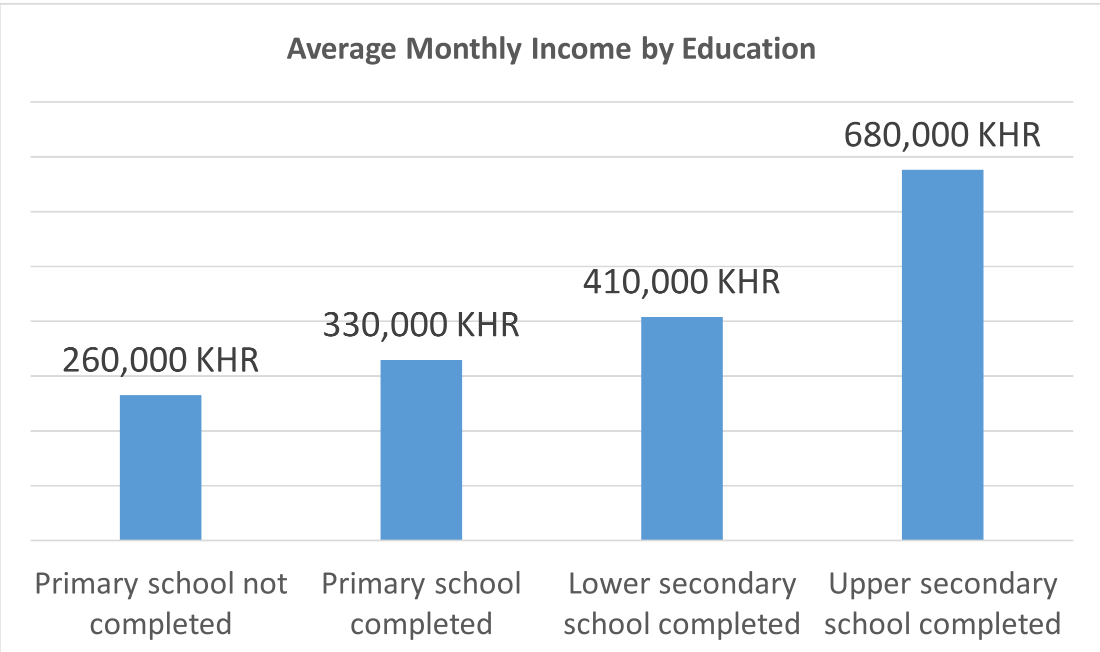
Figure 4.2: English version of leaflet used in door-to-door intervention at baseline.

Community Preschool

Your village has been selected to benefit from an improved community preschool supported by the Cambodian Ministry of Education. To make sure that children have enough space to learn and thrive, the preschool will have its own building and equipment. A trained teacher will prepare children of age 3-5 for primary school for 2 hours per day. **If your child is of age 3-5, it is at the right age to benefit from the preschool. Preschool is free for all children.**

It is important for your child that it constantly learns new things. Preschool education can help children to become more intelligent and well-behaved. The community preschool is a place where children can learn how to interact with each other and learn about honesty, respect, sharing and perseverance. They will also learn about numbers, letters and words. Visiting a community preschool can help your child to stay in school longer and to do well in his/her future.

School education is very important for children. Data from the Cambodian Socioeconomic Survey 2009 has shown that children who stay in school longer are likely to earn more in the future.



Source: Own illustration.

Figure 4.3: English version of leaflet used in door-to-door intervention at midline.

Community Preschool

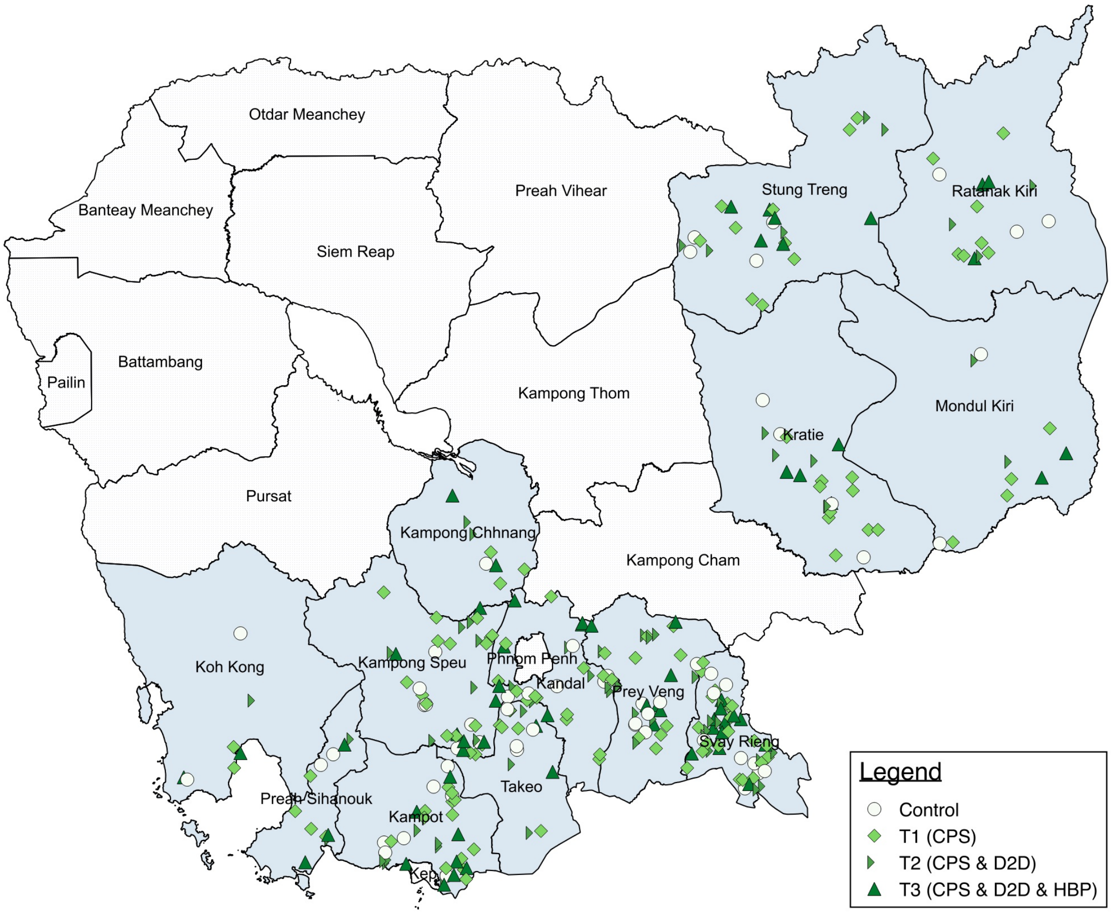
A stimulating environment is crucial for optimal development of your child. Preschool education can contribute to a better future for your child by providing new learning experiences every day.

At preschool, a trained teacher will help your child to learn important values such as respect, sharing and perseverance. Children will also be prepared for primary school by learning about numbers, letters and words.

If you have a child age 3-5, you can enroll your child at preschool!

Source: Own illustration.

Figure 4.4: Location of treatment and control group villages



Source: Own illustration.

Figure 4.5: Preschool enrollment at day of midline and endline survey

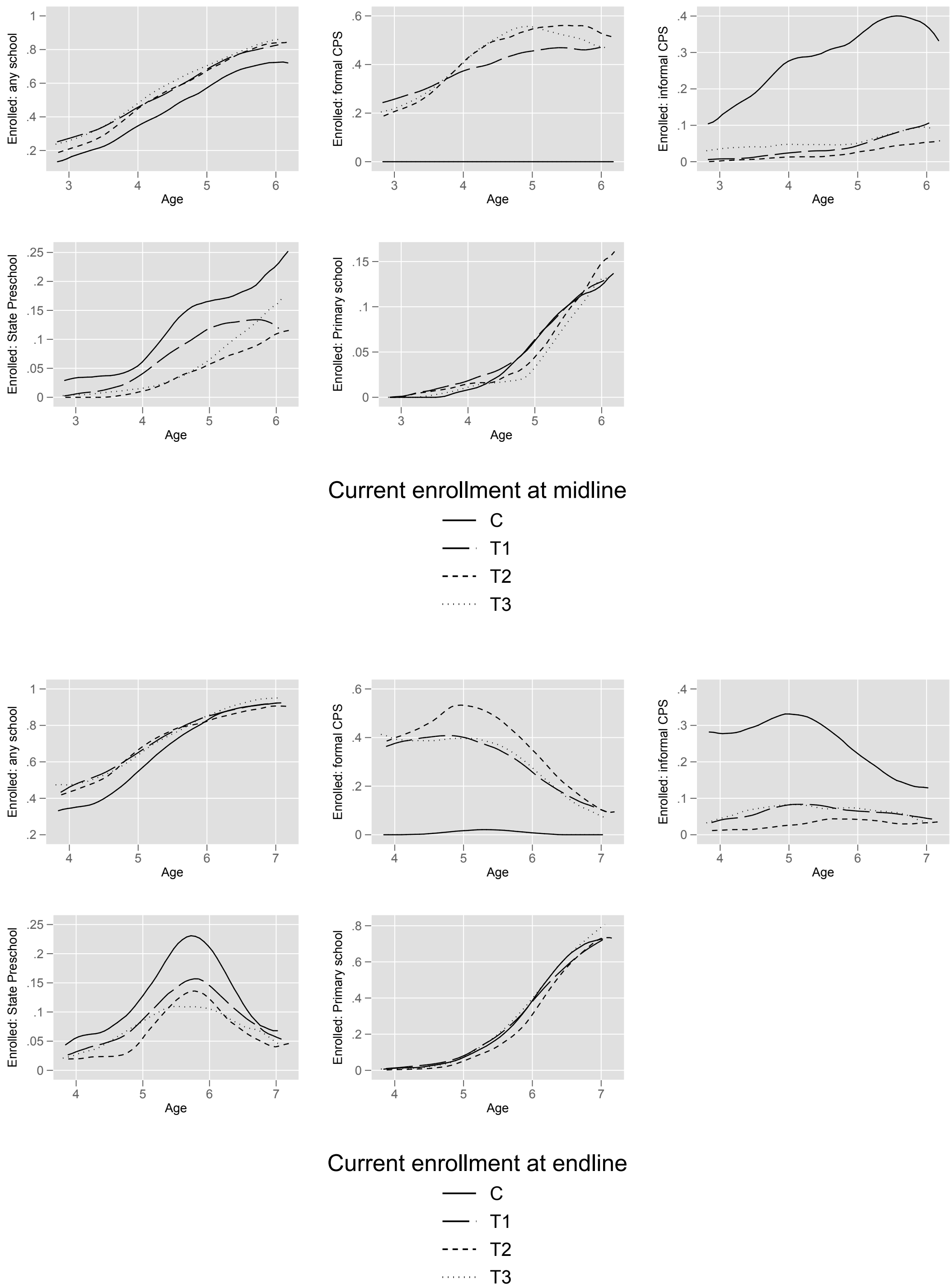
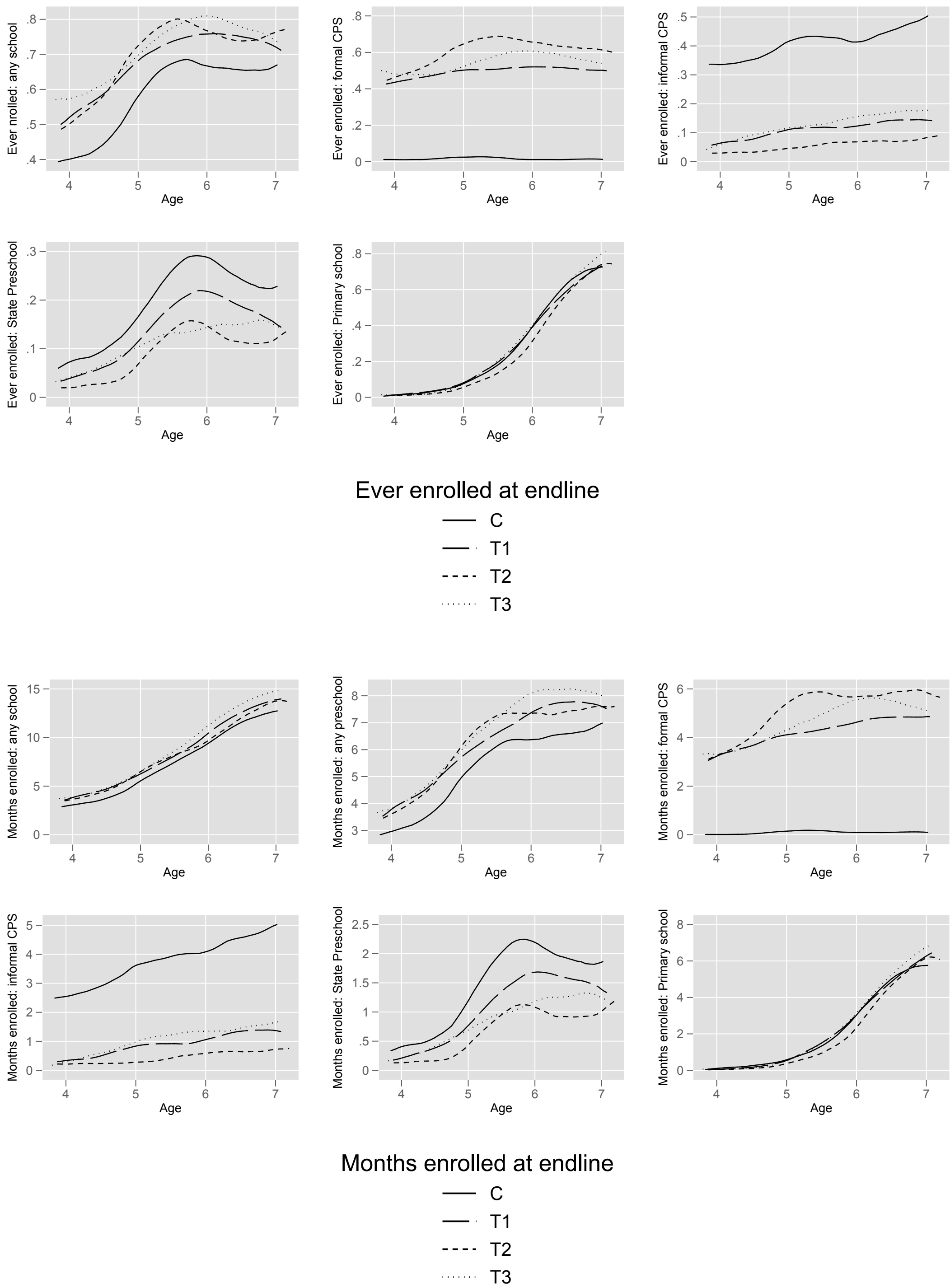


Figure shows separate local polynomial regressions using an epanechnikov kernel and a bandwidth of 3.6 months.

Figure 4.6: Total enrollment at endline by type of school



Total months enrolled at endline measures total months of enrollmentat current school and all previous schools. Figure shows separate local polynomial regressions using an epanechnikov kernel and a bandwidth of 3.6 months.

4.B.2 Additional Tables

Table 4.15: Full list of balance in baseline covariates (1/2)

	N	C	T1-C	T2-C	T3-C	Any T-C
	(1)	(2)	(3)	(4)	(5)	(6)
Child characteristics						
Cognitive development index	7491	0.030	-0.001	0.006	-0.006	-0.000
		[0.977]	(0.038)	(0.045)	(0.042)	(0.036)
Executive function score	7491	0	0.036	0.020	0.013	0.026
		[1.000]	(0.037)	(0.044)	(0.041)	(0.035)
Language score	7491	0	-0.010	0.001	-0.026	-0.011
		[1.000]	(0.039)	(0.044)	(0.042)	(0.037)
Mathematics score	7491	0	-0.033	-0.020	-0.014	-0.025
		[1.000]	(0.037)	(0.045)	(0.042)	(0.035)
Fine motor score	7491	0	0.052	0.012	0.027	0.035
		[1.000]	(0.043)	(0.048)	(0.046)	(0.040)
Gross motor score	7491	0	0.031	0.009	0.002	0.018
		[1.000]	(0.034)	(0.039)	(0.038)	(0.031)
Socio-emotional problems (SDQ)	7611	-0.003	0.027	0.042	0.002	0.024
		[0.996]	(0.043)	(0.048)	(0.045)	(0.040)
Age (yrs) w. decimals	7632	3.476	0.008	0.020	-0.039	-0.001
		[0.824]	(0.029)	(0.032)	(0.031)	(0.027)
Female	7632	0.506	-0.020	-0.030*	-0.004	-0.019
		[0.500]	(0.017)	(0.018)	(0.017)	(0.015)
Stunted (lhfa<2sd)	7473	0.341	0.018	0.020	0.032	0.022
		[0.474]	(0.018)	(0.022)	(0.022)	(0.017)
Household characteristics						
Household size	7632	5.966	-0.030	0.064	-0.004	0.001
		[2.110]	(0.103)	(0.133)	(0.113)	(0.097)
Wealth score	7631	-0.047	-0.001	0.036	0.089	0.032
		[0.971]	(0.066)	(0.076)	(0.067)	(0.059)
Monthly income > 100 USD	7632	0.464	-0.010	-0.068*	-0.034	-0.031
		[0.499]	(0.032)	(0.040)	(0.036)	(0.030)
Farming activity	7630	0.822	0.002	0.000	0.023	0.007
		[0.383]	(0.023)	(0.030)	(0.025)	(0.022)
Maximum years of education	7598	6.044	-0.322	-0.473*	-0.044	-0.287
		[3.700]	(0.218)	(0.246)	(0.234)	(0.201)

Differences in means are based on OLS regressions at the child level on province fixed effects and three treatment group dummies (columns 3-5) or a joint treatment group dummy (column 6). Robust standard errors clustered at village level in parentheses. Control group standard deviations in square brackets. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.16: Full list of balance in baseline covariates (2/2)

	N	C	T1-C	T2-C	T3-C	Any T-C
	(1)	(2)	(3)	(4)	(5)	(6)
Caregiver characteristics						
Caregiver female	7629	0.890	0.024**	0.029**	0.017	0.023**
		[0.313]	(0.012)	(0.013)	(0.014)	(0.011)
Caregiver age	7629	40.777	0.415	-1.188	0.173	-0.054
		[14.688]	(0.644)	(0.742)	(0.680)	(0.584)
Caregiver biological parent	7542	0.596	-0.024	0.028	-0.003	-0.005
		[0.491]	(0.022)	(0.026)	(0.024)	(0.020)
Caregiver years of education	7621	4.681	-0.286	-0.264	0.292	-0.127
		[11.120]	(0.428)	(0.450)	(0.495)	(0.392)
Caregiver Raven's score	7558	0.050	-0.070	0.011	-0.006	-0.033
		[1.017]	(0.045)	(0.048)	(0.049)	(0.041)
Cognitive parenting	7611	-0.006	0.018	-0.012	0.073*	0.025
		[1.002]	(0.040)	(0.049)	(0.044)	(0.037)
Socioemotional parenting	7612	-0.008	-0.021	-0.112**	-0.022	-0.044
		[1.003]	(0.038)	(0.048)	(0.044)	(0.035)
Negative parenting	7613	0.002	0.052	0.092*	0.013	0.052
		[0.999]	(0.043)	(0.047)	(0.044)	(0.038)
Baseline program attendance						
Attending preschool	7612	0.153	0.073***	0.063**	0.108***	0.080***
		[0.360]	(0.023)	(0.027)	(0.026)	(0.021)
Attending CPS or IPS	7612	0.123	0.068***	0.057**	0.101***	0.074***
		[0.329]	(0.023)	(0.026)	(0.026)	(0.021)
Attending SPS	7612	0.030	0.005	0.006	0.008	0.006
		[0.171]	(0.009)	(0.011)	(0.010)	(0.008)
Careg. participated in any HBP	7609	0.104	0.031*	0.055***	0.184***	0.078***
		[0.306]	(0.017)	(0.020)	(0.023)	(0.017)
Caregiver received any D2D	7609	0.017	0.003	0.012*	0.021***	0.010*
		[0.130]	(0.006)	(0.007)	(0.008)	(0.006)

Differences in means are based on OLS regressions at the child level on province fixed effects and three treatment group dummies (columns 3-5) or a joint treatment group dummy (column 6). Robust standard errors clustered at village level in parentheses. Control group standard deviations in square brackets. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.17: Attrition of eligible children at midline and endline

	Midline			Endline		
	(1)	(2)	(3)	(1)	(2)	(3)
T1	0.019	-0.062	0.021*	-0.003	0.067	-0.001
T2	0.008	0.058	0.009	0.008	0.091	0.009
T3	-0.008	0.001	-0.006	-0.002	0.050	0.000
Cognitive development index		0.006			-0.010	
Age		-0.015			0.008	
Height-for-age z-score		-0.005			-0.009	
Multidim. poor		0.002			0.016	
Caregiver years of education		0.001			-0.001	
T1 * Cognitive development index		-0.015			0.006	
T1 * Age		0.014			-0.020	
T1 * Height-for-age z-score		-0.007			0.006	
T1 * Multidim. poor		0.028			0.001	
T1 * Caregiver years of education		0.001			0.001	
T2 * Cognitive development index		-0.002			-0.006	
T2 * Age		-0.007			-0.007	
T2 * Height-for-age z-score		0.009			0.041***	
T2 * Multidim. poor		-0.011			0.019	
T2 * Caregiver years of education		-0.002			-0.001	
T3 * Cognitive development index		0.001			0.008	
T3 * Age		0.001			-0.015	
T3 * Height-for-age z-score		0.004			0.005	
T3 * Multidim. poor		-0.004			0.003	
T3 * Caregiver years of education		-0.001			0.001	
Control group mean attrition	0.102			0.0956		
Joint F-test (p-values):						
Baseline controls (without interaction)		0.847			0.190	
T1 interactions with baseline controls		0.525			0.423	
T2 interactions with baseline controls		0.676			0.0434	
T3 interactions with baseline controls		0.986			0.912	
Additional sampling at midline	No	No	Yes	No	No	Yes
Observations	7632	7632	7693	7632	7632	7693

Table shows OLS regressions with midline and endline attrition as dependent variable using all children eligible for testing at baseline (columns 1 and 2) plus children that were added to the sample at midline (column 3). All regressions also control for province fixed effects. Robust standard errors are clustered at village level. Missing baseline control variables are replaced by the control group mean. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.18: Teacher characteristics and preschool equipment

	CPS (1)	IPS (2)	SPS (3)	IPS-CPS (4)	SPS-CPS (5)	N (6)
Teacher characteristics	**-0.37**	**-0.22**	**1.41**	0.148	1.781***	328
Female	0.94	0.97	0.84	0.03	-0.10*	329
Age	40.46	40.92	34.05	0.45	-6.42***	329
Years since first teaching experience	6.04	6.46	5.97	0.41	-0.08	329
Completed primary school (6)	0.81	0.78	0.98	-0.03	0.18***	329
Completed lower secondary school (9)	0.43	0.39	0.86	-0.04	0.42***	329
Completed upper secondary school (12)	0.13	0.14	0.68	0.01	0.55***	329
No teacher training	0.12	0.15	0.29	0.04	0.17***	329
1-4 weeks of teacher training	0.13	0.32	0.27	0.19***	0.14**	329
5-8 weeks of teacher training	0.70	0.46	0.03	-0.24***	-0.67***	329
More than 8 weeks of teacher training	0.05	0.07	0.41	0.01	0.36***	329
Had practical teacher training	0.21	0.20	0.41	-0.00	0.20***	329
Trained as prim./sec. school teacher	0.02	0.05	0.44	0.03	0.42***	329
Nonverbal reasoning test (Raven's)	-0.10	-0.04	0.41	0.05	0.51***	328
Salary for teaching position (USD)	60.74	67.32	250.22	6.58	189.48***	329
Teacher fully paid, regularly	0.75	0.85	0.95	0.10*	0.20***	329
Teacher fully paid, irregularly	0.22	0.15	0.03	-0.07	-0.19***	329
Equipment	**0.08**	**-0.51**	**0.22**	**-0.592*****	**0.138**	**326**
Table and chair for teacher	0.99	0.34	0.87	-0.65***	-0.12***	329
Storage for teacher	0.97	0.19	0.52	-0.78***	-0.44***	329
Tables and chairs for children	0.95	0.32	0.67	-0.63***	-0.29***	329
Childrens' tables and chairs appropriately sized	0.78	0.29	0.34	-0.48***	-0.44***	327
Blackboard/whiteboard and markers/chalks	0.95	0.83	0.97	-0.12**	0.02	329
Electricity access	0.07	0.25	0.24	0.19***	0.17***	329
Field, playground or school yard	0.62	0.61	0.83	-0.01	0.21***	329
Equipment for gross-motor activities on school yard	0.38	0.32	0.30	-0.05	-0.08	329
First aid kit	0.31	0.12	0.33	-0.20***	0.02	329
Functional water source	0.46	0.64	0.81	0.18**	0.35***	329
Functional drinking water source	0.61	0.54	0.67	-0.07	0.05	329
Hand washing facility	0.54	0.37	0.57	-0.16**	0.04	329
Toilet facility	0.27	0.42	0.90	0.15**	0.63***	329
Writing utensils	0.94	0.80	0.89	-0.14**	-0.05	329
Writing utensils used by children	0.61	0.63	0.79	0.01	0.18***	329
Art materials	0.91	0.71	0.81	-0.20***	-0.10*	329
Art materials used by children	0.56	0.53	0.70	-0.03	0.14**	329
Fantasy play materials	0.65	0.29	0.25	-0.36***	-0.40***	329
Fantasy play materials used by children	0.40	0.19	0.08	-0.21***	-0.32***	329
Educational toys/math materials	0.77	0.44	0.67	-0.33***	-0.10	329
Educational toys/math materials used by children	0.43	0.25	0.46	-0.18***	0.03	329
Number of schools	207	59	63			

Columns 1—3 show averages by type of preschool. Columns 4-5 show differences between types of preschools. Differences are based on regressions of dependent variables on binary preschool type variables using robust standard errors. Summary scores are prepared using the first principal component of the individual variables and are standardized to a mean of zero and a standard deviation of one. Observations with missing values are dropped from the summary scores. Variables with a *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.19: Classroom setting and teaching practices

	CPS (1)	IPS (2)	SPS (3)	IPS-CPS (4)	SPS-CI (5)
Classroom setting	**0.35**	**-1.07**	**-0.18**	**-1.419*** **	**-0.539***
Length of class (minutes)	113.78	137.53	172.49	23.74	58.71***
Total length of breaks (minutes)	43.22	21.07	44.22	-22.15*	1.00
Number of children enrolled in this class	25.09	22.56	27.57	-2.53**	2.48**
Children present	17.66	14.98	21.21	-2.68**	3.54***
Num. of teachers in classroom	1.01	0.98	0.98	-0.03	-0.03
Num. of assistants in classroom	0.03	0.03	0.00	0.00	-0.03**
Num. of other adults in classroom	1.03	0.73	0.31	-0.30	-0.72***
Teacher follows curriculum to teach class	0.68	0.46	0.49	-0.22***	-0.18**
Teacher documents children's development regularly	0.38	0.19	0.37	-0.19***	-0.01
Teacher documents children's attendance	0.87	0.68	0.86	-0.20***	-0.02
Curriculum content and pedagogy	**-0.05**	**-0.23**	**0.38**	**-0.179**	**0.428****
Activities supporting development of maths skills$^{\times}$	0.82	0.80	0.71	-0.02	-0.10
Quality of maths activities [1-4]	2.68	2.47	2.89	-0.21	0.21
Activities supporting development of literacy skills$^{\times}$	0.73	0.76	0.89	0.03	0.16***
Quality of literacy activities [1-4]	2.66	2.47	3.39	-0.19	0.74***
Activities supporting development of expressive language skills$^{\times}$	0.88	0.80	0.78	-0.09	-0.11*
Quality of expressive language activities [1-3]	2.37	2.28	2.67	-0.09	0.31**
Activity: reading of storybook$^{\times}$	0.54	0.47	0.46	-0.07	-0.08
Quality of storybook activities [1-6]	3.54	2.93	3.86	-0.62*	0.31
Activities supporting development of general knowledge$^{\times}$	0.85	0.90	0.87	0.05	0.03
Quality of teaching during general knowledge activities [1-6]	3.78	3.89	4.00	0.10	0.22
Activities supporting development of fine motor skills$^{\times}$	0.39	0.41	0.57	0.02	0.18**
Quality of teaching during fine motor skills activities [1-3]	2.04	2.09	2.31	0.05	0.27**
Activities supporting development of gross motor skills$^{\times}$	0.67	0.58	0.75	-0.09	0.08
Quality of gross motor skills activities [1-3]	1.63	1.56	1.85	-0.07	0.22
Time of gross motor skills activities (minutes)$^{\times}$	7.87	8.75	10.42	0.88	2.55**
Quality of the teacher's use of theme [0(no theme used)-4]$^{\times}$	2.06	2.13	2.17	0.06	0.11
Teacher-child interactions	**-0.06**	**-0.10**	**0.29**	**-0.044**	**0.347****
The teacher enjoyed teaching [1-3]	2.58	2.49	2.60	-0.09	0.02
The teacher showed negative attitudes [0-2.5]	0.06	0.07	0.06	0.00	0.00
Quality of the disciplinary strategies used by the teacher [0-4]	2.95	3.02	3.13	0.07	0.18
More than 4 occurences of negative interactions	0.44	0.53	0.51	0.09	0.07
More than 8 occurences of encouragements	0.43	0.44	0.60	0.01	0.17**
Children wait more than 5 minutes without any specific activity	0.15	0.10	0.10	-0.05	-0.06
The teacher correct student's work and give feedbacks [0-3]	1.86	2.00	2.32	0.14	0.46***
Children ever left without supervision	0.14	0.20	0.24	0.06	0.09
Quality of the engagement of children [0-4]	2.57	2.76	3.25	0.20	0.69***
Teacher's awareness of children's individual needs [0-3]	1.41	1.36	1.44	-0.05	0.04
Teacher's behavior with respect to gender equality [0-4]	3.26	3.19	3.35	-0.08	0.09
Class is interrupted at least once	0.58	0.58	0.62	-0.00	0.04
Presence of disturbing noise [1-3]	1.36	1.47	1.27	0.11	-0.09
Number of schools	207	59	63		

Columns 1—3 show averages by type of preschool. Columns 4-5 show differences between types of preschools. Diff based on regressions of dependent variables on binary preschool type variables using robust standard errors. Summary prepared using the first principal component of the individual variables and are standardized to a mean of zero and deviation of one. Observations with missing values are dropped from the summary scores. Variables marked with $^{\times}$ we from the principal component analysis to avoid missing values in summary scores. Activities which were not taught we a quality score of 0. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.20: ITT effects on individual tests

	Midline (age 3–5)				Endline (age 4–6)			
	Any T	T1	T2	T3	Any T	T1	T2	T3
Executive function								
Head-knee task	0.004	0.009	-0.028	0.027	0.027	0.010	0.007	0.078**
Forward digit-span test	0.029	0.043	0.009	0.024	0.027	0.031	0.009	0.036
DCCS	0.045	0.049	0.043	0.040	-0.018	0.000	-0.030	-0.039
Cancellation task	0.060**	0.075***	0.040	0.053*	-0.033	-0.023	-0.081**	-0.005
Language								
TVIP	0.018	0.023	0.014	0.015	0.009	0.012	0.001	0.011
Naming items	0.027	0.028	-0.009	0.061	0.008	0.017	-0.009	0.007
Short-story	0.033	0.043*	0.017	0.032	0.024	0.022	0.038	0.016
Reading concepts	0.025	0.047	-0.008	0.017	0.049*	0.065**	0.064*	0.004
Letter knowledge	0.050	0.041	0.101	0.015	0.028	0.053	-0.018	0.027
Name writing					0.019	0.031	-0.012	0.025
Initial letter identification					0.047	0.062*	0.028	0.039
Reading words					0.009	0.037	-0.033	-0.001
Early numeracy								
Measurement concepts	-0.014	-0.011	-0.031	-0.003				
Verbal counting	0.087***	0.105***	0.054*	0.086***	0.050*	0.063*	0.015	0.060
Quantitative comparison	0.019	0.020	-0.022	0.055*	0.003	0.005	-0.000	0.003
Number knowledge	0.062**	0.071**	0.061	0.046	0.034	0.044	0.000	0.049
Shape recognition	0.005	-0.006	-0.000	0.028	-0.001	-0.018	0.008	0.020
Arithmetic problem					-0.009	-0.007	-0.017	-0.004
Spatial vocabulary					-0.048*	-0.059**	-0.027	-0.047
Fine motor								
Copying	0.038	0.043	0.033	0.035	-0.042	-0.028	-0.053	-0.056
Draw-a-person	0.044	0.072**	-0.013	0.050	-0.022	0.008	-0.077**	-0.026
SDQ								
Emotional symptoms	-0.021	-0.007	-0.033	-0.034	-0.009	-0.011	-0.042	0.028
Conduct problems	-0.063**	-0.045	-0.085**	-0.074*	-0.015	-0.026	-0.015	0.004
Hyperactivity/inattention	-0.078**	-0.088**	-0.040	-0.097**	0.008	0.017	-0.001	-0.001
Peer problems	-0.053	-0.028	-0.088*	-0.065	0.014	0.044	0.025	-0.054
Prosocial	0.033	0.020	0.003	0.085**	0.026	0.022	0.025	0.034

All regressions control for individual baseline test scores, child age, child age squared, and province fixed effects. Additional control variables include gender, baseline height-for-age z-score, household size, household wealth quintile dummy variables, and baseline caregiver age, gender, Raven's score and parenting scores. Missing baseline covariates are replaced by the sample mean and interacted with a missing covariate dummy. Standard errors clustered on village level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.21: Impact of the program on parenting domains

	Cognitive parenting	Socio-emotional parenting	Negative parenting
Midline (age 3–5)			
T1	0.064*	0.018	0.002
	(0.034)	(0.043)	(0.035)
T2	0.036	0.040	0.008
	(0.037)	(0.048)	(0.041)
T3	0.070*	0.053	0.035
	(0.038)	(0.046)	(0.041)
Any treatment	0.058*	0.033	0.012
	(0.030)	(0.039)	(0.031)
Observations	6993	6993	6993
Endline (age 4–6)			
T1	-0.016	0.061	-0.002
	(0.035)	(0.038)	(0.038)
T2	-0.011	0.012	0.039
	(0.042)	(0.045)	(0.038)
T3	-0.016	0.035	0.004
	(0.043)	(0.043)	(0.045)
Any treatment	-0.015	0.041	0.010
	(0.033)	(0.035)	(0.035)
Observations	7017	7017	7017

All regressions control for baseline value of dependent variable, child age, child age squared, and province fixed effects. Additional control variables include gender, baseline height-for-age z-score, household size, household wealth quintile dummy variables, and baseline caregiver age, gender, Raven's score, and parenting scores. Missing baseline covariates are replaced by the sample mean and interacted with a missing covariate dummy. Standard errors clustered on village level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4.22: Impact of the program on parenting domains by subgroups

Baseline characteristics	Midline (age 3–5) Cognitive parenting	Socio-emotional parenting	Negative parenting	Endline (age 4–6) Cognitive parenting	Socio-emotional parenting	Negative parenting
Female	0.067*	0.068	0.065	-0.032	0.042	0.089*
	(0.037)	(0.055)	(0.041)	(0.046)	(0.049)	(0.047)
Male	0.052	-0.003	-0.034	0.008	0.037	-0.071
	(0.040)	(0.045)	(0.045)	(0.041)	(0.049)	(0.043)
Age 2	0.048	0.030	0.024	0.010	0.047	0.003
	(0.051)	(0.070)	(0.056)	(0.050)	(0.052)	(0.058)
Age 3	0.026	0.027	0.024	0.001	0.038	0.043
	(0.053)	(0.052)	(0.054)	(0.044)	(0.050)	(0.047)
Age 4	0.099*	0.046	0.003	-0.052	0.036	-0.008
	(0.053)	(0.065)	(0.052)	(0.061)	(0.062)	(0.053)
Stunted	0.059	0.048	-0.024	0.024	0.039	0.031
	(0.050)	(0.064)	(0.053)	(0.047)	(0.067)	(0.058)
Not stunted	0.032	0.019	0.022	-0.034	0.036	0.002
	(0.039)	(0.041)	(0.039)	(0.040)	(0.036)	(0.036)
Wealth quartile 1	0.047	0.021	0.025	-0.069	0.054	-0.045
	(0.055)	(0.076)	(0.064)	(0.059)	(0.075)	(0.069)
Wealth quartile 2	0.060	0.055	-0.031	0.013	0.030	0.055
	(0.053)	(0.059)	(0.059)	(0.061)	(0.062)	(0.066)
Wealth quartile 3	-0.025	0.027	0.024	-0.028	-0.010	-0.008
	(0.069)	(0.069)	(0.061)	(0.065)	(0.064)	(0.060)
Wealth quartile 4	0.156**	0.055	0.070	0.037	0.091	0.051
	(0.069)	(0.066)	(0.065)	(0.065)	(0.061)	(0.061)
Educ. backgr. quartile 1	0.034	0.036	0.006	0.037	0.254***	-0.014
	(0.054)	(0.060)	(0.055)	(0.051)	(0.064)	(0.056)
Educ. backgr. quartile 2	0.008	0.059	0.109*	-0.044	-0.052	0.099
	(0.054)	(0.070)	(0.060)	(0.056)	(0.064)	(0.064)
Educ. backgr. quartile 3	0.071	0.137*	0.007	0.032	-0.034	0.003
	(0.054)	(0.075)	(0.066)	(0.077)	(0.068)	(0.071)
Educ. backgr. quartile 4	0.143**	-0.074	-0.025	-0.024	-0.003	-0.021
	(0.072)	(0.058)	(0.062)	(0.063)	(0.060)	(0.064)
Cognition quartile 1	0.001	0.087	-0.021	-0.002	0.056	-0.044
	(0.050)	(0.079)	(0.059)	(0.057)	(0.058)	(0.060)
Cognition quartile 2	0.029	-0.027	-0.052	-0.034	0.046	0.071
	(0.062)	(0.065)	(0.063)	(0.062)	(0.065)	(0.061)
Cognition quartile 3	0.002	-0.045	0.033	0.056	0.011	0.057
	(0.061)	(0.072)	(0.051)	(0.060)	(0.062)	(0.058)
Cognition quartile 4	0.142**	0.080	0.055	-0.068	0.026	-0.041
	(0.068)	(0.063)	(0.055)	(0.069)	(0.062)	(0.057)

Tables shows estimates from separate regressions of an outcome variables on a joint treatment group (T1-T3) dummy variable and control variables. Education background refers to the maximum years of schooling by a household member or caregiver. All regressions control for individual baseline test scores, child age, child age squared, province fixed effects, gender, baseline height-for-age z-score, household size, household wealth quintile dummy variables and baseline caregiver age, gender, Raven's score, and parenting scores. Missing baseline covariates are replaced by the sample mean and interacted with a missing covariate dummy. Standard errors clustered on village level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

4.B.3 Child assessments

This section summarizes the individual tests, their distributions and the scoring methods used in this paper. An in-depth discussion of the tests, scoring methods, cultural adaptations and pretesting procedures can be found in Berkes et al. (2019).

To ensure that children correctly understood the tests and that the test were reliable, the research team pretested every instrument at least three times before collecting data in the sample villages. The survey firm translated the questionnaires into Khmer and an independent third party back-translated them into English which led to further refinements in the instruments. The final child assessments included a total of 15 individual tests at baseline, 17 at midline and 20 at endline.

Before constructing the composite scores of child test domains, individual tests were first scored and standardized thus ensuring that all tests contributed equal variance to their composite score. Scoring was done by assigning 1 point for each correct response and summing up these points to create an individual score for each test. When a child was unable to complete the practice trial of a test, a score of zero was assigned for this test as long as the child participated in the other tests. Standardization of each test score was done with the control group mean of the same wave by subtracting the mean and dividing by the standard deviation of this wave. All standardized test scores of one domain (e.g. executive function) were then summed into a domain score and standardized again by subtracting its sample mean and dividing by the sample standard deviation of the domain score for better interpretability. After these steps, we obtained the following composite scores:

1. **Executive function:**
   1.1. The construct inhibitory control is assessed with the head-knee task. The test has two stages. In the first stage, the child stands in front of the enumerator and is asked five times to touch his/her head or knees. In a second stage, the child is asked to do the opposite of what the enumerator says.

   1.2. Working memory (short-term auditory memory) is assessed with a forward digit span test in which children have to repeat sequences of digits which increase in length.

   1.3. The Dimensional Change Card Sort test is used as a measure of cognitive flexibility. We followed the procedures outlined in Zelazo (2006) using cards with two colours (blue and red) and two pictures (boat and rabbit). To reduce the burden on tested children, we followed the protocol with the exception that children needed to pass the pre-switch phase (at least 5 out of 6 correct) in order to participate in the post switch phase. The border version of the test was only administered at endline. The demonstration phase of the test included one practice trial. As per protocol, this practice trial was not used to determine whether a child is eligible for the test as it could have performed well by chance.

1.4. We use a self-developed cancellation task to measure sustained attention. In this test, children see a printed matrix with different symbols and are asked to cross-out all symbols that match the given one (e.g. cross out all flowers). When completed, a larger matrix is given and a new symbol has to be crossed-out. The test continues until a child has completed 4 matrices, crossed out more wrong than correct images in a matrix, until the child loses attention, or states that it is done. The test was scored by using the difference between correctly and incorrectly crossed out images.

2. **Language:**

2.1. Receptive vocabulary skills are assessed 78 with a test derived from the TVIP. In this test children are asked to match a word to one out of four pictures. The version used in the Cambodian context was culturally adapted during piloting and validation exercises prior to baseline data collection and with the support of key informants. The final instrument includes 82 pictures with a rule that the test stops after 6 out of the last 8 pictures were wrong. All other language development tests were taken from the MELQO.

2.2. Expressive language skills are assessed by asking children to name up to 10 things that can be eaten and up to 10 animals they know. The final score is the number of recalled items.

2.3. Receptive language is assessed with a listening comprehension test in which a short story (116 words) is read to the child. After reading the story, the child is asked five questions about the content of the story.

2.4. Knowledge of reading concepts is assessed by showing a children's storybook and asking how the book should be opened and where and in which direction one should start reading the story.

2.5. Reading skills are assessed with a letter name knowledge test in which children have to identify common letters of Khmer script.

2.6. *Endline only:* A name writing test was conducted to assess whether children were able to write their own name.

2.7. *Endline only:* An initial letter identification test was conducted in which children were asked to name the first alphabet letter letter of words that were read to the child

2.8. *Endline only:* Reading skills were assessed by asking the child to read out loud different printed words.

3. **Early numeracy:**

3.1. *Midline only:* The tests for early numeracy includes a self-developed test for measurement concepts, e.g. if the child understands concepts such as tallest/shortest, in which the child had to point to different printed objects.

3.2. In a test for verbal counting, children had to count up to 30.

3.3. Numbers and operations are also administered with a self-developed quantitative comparison test where children had to compare the number of printed objects on two sides of a page.

3.4. A number identification test analogous to the letter name knowledge test was used.

3.5. A self-developed shape recognition test was used to test if children are able to identify basic geometric shapes.

3.6. *Endline only:* Children were asked to read printed arithmetic problems and say the correct answer (e.g. 2+1).

3.7. *Endline only:* A spacial vocabulary test was conducted in which the child was shown 4 pictures with a ball and a chair. The child was asked to point to the correct picture with the ball either on, under, in front or next to the chair.

4. **Fine-motor development:**

4.1. A drawing test, where children copy shapes, like circles or squares, was used to assess fine-motor skills.

4.2. A draw-a-person test

5. **Gross-motor development (midline only):** The Malawi Developmental Assessment Tool (MDAT) was used for assessment of grossmotor skills.

6. **Socio-emotional problems:** The recommended method was used to create a total difficulties score, i.e. summing up scores of the individual subcomponents without standarizing first. The subcomponents are:

6.1. Emotional symptoms

6.2. Conduct problems

6.3. Hyperactivity/inattention

6.4. Peer problems

Figure 4.7: Distribution of midline cognition composite scores and individual tests.

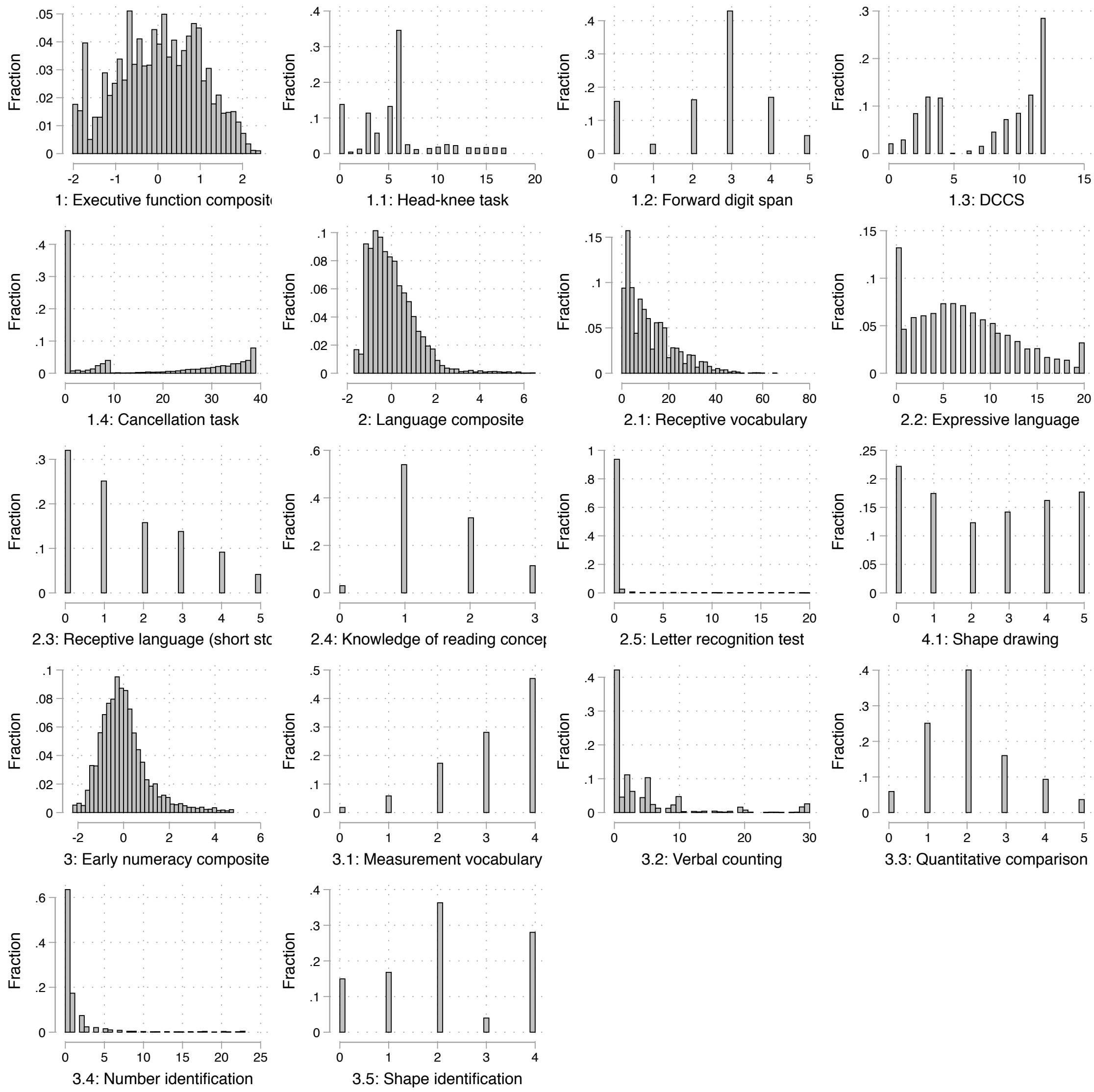


Figure shows empirical distribution of composite and raw midline cognition test scores.

Figure 4.8: Distribution of midline motor-development and socio-emotional composite scores and individual tests.

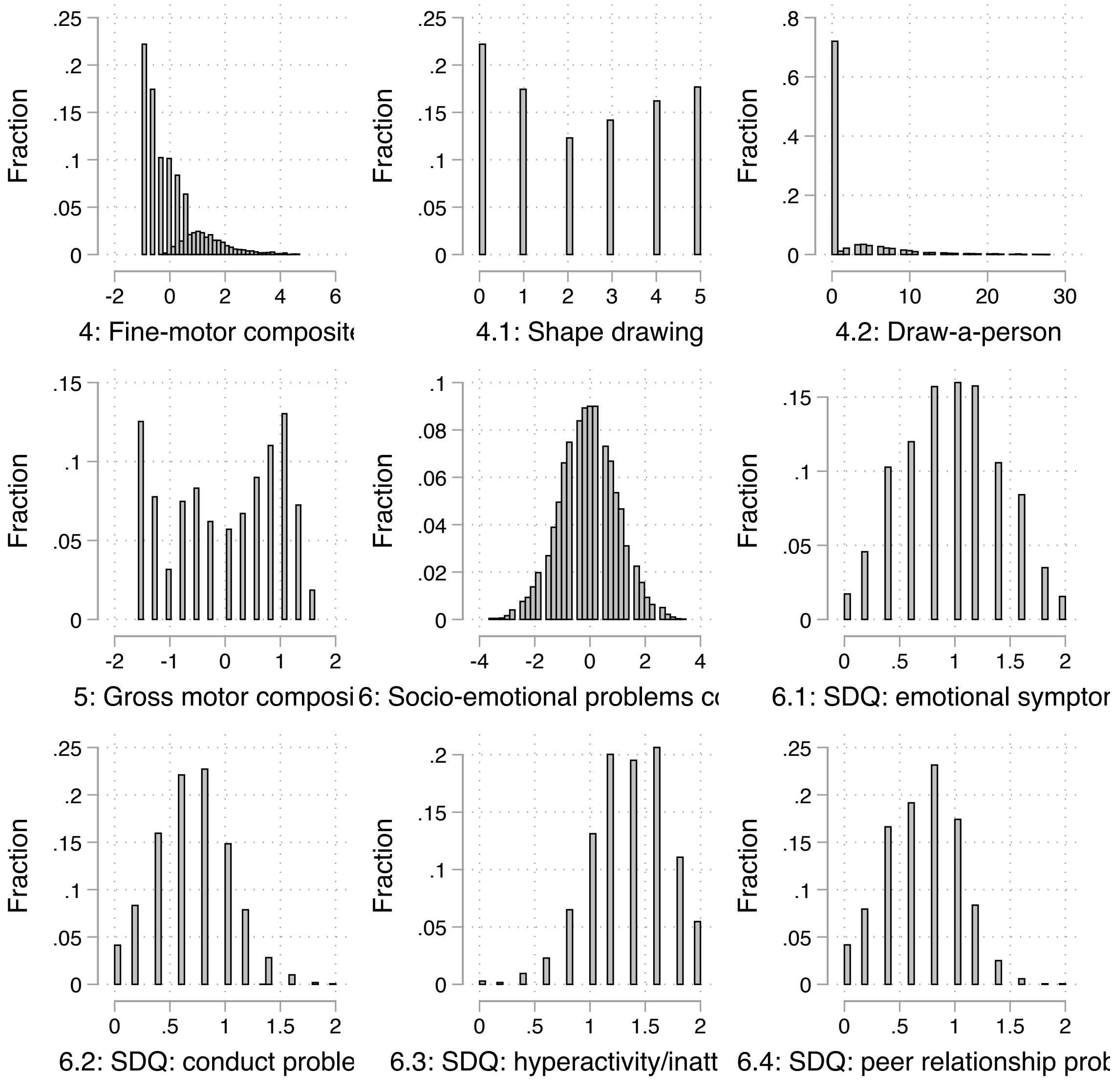


Figure shows empirical distribution of composite and raw motor-development and socio-emotional midline test scores.

Figure 4.9: Distribution of endline cognition composite scores and individual tests.

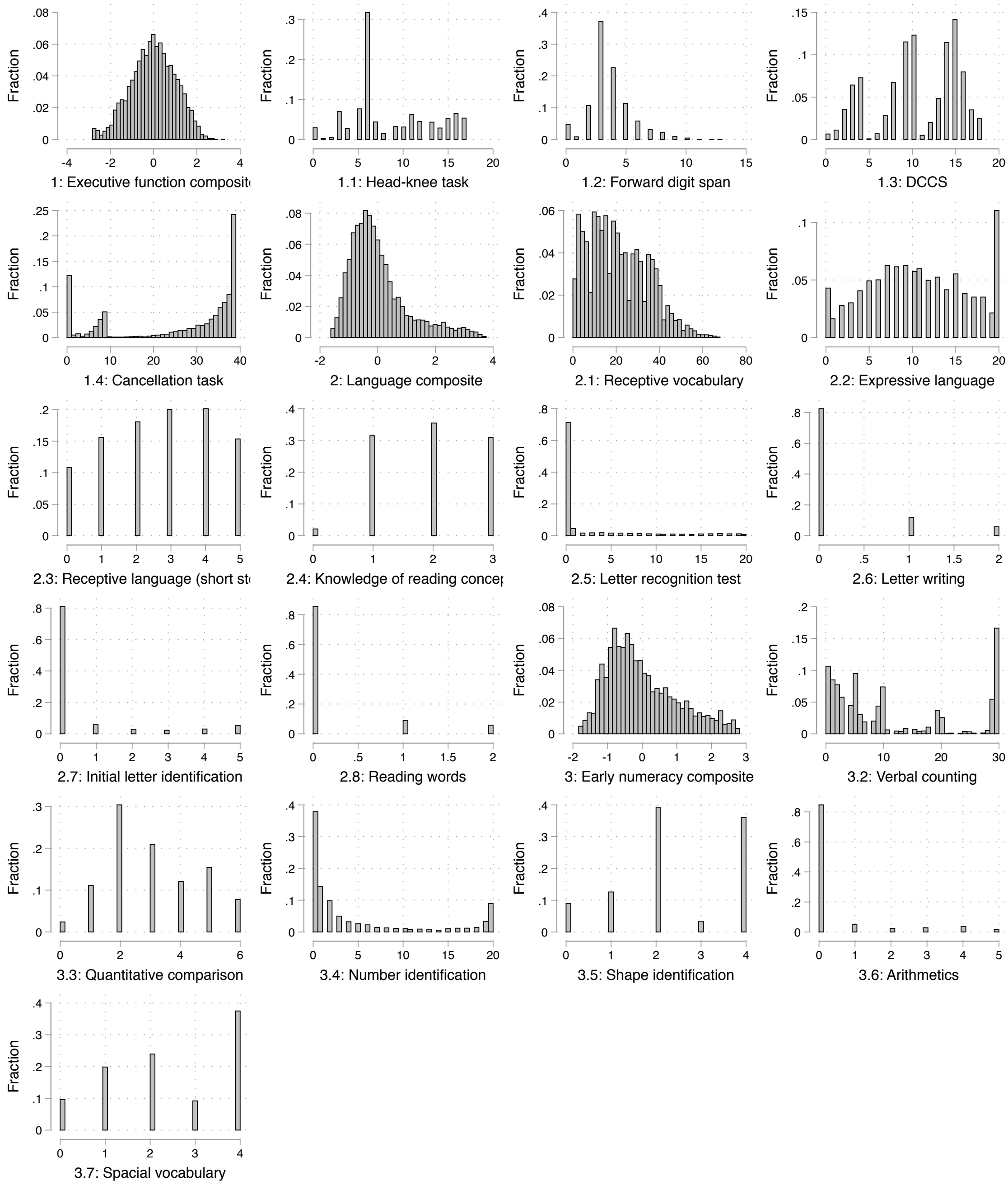


Figure shows empirical distribution of composite and raw endline cognition test scores.

Figure 4.10: Distribution of endline motor-development and socio-emotional composite scores and individual tests.

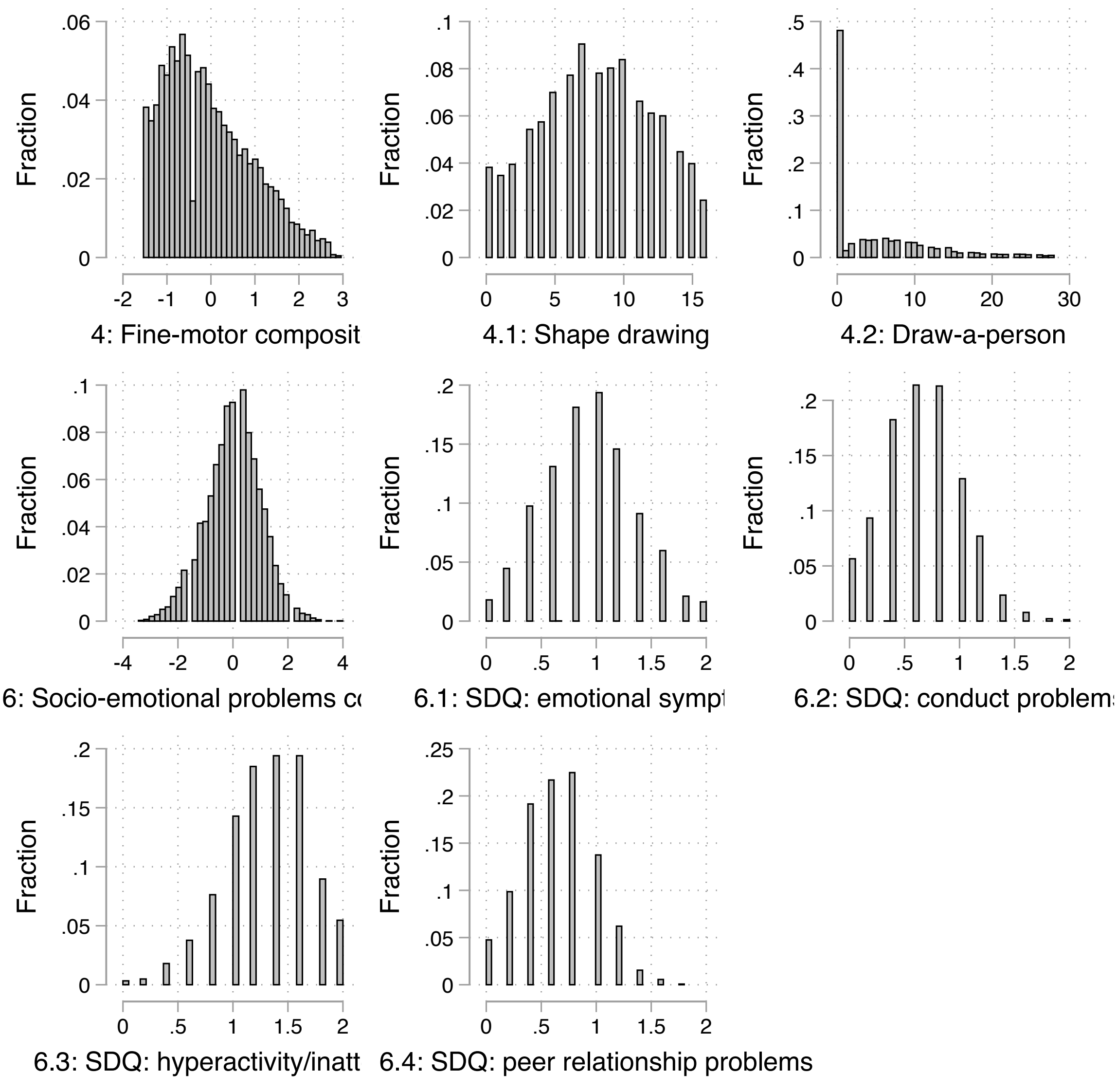


Figure shows empirical distribution of composite and raw motor-development and socio-emotional endline test scores.

CHAPTER 5

Estimating Preschool Impacts when Counterfactual Enrollment Varies: Bounds and Conditional LATE[1]

5.1 Introduction

Development programs, such as large infrastructure plans, new financial institutions, or new technologies, are often introduced in a context where access to similar services already exists. In program evaluation, the presence of these close substitute programs generate a variation in the counterfactual enrollment likely to affect the interpretation of the standard treatment effect parameters – intention-to-treat (ITT) and local average treatment effect (LATE). Specifically in this context, ITT or LATE will cloud the treatment effect differences between individuals who would have benefited from a close substitute program and those who would not (Heckman et al., 2000; Kline and Walters, 2016). While standard treatment effect parameters remain internally valid and relevant estimates of the overall policy impact, clarifying how the effect of a policy depends on close substitutes is critical for producing evidence that is comparable across studies and for making appropriate policy recommendations.

In this paper, we estimate the impact of a preschool program in a Cambodian context, where newly built formal preschools (or community preschools) compete with existing alternative childcare arrangements. In this program – jointly conducted by the World Bank and the Cambodian Ministry of Education, Youth, and Sports – the construction

[1]This chapter is joint work with Adrien Bouguen (University of Santa Clara). The authors thank the World Bank, in particularly Deon Filmer, Tsuyoshi Fukao, Simeth Beng and Samuel Fishman for their support; the SIEF team and Alaka Holla for their financial support; the Ministry of Education in Cambodia, and specifically Sok Sokhom and Lynn Dudley, who made this research possible; and Angkor Research and John Nicewinter, Ian Ramage, Benjamin Lamberet, and Kimhorth Keo, for the stellar fieldwork. Many researchers also contributed to this research through their very useful comments: Craig McIntosh, Patrick Kline, Christopher Walters, Karen Macours, Diego Vera, Markus Frölich, Paul Gertler, Supreet Kaur, Edward Miguel, Katja Kaufmann, Elisabeth Sadoulet, Alain de Janvry, Clément de Chaisemartin, Antoine Camous, Harald Fadinger, Luc Behaghel and Marc Gurgand.

of the community preschools was randomly assigned among villages with alternative forms of preschool (or alternative preschools). The study, therefore, creates a variation in the counterfactual enrollment and, as a result, two sub-populations of compliers: children who would have stayed in home care in the absence of the program (or home compliers) and children who would have attended alternative preschools (or alternative compliers). Consequently, the ITT effects and traditional LATE presented in this paper measure the effectiveness of the new community preschools in comparison to a mix of home care and alternative preschools. In this paper, we propose to go beyond standard treatment effect parameters and develop strategies to isolate the specific contribution of the home compliers and the alternative compliers. Specifically, we will suggest that isolating the effect on home compliers is of prime importance for the early childcare development (ECD) literature.

The presence of close substitute programs is not a unique characteristic of our study. To our knowledge, every large-scale randomized controlled trial conducted to measure preschool effects in a low-income country is implemented in an environment where alternative care arrangements are present. For instance, in a previous preschool experiment conducted in Cambodia from 2008 to 2010, Bouguen et al. (2018a) find that 11% of the control group attended a preschool. Similarly, 8% of a control group in Mozambique attend preschool (Martinez et al., 2017a), 16% do so in The Gambia (Blimpo et al., 2019), and 15% in Indonesia (Brinkman et al., 2017). In the US, 40% of families that lost a lottery to enroll in Head Start ultimately benefited from a close substitute program (Puma et al., 2012). The fact that all of these articles present different degrees of substitution, along with the fact that the quality of alternative childcare programs is often unknown, makes it impossible to draw general conclusions regarding the effectiveness of preschool interventions. Consequently, while Martinez et al. (2017a) find strong effects of preschool attendance on child outcomes, Bouguen et al. (2018a), Bouguen et al. (2013), Blimpo et al. (2019), and Brinkman et al. (2017) in low-income countries, and Puma et al. (2012) in the US, find no effects or only small effects. We interpret this lack of consistency in the literature, at least partially, as a result of the specific substitution patterns that affect every preschool study.

The contribution of this paper, therefore, is twofold. First, we provide an important contribution to the reduced-form literature on preschool impact by using a large and well-implemented preschool construction program conducted in Cambodia between 2015 and 2018. In this paper, we present results from one of the treatment branches after one year of implementation. Berkes et al. (2019b) provide a more comprehensive description of the program's impacts, including results on other treatment branches and two years after construction began. Second, using new empirical strategies and detailed information about the alternative forms of preschool available to parents, we isolate the impact of the program on home compliers from the impact on alternative compliers. We argue that the impact on home compliers is a critical parameter in the (ECD) literature and that the failure to isolate both sub-treatment effect parameters contributes to the ongoing confusion in the debate about the effectiveness of ECD in low-income countries.

We start our analysis by looking at the reduced-form estimates. One year after the construction of the community preschools (CPS), the ITT effect on three- to five-year-old children varies from 0.046 to 0.061 standard deviations (SD) on a large set of child development measures (executive function, language, numeracy, fine-motor, and socio-emotional development). With the exception of socio-emotional development, all effects are statistically significant. While small in magnitude, the overall ITT effect is driven by a 39 percentage point (pp) increase in CPS enrollment and an 11 pp increase in overall school enrollment. A larger ITT effect among five year old kids, from 0.074 SD (executive function) to 0.169 SD (numeracy), is driven by a larger take-up rate of 47 pp. Our results show that a well-implemented, at scale, ECD program, conducted by Cambodian teachers, Cambodian trainers, and piloted by the National Ministry of Education in Cambodia, is able to significantly improve cognitive and non-cognitive child outcomes, at least in the short run.

We then document a large degree of program substitution, using detailed information about the alternative forms of preschool available to parents. In the absence of the construction program, many children would have enrolled in other preschool programs. Hence, the reduced-form effects reflect both the treatment effect of CPS attendance on children who would have stayed at home – the effect on home compliers – but also the effect from enrolling at CPS instead of enrolling in another existing preschool or alternative preschool (APS) – the effect on alternative compliers.

To further investigate the underlying causal effects of the preschool program, we first show that the treatment effects are larger when the predicted share of home compliers increases. This finding suggests that treatment effects are indeed larger for home compliers than for alternative compliers and that the presence of alternative preschools generates an unobserved and, yet fundamental, form of treatment heterogeneity – the *heterogeneity of counterfactual enrollment*.[2] We then show that, under plausible assumptions, the effect on home compliers can be bounded between the traditional local

[2]Note that the heterogeneity of counterfactual enrollment is distinct from the concept of essential heterogeneity introduced by Heckman et al. (2006). The heterogeneity of counterfactual enrollment comes from the enrollment behavior that the individual would have had in absence of the program while essential heterogeneity comes from the self-selection of participants based on the partial knowledge of their response to the treatment (sorting on the gain).

average treatment effect (LATE) of going to CPS and the LATE of going to any preschool. With these bounds, the effect on home compliers, who attended the new CPS program for about 10 months, varies between 0.13 SD and 0.45 SD (on our cognitive index) or between 0.14 SD and 0.39 SD when we use additional baseline variables to narrow our bounds. Finally, we use an empirical technique previously applied elsewhere, the conditional LATE (Kline and Walters, 2016; Hull, 2018), to obtain point estimates of the effects on home and alternative compliers. We find consistent evidence that the effect on home compliers is around 0.2 SD on a child development aggregate score – between 0.16 SD and 0.32 , depending on the set of instruments used to isolate the conditional LATE. The effect on alternative compliers is positive but smaller and indistinguishable from zero.

Our article directly relates to the strand of applied literature that discusses the interpretation of treatment parameters in the presence of close substitutes (Heckman et al., 2000; Feller et al., 2016; Kline and Walters, 2016; Hull, 2018; Kirkeboen et al., 2016). As described by Kline and Walters (2016), in the preschool context, the local average treatment effect is a weighted average of the effects on home and alternative compliers. Yet, these *sub-LATEs* parameters cannot directly be derived, as the counterfactual care arrangement is not observed for individual children in the treatment group.

Depending on the objectives of the researcher, the identification of sub-LATEs might not be of prime concern. As noted by Kline and Walters (2016), program substitution can even be seen as an opportunity when estimating the cost-effectiveness of a similar policy. When the substitution patterns replicate those that would have been found in an ecological environment, then ITT and the standard LATE are the policy relevant parameters. Failure to isolate sub-LATEs and, in particular, failure to isolate the effect on home compliers is, nevertheless, an important limitation. First, the justification for ECD interventions relies heavily on the idea that formalized ECD programs should compensate unfavorable early environments at home (Cunha et al., 2010; Heckman, 2010; Cunha and Heckman, 2007; Cunha et al., 2013). This idea prevails in the United States (Campbell et al., 2002; Currie, 2001; Heckman et al., 2010) and in low-income countries (Gertler et al., 2014; Walker et al., 2011). Preschool interventions, nutrition supplementation, and cognitive stimulation programs for children aged 0–6 are usually seen as ways to compensate for detrimental factors in the home environment. Failure to isolate the benefit of preschool versus home environment is very detrimental to the ECD literature.

Second, many influential empirical papers in the early childcare literature implicitly report the impacts on the home compliers. The Jamaica study (Grantham-McGregor et al., 1991) in low-income countries and the Perry Preschool Project (Anderson, 2008) in the US, which constitute the empirical foundation for new ECD interventions, implicitly measure effects on home compliers. Comparing more recent at-scale programs with these studies on the basis of reduced-form estimates is inappropriate if children in the control group have access to close substitutes.[3] More generally, since the magnitude of standard treatment parameters crucially depends on local conditions – including rate of substitution and substitute programs' quality – the ITT and LATE are likely to be systematically incomparable across contexts. Instead, the effect on home compliers does not depend on close substitutes and comparability can be assessed using commonly available socioeconomic characteristics, e.g. parental education, poverty, and stunting rates.[4] Any aggregate meta-statistic about the effectiveness of ECD interventions that does not take substitution into account is of limited value. With the increased concern around reproducibility and the revived interest around meta-analysis (Meager, 2018), we believe this is a crucial limitation.

Third, in order to make appropriate policy recommendations, understanding the expected substitution patterns and isolating the sub-LATEs is complementary to a reduced-form analysis. If the share of home compliers is small, or if the effect on alternative compliers is null or negative, then large treatment effects on home compliers are entirely consistent with, for instance, low and insignificant ITT effects. The sub-LATE analysis informs policymakers that the same program, targeting, for instance, home compliers, could generate substantial impacts. It could further mean that additional demand-side interventions (information, cash transfers, nudges, free lunch, free transportation) should be implemented to attract those children who would benefit most from the program.

This paper proceeds as follows. First, we describe institutional details, the experimental design, and the studied sample. Second, we present the empirical framework that is used to analyze the data. We focus on the relationship between ITT, LATE, and sub-LATE parameters. Third, we present our reduced form estimates: the adherence to the experimental protocol, the preschool participation, and the impact on children's performance after one year of preschool. In the fourth section, we present our estimations of the treatment effect on home compliers. We discuss the validity of our bounds and then provide one alternative strategies (conditional LATE) to point estimate the sub-LATEs.

[3] In fact, when the counterfactual care arrangement is a close substitute for a majority of compliers, the study might be more comparable to quality interventions, such as the study by Ozler et al. (2018), which evaluates the impact of preschool quality improvement on child performance in Malawi and who implicitly measure an effect on alternative compliers.

[4] Similarly, although less easily observed, characteristics of the close substitute program can be used to assess the comparability of the $LATE_{ac}$.

5.2 Background, Data, and Design

5.2.1 Recent ECD Program Development in Cambodia

Despite robust economic growth since 2000, Cambodia remains one of the least developed countries in Southeast Asia, with a GDP per capita estimated at $1,384 in 2017 ($4,000 in PPP terms). The country also faces multiple challenges in the education sector. With a preschool enrollment rate in 2009 of 40% among five-year-olds (MoEYS, 2014), the country fares poorly in comparison to its neighbors, Thailand and Vietnam.[5] To increase the capacities and quality of its education system, the government of Cambodia, with the support of the World Bank, launched an education expansion program for the 2014—2018 period, called the Global Partnership for Education II (GPE II). Berkes et al. (2019b) provides further details about previous education expansion programs (GPE I).[6] GPE II, and the education sector in Cambodia. This paper focuses on the part of the GPE II that includes the construction of community preschools (CPS).

5.2.2 Formal Community and Alternative Preschool Programs

Before GPE II, two distinct types of public preschools existed in Cambodia: state preschools (SPS) and (informal) community preschools. Since community preschools lacked uniform quality standards, we refer this type of preschool as informal (community) preschools (IPS). In this article, we consider both IPS and SPS as alternative preschools (APS).[7,8] GPE II introduced a new type of community preschool with a uniform quality standard, which we refer to as (formal) community preschool (CPS).

State preschools are financed by the Ministry of Education, Youth, and Sports (MoEYS) (see Figure 5.1 for pictures of a typical SPS facility). SPS teachers benefit from two years of formal training in a MoEYS teacher training center in Phnom Penh. They receive a monthly salary of about 180 $ in 2017 to teach for three hours a day, five days a week. As almost all SPS are attached to a public primary school, SPS have access to properly equipped classrooms, as well as teaching materials, play materials, and sanitary facilities.

In contrast, informal community preschools are not typically attached to a primary school. Local communities establish IPSs and cover operational costs. This includes

[5] Source: Data from UNESCO Institute for Statistics.

[6] Bouguen et al. (2018a) analyzed the impact of the preschool construction funded by GPE I.

[7] According to government data (MoEYS, 2017), out of 7,241 preschool facilities in Cambodia in 2016, 55% were SPS, 39% were IPS, and 6% were private preschools. However, these preschools are not evenly distributed across the country and 38% of the 1646 communes in Cambodia had no preschool facility.

[8] See Bouguen et al. (2013) for an impact evaluation of each type of preschool developed in the wake of GPE I.

the IPS teacher salary, which is at the discretion of the local commune council. It varies from $30 to $50 per month, with most IPS teachers relying on additional sources of income. IPS teachers are trained for about 35 days by provincial education departments before they begin working. Teachers are required to provide a 2-hour preschool class, five days a week. The quality of IPS can differ substantially across villages as, until 2018, communes were required to establish IPSs using their own funds. Consequently, IPS classes are often held in a teacher's home, in a community hall, or a pagoda (see Figure 5.2). IPSs often lack appropriate equipment, such as teaching and play materials or sanitary facilities. In most cases, IPSs even lack the most rudimentary equipment, such as tables and chairs.

To increase preschool access and to improve the unsatisfactory quality of IPSs, the Cambodian government agreed to use the GPE II grant to establish 500 new formal community preschools. Some of these CPS replaced existing informal arrangements; others were established in villages that had no previous preschool or were too large to be serviced by one preschool alone. Unlike IPSs, a CPS benefits from uniform quality standards, such as a standardized building (see Figure 5.3), directly financed by the GPE II. CPS have a capacity of 25 children and are fully equipped with tables, chairs, a blackboard, and teaching materials. In partnership with GPE representatives, MoEYS is responsible for the curriculum, teacher recruitment, and teacher training, as well as the monitoring of the running facility, including regular payment of teacher salaries. The CPS teacher is usually a (female) community member who receives training from the ministry and gives a two-hour class each day, five days a week, to children aged three to five years.

5.2.3 Randomization and Data

The evaluation of the CPS program is based on a cluster randomized controlled trial.[9] All sample villages are situated in the south and northeast parts of Cambodia, as the western part of the country had already been covered by previous expansion plan (see map in Figure 5.4). Eligibility criteria for villages to participate in the study were demand for a CPS, a high poverty rate, and a high number of children between the ages of 0 and 5.

The total study sample comprises 305 villages. Before baseline, we randomly assigned these villages to different treatment branches: a control group (58 villages), which received no GPE II intervention; and a CPS treatment group (120 villages), which received a CPS. An additional 127 villages received a CPS plus a demand-

[9]The study was pre-registered at the AEA's Social Science Registry (AEARCTR-0001045).

side intervention.[10] These demand-side interventions were in part implemented during follow-up data collection in 2017 and, hence, their impact is evaluated on the basis of a follow-up in 2018; this is the focus of a separate article (Berkes et al., 2019b).

Table 5.1 provides an overview of data collection activities and the timing of preschool construction. The analyses presented in this paper are based on two main waves of data collection: a baseline data collection in 2016 and an initial follow-up in 2017. Additionally, a brief monitoring survey was conducted in late 2016 to confirm that preschool construction proceeded as scheduled. With 86% of CPS constructed before follow-up, Table 5.1 confirms that the construction plan was almost perfectly respected. Yet, despite our effort to conjointly deploy the preschool construction and baseline survey, in 17% of the treatment group villages, the CPS was already available at baseline. Conducting a social experiment on school construction is challenging, since conducting baseline too early (before any construction) would have increased the risk, in case of construction delay, that our baseline sampled children would have been too old to attend the newly built preschools.[11] Inversely, conducting the baseline too late would have resulted in baseline measures that are already affected by the program. In Section 5.2.4, we discuss the implications of the slight overlap between the baseline survey and construction.

During the baseline data collection exercise in 2016, our survey firm sampled up to 26 eligible households per village.[12] Eligible households are composed of at least one child between 24 and 59 months old at baseline. Thus, eligible children were between three to five years old at follow-up.

Our survey instruments include a village, teacher, household, and caregiver survey, as well as a child assessment.[13] The village and teacher surveys serve as sources of information about village and preschool infrastructure. The household survey captures

[10] We randomly assigned the remaining 127 villages to two variants of the demand-side interventions (an awareness campaign or an awareness campaign plus a parenting program) to stimulate preschool enrollment. We performed the randomization with province-level stratification on a list of 310 eligible villages provided by MoEYS. Of these, 60 were assigned to the control group, 123 to T1, 63 to T2, and 64 to T3. Unfortunately, the list contained erroneous village names, with 5 either duplicated or unable to be identified following randomization. Therefore, the total number of villages decreased to 305. We treated this drop-out as random and did not replace the villages.

[11] As described in Bouguen et al. (2013), construction delays occurred in a previously evaluated program in Cambodia, which considerably reduced take-up and statistical precision.

[12] They used an adapted version of the EPI walk to sample the household. EPI refers to the Expanded Programme on Immunization of the World Health Organization; see e.g. Henderson and Sundaresan (1982).

[13] The caregiver is defined as the direct relative (parent, grandparent, aunt/uncle, or adult sibling) who takes care of the child most of the time. In most cases, the caregiver is a biological parent (60.4% at baseline, 58.7% at follow-up). In the provinces of Kampong Speu, Kandal, Prey Veng, Svay Rieng, and Takeo, the caregiver is often a grandparent. These are provinces with relatively high levels of manufacturing and mothers are frequently absent during the day.

information about household wealth, income, and other socioeconomic measures. The caregiver survey is used to obtain information about parenting practices, a fluid intelligence measure of the caregiver (based on Raven's Progressive Matrices), and detailed information about the child (for example, preschool enrollment history). Parental-reported versions of the Strengths and Difficulties Questionnaire (SDQ) and the social development scale of the Malawi Development Assessment Tool (MDAT) were administered to caregivers to obtain a measure of socio-emotional development of the children. Additionally, a comprehensive child assessment was conducted. The battery of child tests measure five crucial domains of cognitive and physical child development: executive function, language, numeracy, as well as both fine- and gross-motor development.[14] Most child tests stem from the Measuring Early Learning Quality and Outcomes project (MELQO). MELQO tools are designed to provide a starting point for national-level adaptation of global measures of child development (see UNESCO (2017) for an overview) and demonstrate adequate internal validity (Fernald et al., 2017; Berkes et al., 2019).[15] Additionally, anthropometric measurements (height and weight) are taken from all tested children.

5.2.4 Sample Description and Cognitive Inequality

A summary of the study sample is presented in Table 5.2. The baseline sample includes 4075 households and 4393 children aged between 2 and 4 in 178 villages.[16] For 4315 out of 4393 children, consent to participate in the child assessments was obtained from the caregivers and children.[17] Table 5.2 also gives an overview of the households interviewed at the follow-up in 2017. The attrition rate, 8.8% for household attrition, can almost entirely be explained by seasonal or permanent relocation of households, since the study does not follow up on households that move beyond the boundaries of the sample villages. Attrition is slightly larger for children (10.6%) but the difference of attrition between the treatment and control group remains small (1.9%) and insignificant.

Table 5.3 (household and caregiver characteristics) and Table 5.4 (child characteristics) show a balance in variables between treatment and control group separately for the baseline sample and the sample of households who participated in baseline and follow-up. The tables show that the variables are balanced at baseline and remain bal-

[14] We discuss cultural adaptation, content, and scoring of all child test scores and the parental practices measures at length in Berkes et al. (2019).

[15] Our version of the test is available upon request.

[16] Unless otherwise stated, all numbers in this paper refer to the sample of 178 villages without the additional treatment groups. On the full sample of 305 villages, the sample includes 7053 eligible households and 7546 children between 2 and 4 years of age.

[17] The 78 eligible children without baseline test scores are balanced across treatment and control (2% versus 1.67%).

anced after taking attrition into account (*Baseline and Follow-up Sample* panel). One exception is preschool enrollment caused by the slight overlap of preschool construction and timing of the baseline survey (cf. Table 5.1). Treatment children were 6.6 pp more likely to be enrolled in preschool (last panel, Table 5.4). As discussed before and indicated in Table 5.1, the difference is due to the fact that in 17% of treatment villages, the CPS was completed briefly before the baseline survey. Since the treatment children only spent 11 more days in preschool than the control children and since we do not measure any developmental difference between treatment and control at baseline, we consider the difference as negligible.

Table 5.3 shows variables that characterize the socioeconomic background of our sample population. Households are generally poor – 41% are considered as poor, according to our multidimensional poverty index.[18] In our sample, 55% of households live on less than $100 per month. Caregivers, on average, have years of formal education. Based on WHO Child Growth Standards, 34% of tested children are stunted and 10% suffer from wasting.

Child test scores are strongly associated with socioeconomic background characteristics. As described more at length in Berkes et al. (2019), children in the top wealth quintile perform, on average, between 0.46 and 0.7 SD better than children in the bottom quintile. Schady et al. (2015) find similar results in South America. The gap that separates children age 3–5 from the bottom quintile and the top quintile corresponds to about 6–12 months of cognitive development. Thus, wealthier children are up to one year ahead of poorer children in development once they reach primary school age.

5.2.5 Preschool Quality

We use village survey data to show differences in quality measures between the types of preschools at baseline (Table 5.5) and follow-up (Table 5.6).[19] Table 5.5 documents that SPS are significantly different from CPS and IPS: SPS are larger (6 additional children when compared to an IPS, which serves around 21 students) and they have more equipment, such as chairs, tables, and blackboards. SPS also have fewer significant problems, as reported by the village chief. SPS teachers benefited from more training days (+152 days, or about three times as much) and they are more likely to be paid

[18] We construct a binary poverty index using baseline data and an adapted version of the method by (Alkire and Santos, 2010). A household is considered poor if it is deprived in at least 30 percent of the weighted indicators for health, education, and living standards.

[19] Note that the full study sample of all 305 villages is used in these tables to maximize statistical power. Since CPSs were also constructed in the two other treatment branches and since both SPS and IPS are present, they can be used to document preschool quality. Further, note that, as shown in Table 5.1, only a handful of CPS were already open at baseline, while almost all CPS were completed at follow-up. Hence, Table 5.6 is better suited to assess the final quality of CPS.

regularly with a significantly higher salary (on average, \$90 per month versus \$35 for IPS teachers). Already at baseline, the quality of CPSs appears better than that of IPSs: CPSs have better, more spacious, buildings and enjoy more resources. In addition, their teachers were also paid more regularly. Yet village chiefs considered CPS and IPS teachers as comparable in terms of salary and training.

At follow-up (Table 5.6), SPSs still offered a higher quality than CPSs and IPSs, but CPSs quality had further increased. CPS buildings are still reported to be larger and of better quality than IPSs, but this time, CPSs are reported to have more tables, chairs, and additional learning materials. Indeed, at the time of the follow-up survey, almost all CPS equipment had been delivered. Yet, again, in terms of teacher quality, the difference in IPSss and CPS is small, at least in the eyes of the village chief. Teachers in IPSs and CPSs seem to have seen their situations improve in similar fashions: preschool teachers are more regularly and better paid at follow-up than they were at baseline. Additional information on the difference between IPSs, SPSs, and CPSs, relying on in-class observations and additional follow-up surveys, are available in (Berkes et al., 2019). While (Berkes et al., 2019) confirm that IPS and CPS teachers share many characteristics (age, gender, education, experience), the equipment, the class setting (time hours effectively teaching), as well as the educational content (following curriculum...) is reported to be significantly superior in CPSs than in IPSs. SPSs remain superior to CPSs and IPSs across all measured dimensions. In all, Tables 5.5 and 5.6 indicate that the program has significantly improved the infrastructure quality of the community preschools: CPSs have more materials and better premises. Yet the teaching quality – arguably the most important factor in early childhood development – remains comparable in the eyes of village chiefs.

5.3 Empirical Framework

As explained in the previous section, we evaluate the impact of the CPS in the context where alternative preschools (APS), i.e. SPS and IPS, are also available. In this section, we outline the empirical framework and list the strategies we implement to identify the relevant treatment parameters. We extensively discuss the identification strategies and their assumptions in the Appendix.

5.3.1 Identifying Effects of CPS using ITT and LATE

To estimate the intention-to-treat effect (ITT) and the local average treatment effect (LATE) of attending CPS, here called $LATE_{cps}$, we define Z_i to be the instrumental variable that takes the value 1 for children in treatment villages and 0 otherwise.

Further, we define D the participation variable that takes value c, a or h depending on whether the child is enrolled in CPS (c), in APS (a) or is staying at school (h). As described in the Appendix, the ITT is valid under Assumption A1 (independence) and A2 (SUTVA), which are very likely to be valid in our setting. We estimate ITT effects using the following regression:

$$Y_{iv} = \beta_0^{ITT} + \beta_1^{ITT} Z_v + \boldsymbol{W_{iv}}\boldsymbol{\beta_2^{ITT}} + \mu_v^{ITT} + \epsilon_{iv}^{ITT} \tag{5.1}$$

where Y_{iv} denotes the *observed* outcome of child i and village v, Z_v the *observed* treatment assignment, and $\boldsymbol{W_{iv}}$ a set of control variables. μ_v and ε_{iv} are the village-specific error term component and the unobserved within-village error component, both assumed to be uncorrelated with $\boldsymbol{W}$ and Z. We use standard errors clustered at the village level to account for the randomization implemented at the village level. The outcomes of interest Y, specified in the pre-analysis plan (Berkes et al., 2017), include (i) the school construction collected at the village level (see supra Section 5.4.1.1); (ii) the enrollment in, and months of exposure to, each childcare arrangement (see supra Section 5.4.1.2); and (iii) the children's cognitive and socio-emotional performance as well as the parental response to the school construction (see supra Section 5.4.1.3).

We estimate equation (5.1) using different sets of control variables, always including province fixed effects due to stratification at the province level. In a second specification, we include all children baseline test scores, as well as the age and the gender of each child.[20] Since controlling for $\boldsymbol{W}$ considerably reduces the residual variation and was pre-announced in our pre-analysis plan, this specification is our preferred estimation. In a third specification we implement a double LASSO procedure following Belloni et al. (2013) and Chernozhukov et al. (2018). We start by selecting 196 covariates available at baseline and create a set of dummy indicators for categorical variables, and create the square of each covariate, for a total of 450 variables. Finally, we suppress all perfectly collinear variables and use the remaining 265 variables for the double LASSO estimation. As we show below, the double LASSO estimation slightly improves the standard errors but does not change the interpretation of the point estimates.

We estimate the local average treatment effect CPS ($LATE_{cps}$) using following equation:

$$Y_{iv} = \beta_0^{LATE_{cps}} + \beta_1^{LATE_{cps}} \mathbb{1}_{\{D_i=c\}} + \boldsymbol{W_{iv}}\boldsymbol{\beta_2^{LATE_{cps}}} + \mu_v^{LATE_{cps}} + \epsilon_{iv}^{LATE_{cps}} \tag{5.2}$$

[20] We replace missing test scores with the sample mean and interact them with a dichotomous variable indicating a missing value. Age is measured as a trimester fixed effect and also imputed if missing.

where $\mathbb{1}_{\{D_i=c\}}$, which takes value 1 when the child go to CPS (0 otherwise), is instrumented using Z_v. As shown in Appendix 5.C.2, the $LATE_{cps}$ is valid under additional assumption A3 (non-zero average causal effect), A4 (extended exclusion restriction) and A5 (extended monotonicity assumption). The LATE estimation is valid, yet, due to the different counterfactuals in this context, there are implicit assumptions behind this standard model which we, following the framework by Kline and Walters (2016), discuss in Section 5.3.2. The main takeaway is that by defining $D_i \in \{c, a, h\}$ as capturing enrollment into CPS, into APS, or as not being enrolled in preschool (home care), two different types of compliers and never-takers exist:

1. a-never takers (ANT): $D_i(0) = a$, $D_i(1) = a$,
2. h-never takers (HNT): $D_i(0) = h$, $D_i(1) = h$,
3. a-compliers (AC): $D_i(0) = a$, $D_i(1) = c$,
4. h-compliers (HC): $D_i(0) = h$, $D_i(1) = c$,

In Figure 5.5 we provide a visual representation of the different sub-populations in our sample under the stated assumptions. In the absence of the intervention, i.e. the control group scenario (left panel), CPS compliers would have either stayed home or attended an APS. After the CPS construction occurred in the treatment group (right panel), children in the treatment group give rise to the four groups of children, later referred to as principal strata (Feller et al., 2016).

While it is usually assumed that the counterfactual to the participation into the treatment is homogeneous, in our context, $D_i(0)$ can be either a or h. This creates a form of heterogeneity that is not directly observable. Consequently, any traditional heterogeneous treatment analysis would confound the heterogeneity of the (baseline) *observed characteristics* with the *heterogeneity of counterfactual enrollment* if the baseline observed characteristics used to measure the observed heterogeneity correlate with the counterfactual enrollment behavior. In the following, we describe how to take close substitutes better into account.

5.3.2 Substitution, sub-LATEs and Principal Strata

Under the defined assumptions (A1–A5), the local average treatment effect of going to CPS, here called $LATE_{cps}$, is identified and given by:

$$LATE_{cps} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i = c|Z_i = 1] - E[D_i = c|Z_i = 0]}.$$

In presence of a close substitute, the $LATE_{cps}$ can be further decomposed into two sub-LATEs (Kline and Walters, 2016):

$$LATE_{cps} = S_{ac}LATE_{ac} + (1 - S_{ac})LATE_{hc} \tag{5.3}$$

where $LATE_{ac} \equiv E[Y_i(c) - Y_i(a)|D_i(1) = c, D_i(0) = a]$ and $LATE_{hc} \equiv E[Y_i(c) - Y_i(h)|D_i(1) = c, D_i(0) = h]$ give the average treatment effect on a and h compliers. Importantly, S_{ac}, the share of a-compliers (within the group of compliers), is identified and given by Kline and Walters (2016):

$$S_{ac} = \frac{P(D = a|Z = 0) - P(D = a|Z = 1)}{P(D = c|Z = 1) - P(D = c|Z = 0)} \tag{5.4}$$

Figure 5.5 provides a visual representation of the parameters in Equation (5.3): the share of a-compliers is visually represented by the a-compliers region divided by the region occupied by any compliers, and $LATE_{cps}$ is a weighted average of both sub-LATEs, $LATE_{ac}$ and $LATE_{hc}$.

Finally, we provide in Figure 5.6 a modification of the notation in order to better describe and compare the principal strata. Let Y_p^z be the expected outcome variable (typically here the performance of the child at a development test) in experimental branch z (1 for treatment, 0 for control) and for principal strata p. p takes the four values: a-compliers (ac), h-compliers (hc), a-never-takers (ant), and h-never-takers (hnt). Hence:

$$\begin{aligned}
Y_{ac}^z &= E[Y(z)|D(1) = c, D(0) = a] \\
Y_{hc}^z &= E[Y(z)|D(1) = a, D(0) = h] \\
Y_{ant}^z &= E[Y(z)|D(1) = a, D(0) = a] \\
Y_{hnt}^z &= E[Y(z)|D(1) = h, D(0) = h]
\end{aligned}$$

To summarize, Table 5.8 maps the different principal strata for each value of Z and of D as well as their expected value notation under A5. Using this notation and Figure 7, each sub-LATE can be written:

$$\begin{aligned}
LATE_{ac} &= Y_{ac}^1 - Y_{ac}^0 \\
LATE_{hc} &= Y_{hc}^1 - Y_{hc}^0 \\
LATE_{ant} &= Y_{ant}^1 - Y_{ant}^0 = 0 \\
LATE_{hnt} &= Y_{hnt}^1 - Y_{hnt}^0 = 0
\end{aligned} \tag{5.5}$$

again with the effects on non-compliers equals to zero. Similar to Feller et al. (2016), we also introduce A_d^z, the weighted average of Y_p^z for the children staying at home in the control, going to APS in the control or going to CPS in the treatment:

$$\begin{aligned} A_c^1 &= S_{ac} * Y_{ac}^1 + (1 - S_{ac}) * Y_{hc}^1 \\ A_a^0 &= S_{ant} * Y_{ant}^0 + (1 - S_{ant}) * Y_{ac}^0 \\ A_h^0 &= S_{hnt} * Y_{hnt}^0 + (1 - S_{hnt}) * Y_{hc}^0 \end{aligned} \tag{5.6}$$

with S_{hnt} (resp. S_{ant}) the share of h-never-takers (a-never-takers) defined analogous to (5.4).[21] We use (5.6) to describe principal strata at baseline.[22] At baseline, all parameters are identified and specifically both groups of compliers:

$$\begin{aligned} Y_{hc}^0 = Y_{hc}^1 &= (A_h^0 - S_{hnt} * Y_{hnt}^0)/(1 - S_{hnt}) \\ Y_{ac}^0 = Y_{hc}^1 &= (A_a^0 - S_{ant} * Y_{ant}^0)/(1 - S_{ant}) \end{aligned}$$

5.3.3 Bounding $LATE_{hc}$

Equation (5.3) makes explicit the challenges faced by researchers when estimating the impact of a policy in a context of close substitutes. Under A1–A5 (see Appendix), $LATE_{cps}$ is identified, but its sub-LATE components ($LATE_{ac}$ and $LATE_{hc}$) are not. Yet, under plausible assumptions, we can derived bounds for the $LATE_{hc}$ a parameter of particular interest for the preschool literature. The assumption to derive the bounds is:

$$0 \leq LATE_{ac} \leq LATE_{hc} \tag{5.7}$$

i.e., the CPSs offer, on average, a better learning environment than APSs (left hand side of the inequality) and that the returns to CPS are not higher for a-compliers (right hand side of the inequality).

The left hand side of inequality 5.7, $0 \leq LATE_{ac}$, simply implies that switching from APS to CPS is not, on average, detrimental to a-compliers, i.e. using the principal strata notation $Y_{ac}^1 \geq Y_{ac}^0$. Given the resources devoted to CPS in comparison with IPS, as discussed in Section 5.2.5, we believe that the left hand side of 5.7 inequality is a very likely assumption.

[21] $S_{hnt} = \frac{P(D=h|Z=0)-P(D=h|Z=1)}{P(D=h|Z=1)-P(D=h|Z=0)}$ and $S_{ant} = \frac{P(D=a|Z=0)-P(D=a|Z=1)}{P(D=a|Z=1)-P(D=a|Z=0)}$.

[22] In principle (5.6) can be used to describe strata at follow-up as well

The right side of inequality (5.7), $LATE_{ac} \leq LATE_{hc}$, implies that a-compliers do not benefit more from the CPS than the h-compliers, i.e. using principal strata notation $Y^1_{ac} - Y^0_{ac} \leq Y^1_{hc} - Y^0_{hc}$. There are reasons to believe that this is also a likely assumption. Intuitively, h-compliers benefit from a larger improvement of their learning environment than a-compliers, who benefit from some preschool intervention regardless of their treatment status. Consequently, the h-compliers will likely benefit more from CPS than will the a-compliers.

There is no guarantee though. For instance, the CPS teaching content may be more specific to the needs of the a-compliers or the home environment of the h-compliers may be particularly favorable, making CPS returns lower for h-compliers than for a-compliers. While we cannot entirely discard these possibilities, Table 5.9 alleviates some concerns. Using equations (5.6), we compute in Table 5.9 the characteristics of the principal strata at baseline, before anyone had access to any program.[23] While h-never-takers and a-never takers are very different in terms of children and caregivers characteristics (see first three columns), the a and h compliers present no systematic differences (see last column): a and h-compliers have similar parents, similar home environments, and fare similarly in our achievement tests at baseline. Given their similarities at baseline, it would be surprising that the h-compliers would benefit much more from CPS than a-compliers at least enough to compensate the fact that a-compliers benefits from APS in the control group. Thus, table 5.9 provides suggestive evidence that $LATE_{ac} \leq LATE_{hc}$ is a valid assumption.

Under 5.7, we can calculate a lower bound (LB) and upper bound (UB) to $LATE_{hc}$ with:

$$LATE_{ac} = LATE_{hc} \iff LATE^{LB}_{hc} = LATE_{cps} \tag{5.8}$$

The low bound assumes that h-compliers and a-compliers equally benefit, on average, from the CPS intervention . Hence, $LATE_{cps}$ is the local average treatment effect of both sub-populations. Under the upper bound, we assume:

$$\begin{aligned} LATE_{ac} = 0 \iff LATE^{UB}_{hc} &= \frac{LATE_{cps}}{(1 - S_{ac})} \\ &= \frac{\beta^{ITT}_1}{\beta^{FS_c}_1 * (1 - S_{ac})} = \frac{\beta^{ITT}_1}{\beta^{FS_c}_1 * (1 + \frac{\beta^{FS_a}_1}{\beta^{FS_c}_1})} \\ &= \frac{\beta^{ITT}_1}{\beta^{FS_c}_1 + \beta^{FS_a}_1} = \frac{\beta^{ITT}_1}{\beta^{FS_{ps}}_1} \equiv LATE_{ps} \end{aligned} \tag{5.9}$$

[23]Children were aged 2 to 4 at baseline, too young to benefit from any preschool. This is not entirely true, as a very small proportion of compliers (2.54 %) actually benefited from APS.

with $LATE_{cps} = \frac{\beta_1^{ITT}}{\beta_1^{FSc}}$, β_1^{FSc} the CPS first stage parameter described in Appendix equation 5.15, β_1^{FSa} the equivalent first stage parameter for the APS, $\beta_1^{FS_{ps}} = \beta_1^{FSa} + \beta_1^{FSc}$ the first stage parameter that captures the differential *any* preschool take-up, and $LATE_{ps}$ the effect of *any* preschool enrollment instrumented by Z. Essentially, our arguments imply:

$$LATE_{cps} \leq LATE_{hc} \leq LATE_{ps}$$

$LATE_{hc}$ is bounded by $LATE_{cps}$ and $LATE_{ps}$.

5.3.4 Narrowing the Bounds for $LATE_{hc}$

We can narrow the bounds using a set of additional variables orthogonal to Z. We assume:

$$0 \leq LATE_{ac}(\boldsymbol{B}) \leq LATE_{hc}(\boldsymbol{B})$$

which is the equivalent to 5.7 for each value of $\boldsymbol{B}$, the variables orthogonal to Z. We can implement the bounding strategy – calculate $LATE_{cps}$ and $LATE_{ps}$ for each value of $\boldsymbol{B}$ – in the sample cells formed by $\boldsymbol{B}$. We then average across the value of $\boldsymbol{B}$ to recover the unconditional narrow lower and upper bound, using the probability to belong to the cells as weights. The lower bound can also be more flexibly estimated using the following IV regression model:[24]

$$Y_{iv} = \beta_0 + \beta_1^{lb}\mathbb{1}_{\{D_i=c\}} + \beta_2^{lb}\boldsymbol{B}_i + \beta_3\boldsymbol{W}_i + u_{iv}$$

where $\mathbb{1}_{\{D_i=c\}}$ – the dummy for CPS enrollment – is instrumented by $\boldsymbol{B}$, Z, Z*$\boldsymbol{B}$ and controlling for $\boldsymbol{B}$ and $\boldsymbol{W}$. $\boldsymbol{W}$ is a set of control variables not interacted with Z. β_1^{lb} is the parameter of interest that gives the narrow lower bounds. Similarly, the narrow upper bound is estimated using :

$$Y_{iv} = \beta_0 + \beta_1^{ub}\mathbb{1}_{\{D_i=a\ \cup\ D_i=c\}} + \beta_2^{ub}\boldsymbol{B}_i + \beta_3\boldsymbol{W}_i + u_{iv}$$

where $\mathbb{1}_{\{D_i=a\ \cup\ D_i=c\}}$ – the dummy for any preschool enrollment (CPS or APS) – is instrumented by $\boldsymbol{B}$, Z, $\boldsymbol{B} * Z$ and controlling for $\boldsymbol{B}$ and $\boldsymbol{W}$. While any baseline variable can theoretically be included in $\boldsymbol{B}$, the choice of B variables depends on two

[24]To see this, let's $\boldsymbol{B}$ be a dummy variable taking two values. To calculate the narrow lower bound, we jointly estimate an IV regression for B=0 and B=1 and we take the weighted average using the probability to belong to group B=1 and B=0 respectively. Doing so corresponds to an IV regression where $D_i = c$ is instrumented by Z, $\boldsymbol{B}$ and $\boldsymbol{B} * Z$.

potentially conflicting criteria. First, the size of the narrow bounds will depend on the ability of the $\boldsymbol{B}$ variables to predict the enrollment behavior. Second, the $\boldsymbol{B}$ variables should be sufficiently parsimonious to maintain a reasonable sample size in each cell. To balance both criteria, we estimate the narrow bounds using the quintile of the predicted share of alternative compliers $\hat{S}_{ac}(\text{X})$, as defined below in equation (5.11), which allows both a reasonable sample size and a good prediction of the enrollment behavior. We show that using the quintile of the predicted share of alternative compliers significantly narrows the bounds.

5.3.5 Estimating the Share of a-compliers and the Sub-LATE $LATE_{hc}$ using Conditional LATE

Although the bounds rely on a less restrictive set of assumptions, we implement several alternative strategies to identify more precisely the sub-LATE parameters.

Using additional baseline characteristics as instruments and under a constant sub-LATE assumption, we can isolate $LATE_{hc}$ and $LATE_{ac}$. Kline and Walters (2016) – see also Hull (2018) and Feller et al. (2016) – show that sub-LATEs can be identified by interacting the randomly assigned preschool construction program with observed covariates. For example, let $X_i \in \{0, 1\}$ be a dummy variable strongly correlated with APS enrollment but uncorrelated to Z, i.e. any constant or baseline variable strongly correlated with enrollment. For each value of X, the conditional $LATE$ or LATE(X), can be written as:

$$\begin{aligned} LATE_{cps}(1) &= S_{ac}(1)LATE_{ac} + (1 - S_{ac}(1))LATE_{hc} \\ LATE_{cps}(0) &= S_{ac}(0)LATE_{ac} + (1 - S_{ac}(0))LATE_{hc} \end{aligned} \tag{5.10}$$

As we can estimate $LATE_{cps}(X)$ and $S_{ac}(X)$ for each value of X, we can identify the two unknown sub-LATEs by solving the system of equations. As it is apparent in (5.10), the identification relies on the assumption that $LATE_{ac}$ and $LATE_{hc}$ do not depend on X. In other words, the variation of the $LATE_{cps}(X)$ derives entirely from variation in $S_{ac}(X)$, while the sub-LATEs remain constant on X.

Using additional X variables, strongly correlated with the enrollment behavior but uncorrelated with Z, we also define:

$$S_{ac}(\boldsymbol{X}) = \frac{\pi_{ac}(\boldsymbol{X})}{\pi_c(\boldsymbol{X})} \tag{5.11}$$

the predicted share of a-compliers given $\boldsymbol{X}$. $\pi_{ac}(\boldsymbol{X}) = E(\mathbb{1}_{\{D_i=a\}}|Z = 0, \boldsymbol{X}) - E(\mathbb{1}_{\{D_i=a\}}|Z = 1, \boldsymbol{X})$[25] represents the predicted probability of being a-compliers and $\pi_c = \pi_{hc} + \pi_{ac} = E(\mathbb{1}_{\{D_i=c\}}|Z = 1, \boldsymbol{X}) - E(\mathbb{1}_{\{D_i=c\}}|Z = 0, \boldsymbol{X})$ the predicted probability of being a complier (hc or ac), with $\mathbb{1}_{\{D_i=c\}}$ the enrollment in CPS and $\mathbb{1}_{\{D_i=a\}}$ the enrollment in APS. Both π_{ac} and π_c taking the predicted valued of the following linear models:

$$\begin{aligned} \mathbb{1}_{\{D_i=c\}} &= \gamma_0^c + \gamma_1^c \boldsymbol{X}_i + \gamma_2^c Z_i + \nu_v^c + \epsilon_{iv}^c \\ \mathbb{1}_{\{D_i=a\}} &= \gamma_0^a + \gamma_1^a \boldsymbol{X}_i + \gamma_2^a Z_i + \nu_v^a + \epsilon_{iv}^a \end{aligned} \tag{5.12}$$

We use $\hat{S}_{ac}(\boldsymbol{X})$ in section 5.4.2 to analyze how the treatment changes when the share of a-compliers is modified. Intuitively, under assumption 5.7, we expect the impact to be smaller as the share of a-compliers increases. We also use the quintile of the predicted share of a-compliers to narrow the bounds as explained above (Section 5.3.4). We finally use the the predicted share of a-compliers as instruments in the conditional LATE estimation.

Finally, as mentioned, the $LATE_{ac}$ and $LATE_{hc}$ can be estimated using a 2SLS, with $\mathbb{1}_{\{D_i=c\}}$ (enrollment in CPS) and $\mathbb{1}_{\{D_i=a\}}$ (enrollment in APS) as endogenous variables and Z and $\boldsymbol{X} * Z$ as instrumental variables. The structural equation takes the following form:

$$Y_{iv} = \beta_0 + \beta_1 \ \mathbb{1}_{\{D_i=c\}} + \beta_2 \mathbb{1}_{\{D_i=a\}} + \beta_3 \boldsymbol{X}_i + \beta_3 \boldsymbol{W}_i + u_{iv} \tag{5.13}$$

where Y_i is a follow-up outcome, $\boldsymbol{X}$ a set of additional variables orthogonal to the treatment, and $\boldsymbol{W}$ the preferred set of control variables used for the ITT reduced form estimation (province fixed effect, age, gender and baseline test scores). β_1 captures the $LATE_{hc}$, and β_2 captures the effect of going to APS. To derive the $LATE_{ac}$, we

[25] To see this, note that:

$$E(\mathbb{1}_{\{D_i=a\}}|Z = 0, \boldsymbol{X}) = \pi_{ant}(\boldsymbol{X}) + \pi_{ac}(\boldsymbol{X})$$

where $\pi_{ant}(\boldsymbol{X})$ is the predicted probability of being a a-never-taker. Similarly, the a-never-takers are identified as follow:

$$E(\mathbb{1}_{\{D_i=a\}}|Z = 1, \boldsymbol{X}) = \pi_{ant}(\boldsymbol{X})$$

therefore, $\pi_{ac}(\boldsymbol{X})$ is estimable:

$$\pi_{ac}(\boldsymbol{X}) = E(\mathbb{1}_{\{D_i=a\}}|Z = 0, \boldsymbol{X}) - E(\mathbb{1}_{\{D_i=a\}}|Z = 1, \boldsymbol{X})$$

subtract β_2 from β_1. $\mathbb{1}_{\{D_i=c\}}$ and $\mathbb{1}_{\{D_i=a\}}$ are both endogenous and instrumented by:

$$\begin{aligned}\mathbb{1}_{\{D_i=c\}} &= \alpha_0 + \alpha_1 Z_v + \alpha_2 Z_v * \boldsymbol{X}_i + \alpha_3 \boldsymbol{X}_i + \alpha_4 \boldsymbol{W}_i + \mu_v + \epsilon_{iv} \\ \mathbb{1}_{\{D_i=a\}} &= \gamma_0 + \gamma_1 Z_v + \gamma_2 Z_v * \boldsymbol{X}_i + \gamma_3 \boldsymbol{X}_i + \gamma_4 \boldsymbol{W}_i + \phi_v + \nu_{iv}\end{aligned} \tag{5.14}$$

The identification of β_1 and β_2 relies on the independence of Z and $Z * \boldsymbol{X}$ and on the assumption that the h and a compliers have a constant return to preschool on $\boldsymbol{X}$. Importantly, the constant treatment effect assumption means that among compliers, if the treatment effect is heterogeneous, such heterogeneity should be *between* a- and h-compliers groups and not *within* them. In other words, while heterogeneity among compliers is possible, the treatment heterogeneity must be driven by the heterogeneity of *counterfactual enrollment* rather than by the standard heterogeneity of *observed characteristics* (see infra section 5.3.1). Note that when $\boldsymbol{X}$ is made of more than one instrument, the validity of this assumption can be tested using an over-identification test.

One of the practical estimation issues of the conditional LATE – and $S_{ac}(\boldsymbol{X})$ – is that the set of $\boldsymbol{X}$ variables needs to be sufficiently predictive of the substitution behavior to secure a good first stage. In our context, we include village and infrastructure indicators in a first specification (village level indicators include a dummy for above median village population size, as well as land area, baseline presence of a primary school, baseline presence of a secondary school), and we add individual level characteristics in a second stage (individual level characteristics including baseline caregiver Raven score, caregiver education, household poverty dummy and household poverty score). This is our preferred specification and is therefore used to estimate $S_{ac}(X)$ in equation 5.12. To improve the transparency of our approach and comparability to the narrow bounds, we include in the last specification only the quintile of the predictive share of a-compliers.[26]

5.4 Results

5.4.1 Reduced-Form Estimates

5.4.1.1 School Construction

We begin our empirical analysis by documenting in Table 5.10 how treatment assignment affects preschool availability in the sampled villages. At follow-up, none of the control villages benefited from a CPS, while 86% of the treatment villages did. Given

[26] For the estimation of $S_{ac}(\boldsymbol{X})$, we only present the results using the second specification that includes village and individual level characteristics

the constraint inherent to any construction work in low-income countries and the delays that such programs may incur, we consider this a particularly favorable result.[27] The fact that none of the control group villages received a CPS confirms that the Cambodian government strictly respected the study protocol. Less anticipated was the number of alternative preschools available in both treatment and control villages. As there was at least some kind of preschool in 81% of the control group villages, our treatment increases availability of any preschool by just 12 pp.

Furthermore, IPS availability declines by a significant 56 pp in the treatment group, while the availability of state preschools (SPSs) remains approximately unaffected (insignificant -6 pp). As confirmed by our field visits, IPSs were often shut or turned into CPSs as soon as the new preschool building became available. Conversely, SPSs, already a formalized form of preschool, remained available to the children.[28] Table 5.10 confirms that most of the substitution occurred between IPS and CPS availability and that children enrolling in SPS are essentially unaffected by the program. This substitution pattern has important implications for the interpretation of our results: Since IPSs are arguably of much lower quality than SPSs (see Tables 5.5 and 5.6), the fact that a-compliers would have enrolled in IPS if assigned to the control means that a-compliers are likely to contribute positively to the overall treatment effect. Although in the rest of the analysis, we will still consider the substitution pattern to exist between CPSs and APSs, the reader should keep in mind that the vast majority of the substitution is actually occurring between IPSs and CPSs.

5.4.1.2 Preschool Enrollment

To study the enrollment patterns at the child level, we explore the enrollment and exposure of children, separately by preschool type, in Table 5.11. Assignment to the treatment group significantly affects enrollment by about 39 pp (47 pp for the five-year-olds). Since 14% of treatment group villages did not receive a CPS until follow-up, the CPS take-up rate in villages with CPS is about 45 pp (55 pp for five-year-olds) on average. Such level of (differential) enrollment is in line with our most optimistic scenario of the power calculations.[29] It is also higher than the enrollment rates reported

[27] As a comparison, in Bouguen et al. (2018a), differential construction rate was 43 pp. Martinez et al. (2017a), and in Indonesia (Brinkman et al., 2017), all control villages received the program by follow-up. Blimpo et al. (2019) do not provide information at the village level; however, compliance appears to be high in Mozambique (comparable to our setting) but lower in The Gambia.

[28] In villages where both CPS and SPS are available, five-year-olds would often register at SPS, while three- to four-year-old children would register at CPS.

[29] The most optimistic scenario is 51.43%.

in low-income countries elsewhere in the literature.[30] Finally, such high take-up rate confirms the relatively high level of Cambodian parents' interest in preschool education.

We explore the potential reasons for non-enrollment in Table 5.7.[31] The most common reasons relate either to the self-sufficiency and emotional maturity of the child (the child is too young; too active; is afraid; does not speak well enough; or refuses to go) or to various practical reasons (afraid child gets hurt on way to school; no one there to pick up child; school is too far away). Preschool quality (inadequate school facility, unreliable or unqualified teacher) are infrequently evoked, while financial constrains are almost never mentioned as reason for non-enrollment. Finally, note that since classes are typically given between 7 and 9 am, preschools are unlikely to relax labor supply constraints of the mother and may even constitute an additional constraint when the child is enrolled.

Caregivers also reported the enrollment history of each child. Exposure to a CPS led to an increase of about 3.4 months, while it decreased by about 2.4 months for an APS, for an overall increase of about 1 month. This is an important result when interpreting the magnitude of the ITT effects on test scores: even when a-compliers are taken into account, results are driven by a short period of preschool attendance. While statistical power is sufficient to detect even small impact, we cannot expect the ITT results to be very large.

Finally, Table 5.11 provides children-level information about the substitution patterns. As seen at the village level in Table 5.10, assignment to treatment negatively affects the probability to attend APS as well as the APS exposure. As in Table 5.10, this substitution is almost entirely driven by a CPS/IPS substitution. This confirms that the a-compliers are likely to substitute a poorly resourced preschool with the newly built CPS. Table 5.11 provides all the necessary information to calculate S_{ac} as in quation (5.4). Under the assumptions A1-A5, 28.3% of the sample are a-compliers. Hence, among the 38.9% of the sample who complied, the share of a-compliers is 73%, while 27% of the compliers would have stayed at home if the CPS construction had not occurred. Similarly, 70% of the overall exposure in terms of months enrolled is coming from the a-compliers, while 30% is coming from the h-compliers. Note that despite a higher level of CPS enrollment, the five-year-olds do not drive up the share of

[30] Martinez et al. (2017a) report a differential take-up of about 33 pp, 24 pp in Indonesia (Brinkman et al., 2017), 9 pp in The Gambia (Jung and Hasan, 2016), and 25pp in the previous preschool impact evaluation in Cambodia (Bouguen et al., 2018a).

[31] Eliciting reasons for non-enrollment is always challenging: it is often multi-factorial, affected by social desirability, and influenced by the way the question is framed. Thus, we take these descriptive statistics with caution. Here, we first asked which reasons for non-enrollment apply among a list of possible reasons, including the possibility to add an additional reason; second, we let caregivers rank the most important to third most important reasons.

a-compliers: they participate more in CPSs, but they also substitute more from APSs to CPSs (S^5_{ac}=80%).

5.4.1.3 ITT Impact on Child Performance

We assess the ITT impact of the CPS construction in Table 5.12. As indicated in Equation (5.1), we present both the treatment-control differences controlling for province fixed effect in column 1, the treatment coefficients controlled for baseline characteristics in column 2 (province fixed effect, gender, age and baseline test score), and implement the double LASSO selection model (column 3). Given the high predictive power of the baseline variables (R^2 generally above 50%), the inclusion of control variables greatly reduces the standard errors. Since the PAP (Berkes et al., 2017) pre-defined the set of control[32] and the outcome[33] variables we use, column (2) is our preferred estimate. As a robustness check, in column 3 we implement a double LASSO estimation (Chernozhukov et al., 2018). The set of control variables selected via double LASSO changes from outcome to outcome but generally includes several baseline test scores (usually in a more parsimoniously way than in column (2)), gender, age, and a few household wealth measures (poverty measures, house quality score).

ITT results point toward a positive effect of the CPS construction on the performance of children.[34] Children in treatment villages perform about 0.05 to 0.06 SD higher in treatment than those in control villages in cognitive test scores. Results on physical development are generally smaller, although we do detect a significantly impact on weight for age, the anthropometrics measure the most likely to be impacted in the short run. Specifications used in columns (2) and (3) improve precision but do not fundamentally modify the point estimates. To alleviate the issue of multi-hypothesis testing, we aggregate all the cognitive and physical development test scores into two indexes (CD and PD). The effect on the cognitive index becomes more significant and is

[32]We have some minor deviation from the PAP: in the PAP, we loosely indicated province fixed effect, child, and household main characteristics, as well as baseline child performance measures. Our final set of controls include: child gender, child age (trimester fixed effects), province fixed effect, and all baseline child performance measures (test scores). Thus, we are more conservative than the PAP, as we do not include any household characteristics. Since household characteristics are very well balanced, none of them significantly improve, in term of precision, and none modified our results significantly.

[33]In the PAP, we included the gross and fine motor skills together and did not mention the anthropometric measures (we were unsure whether anthropometric measures would be collected). Given the low level of correlation between anthropometric measures, gross motor, and fine motor items, we decided in Table 5.12 to regroup the anthropometric measures and the gross motor test score in the physical development test score. Fine motor skills, which are a prerequisite for some dimensions of literacy skills, such as writing letters, are included in the cognitive measures.

[34]Note that we also find impacts on parental involvement and perceived returns to education with parents spending more time in “cognitive games” and having a higher perceived return to primary and secondary education. We cover these impacts more in details in Berkes et al. (2019b)

evaluated at 0.051 SD, both for our preferred estimation (column 2) and for the double LASSO estimation. The physical development index is not significant. The cognitive effect corresponds to approximately a tenth of the baseline quintile cognitive gap, or tenth of a year of child development between three and five (Berkes et al., 2019). Given the large substitution pattern documented previously, the small magnitude of the CPS impact should be interpreted with care, as it is driven by both a and h compliers that may have very different treatment effects. We return to this point in the following sections.

Table 5.12 also indicates that results are driven by the five-year-olds, with an impact that reaches 0.1 SD (aggregate measure of cognition) or 0.08 with the double LASSO estimation. The fact that five-year-olds are more likely to enroll in a CPS and have a larger share of a-compliers, yet have a similar effect on enrollment into preschool in general (see Table 5.12), seems to suggest that the a-compliers also contribute to the overall effect – in other words, switching from an APS to a CPS is actually valuable for a child's cognitive performance.[35] We discuss below more in details the respective contributions of a and h compliers in the overall treatment effect.

5.4.2 Estimating the sub-LATEs

Although small, the magnitude of ITT effects reflects both a and h compliers who may have benefited differently from the CPS. We start our investigation of the sub-LATEs by suggesting that the heterogeneity of treatment effects is very likely driven by a heterogeneity of counterfactual enrollment. We then introduce our bounding strategy to estimate the $LATE_{hc}$. Finally, we use two strategies (conditional LATE and predicted LATE) to pinpoint the exact magnitude of the sub-LATEs.

5.4.2.1 Share of a-complier and the source of heterogeneity

Before estimating sub-LATEs, we analyze how the $LATE_{cps}$ behaves when $S_{ac}(\boldsymbol{X})$ – the predictive share of a-compliers as defined in Equation (5.11) – changes. To investigate this relation, we first estimate $S_{ac}(\boldsymbol{X})$ using equation 5.12. As mentioned in section 5.3.5, the specification includes the baseline variables used for the ITT as well as a set of baseline variables likely to predict the preschool enrollment behavior (village level characteristics as well as household, parental, and child characteristics).[36] Second, we split our sample into five quintiles using $S_{ac}(\boldsymbol{X})$, the first quintile having the lowest $S_{ac}(\boldsymbol{X})$ values. Last, we estimate the $LATE_{cps}$ for each sub-group. We present our

[35] Another explanation is that the CPS curriculum is more adjusted to the 5-year-olds and, hence, they benefit the most from CPS.

[36] We also estimated $S_{ac}(\boldsymbol{X})$ using a LASSO selection model. It does not change the conclusion.

results in Figures 5.7. Although our estimation is imprecise, due to the small sample size, Figure 5.7 shows that as the share of a-compliers increases, the treatment effects tends to decrease. This is particularly true for scholastic measures, such as language, mathematics and executive function.

However, the pattern is less clear for fine motor skills, with Figure 5.7 being inconclusive, and is entirely inconsistent for socio-emotional scoring. The fact that the $LATE_{cps}$ for socio-emotional performances does not depend on the share of a-compliers may be interpreted as the APS having zero, if not a detrimental, effect on the socio-emotional development of the children, which would in turn make the bounds inconsistent for this dimension. We return to this interpretation in subsequent sections.

In any case, when looking at the aggregate measure of cognitive development, the treatment effect is reduced when a larger predicted share of a-compliers are switching to CPS. We interpret this finding as suggestive evidence that the presence of an alternative preschool produces a heterogeneity of *counterfactual substitute* that fundamentally modify the treatment effect. Given these results, estimating the sub-LATEs appears crucial to provide a comprehensive picture of the preschool impacts.

5.4.2.2 Bounding $LATE_{hc}$

We start the investigation of the sub-treatment effects by estimating the bounds for the $LATE_{hc}$ in Table 5.13. As described in section 5.3.3, the lower bound is equivalent to estimating $LATE_{cps}$ and the upper bound is equivalent to $LATE_{ps}$. We start the analysis by looking at the effect on months of exposure, here measured as the number of months spent at (any) school since birth. Bounds for exposure indicate that the children who would have stayed at home (if the CPS had not been constructed) are now spending between three and nine months at school. For exposure, however, the upper bound is much more likely. Indeed, children who switch from an APS to a CPS are *unlikely* to have experienced a different level of preschool exposure: both a and h compliers spend about the same time at school.[37] As a result, at least for exposure, we can assume $LATE_{ac} = 0$, which corresponds to the upper bound in Equation (5.9). Consequently, the h-compliers are very likely to have spent around 9 months at school in total.[38] Still, according to Table 5.13, the impact on the h-compliers is bounded

[37] Remember that, in theory, the exposure to IPS and CPS should be the same: lasts 9 months per year, 2 hours a day, 5 days a week.

[38] 9 months corresponds quite precisely to the timeline of the impact evaluation, i.e. the difference between the baseline and the first follow-up. This is in itself reassuring that the $LATE_{hc}$ recovers quite precisely this parameter.

between 0.12 SD and 0.45 SD for the cognitive development index and similarly for the individual scores.[39]

We then estimate the narrow bounds using the $\boldsymbol{B}$ variables. As mentioned, we want the bounds to be both predictive of the enrollment behaviors and parsimonious to preserve a reasonable sample size. We use here the quintile dummies of the $S_{ac}(X)$ which create a small number of cells (5) and is by definition predictive of the enrollment behavior. Narrow bounds presented in Table 5.13 are significantly tighter than before. They vary between 0.13 and 0.42 SD and are consistent with bounds estimated on the full sample. The index of cognitive development is estimated to be between 0.14 and 0.39 SD.

Overall, our results confirms that effects on children who would not have enrolled to preschool drives the ITT effects found previously. While the bounds remain large, the additional X variables allow significantly narrow the overall cognitive index between 0.14 and 0.39. The results show that low ITT impacts are consistent with substantial positive effects on home compliers and that low ITT effects should not be interpreted as evidence against sizeable effects of preschool in general.

In the following, we use additional identification strategies to obtain point estimates of the sub-LATE parameters.

5.4.2.3 Results from the Conditional LATE Estimation

Under the constant sub-LATE assumptions outlined in Section 5.3.5, we use additional instrumental variables interacted with Z to estimate the sub-LATE effects as in Equation (5.13). In Table 5.14, we provide a comparison between our bounds and the sub-LATEs for three different sets of additional instruments. The first two columns provide the bounds, as in Table 5.13, with estimates for the $LATE_{hc}$ varying between 0.12 and 0.45 SD. We then estimate the sub-LATEs using, first, a set of village level characteristics and province dummies, and, second, our preferred estimation, where we add caregiver and child level baseline covariates to the instruments. In order to mimic the specification used to estimate the narrow bounds, we finally provide the same conditional LATE approach using $\boldsymbol{B}$ the quintiles of the predicted share of a-compliers as additional instrumental variables.

Although the estimates remain imprecise (essentially due to the small share of h-compliers and the low first stage), results are generally in line with our estimation of the bounds. The $LATE_{hc}$ are for instance very close to the upper bound, confirming our intuition that for school exposure the upper bound is indeed more likely. Also, while

[39]Here again, in results not shown here, the five-year-olds drive the impact, with an upper bound on the aggregate measure as high as 0.99 SD.

the impacts on language, numeracy, and executive functions are bounded between 0.12 SD and 0.45 SD (or 0.13 SD and 0.39 SD for the narrow bounds), our preferred specification for the conditional LATE suggests that the $LATE_{hc}$ effect is around 0.2 SD, i.e. almost twice as high as the low-bound but much lower than the upper bound and fully in line with our estimation of the narrow bounds. For all outcomes and both specifications, the $LATE_{hc}$ is significantly different from the upper bound, yet never significantly different from the lower bound. Further, the $LATE_{ac}$ is located at a much lower level (around 0.06-0.09 SD for language, numeracy, and executive functions). Our third specification using $\boldsymbol{B}$ as additional instrument provides a slightly higher (and less precise) estimation for the $LATE_{hc}$ (0.32 SD for the cognitive index against 0.16 SD for our preferred specification) but remain well in line with our bounding strategies.

Once again, however, the effects on fine motor and socio-emotional skills are not entirely in line with our assumptions. For both outcomes, the $LATE_{ac}$ is estimated to be higher than the $LATE_{hc}$, a violation of the bounds assumption spelled out in equation 5.7. While this finding should be interpreted with care, as $LATE_{hc}$ and $LATE_{ac}$ are not significantly different from each other, the results suggest that APSs are particularly ineffective and potentially detrimental to the child development in term of socio-emotional and fine motor skills. This interpretation is corroborated by Figure 5.7, where neither dimensions are affected by the share of a-compliers. Further, this interpretation resonates with some of the existing empirical evidence on preschool impacts. Enrolling children in preschool too early is sometimes suspected to have negative effects on socio-emotional skills (Baker et al., 2008). Hence, it is possible, in our context, that APSs are so poorly equipped in material and building that they negatively affected socio-emotional performance. As a result, $LATE_{ac}$ is large and (almost) significant, while better equipped preschools like CPS, which offer a more satisfactory education environment, at least do not negatively affect the socio-emotional development of a child. Similarly, the $LATE_{ac}$ on fine motor skills is positive and significant (0.18 SD), while the effect on h- compliers is small; this may also be interpreted as a negative effect of APSs. As shown in Tables 5.5 and 5.6, the main difference between IPS and CPS is in material and infrastructure: IPSs seems to lack the infrastructure necessary to allow children to develop their fine motor skills. For both skills, children may be better off staying at home than going to an APS. As a result, when a CPS is constructed, a-compliers benefit strongly from the intervention, while the effect on the h-compliers remains small and non-significant.

Lastly, conversely to Kline and Walters (2016), we fail to reject the over-identification test in most cases (except for our preferred specification of the executive functions). It suggests that the homogeneity assumption – i.e. the sub-LATE does not vary on $\boldsymbol{X}$

– could be a valid assumption in our context. This is true for all outcomes except for executive function, where the over-identification test rejects the null hypothesis.

Overall, although the estimated sub-LATEs are only valid under restrictive assumptions the evidence from Figure 7, both bounding strategies and conditional LATE converge to the same conclusion. The LATE(X) estimates are consistently closer to the low bound and entirely inconsistent with the upper bound. They are also, in most cases, consistent with our narrow bounds. Taken at face value, Table 5.14 suggests that for 8.5 months of exposure (about one school year), children who would have stayed at home in absence of CPS perform, on average, about 0.16 SD better on the summary index when enrolled in CPS. If targeted to the right compliers, such a program could therefore reduce by about a quarter the total cognitive inequality measured at baseline.

5.5 Conclusion

In this article, we analyze the issue of close substitution and preschool impact in the context where other competing preschool programs (here called alternative preschools) are also available to parents and their children. We show that the presence of close substitute programs generates two different types of compliers: the children who would benefit from an alternative preschool (a-compliers) and those who would stay at home (h-compliers) in the absence of the program. Even though both groups of compliers may be similar in terms of observed characteristics, their local treatment effects are likely to be fundamentally different, because their counterfactual enrollment condition is different. Averaging together the treatment effects on both sub-populations, which is implicitly what standard treatment parameters (ITT and traditional LATE) produce, does not provide a sufficiently comprehensive picture of the way the program affects children's performance. We argue that, in addition to providing reduced-forms estimates, isolating the treatment effect on both sub-populations of compliers (sub-LATEs) is necessary to provide a comprehensive picture of the preschool impacts and make appropriate policy recommendations.

We rely on a large and well-implemented preschool construction program, evaluated using a randomized controlled trial, to produce three important results. First, using the traditional reduced-form tools (here ITT), we show that the preschool construction program increases preschool attendance (here by +39 pp) and improves the performance of three- to five-year-old children (+0.05 SD). Five-year-olds are more likely to enroll (+45 pp) and benefit more from the CPS (+0.11 SD). Second, we show that a large share of the compliers would have attended another preschool in the absence of the program. Interestingly, the presence of alternative preschools is frequent in the preschool literature: all of the existing literature studying preschool impact report similar substitution patterns. In this paper, we argue that failure to identify the sub-LATE parameter and, in particular, the effect on children who would have stayed at home ($LATE_{hc}$ here) is a shortcoming of the current literature on preschool in low-income countries. We show that, with a set of very plausible assumptions, we can derive bounds for the $LATE_{hc}$ and show that, after about 9 months of preschool, a child's performance increases between 0.13 SD and 0.45 SD. With additional baseline predictors, the bounds can be significantly narrowed to 0.14 and 0.39 SD. Using additional instrumental variables and under heavier assumptions (constant treatment effects within compliers type), we estimate the $LATE_{hc}$ is positive and significant at around 0.2 SD (between 0.16 SD and 0.32 SD depending on the instruments used). This result corresponds to an effect of sizable magnitude as it represents a quarter of the cognitive gap measured at baseline between relatively wealthier and poorer children. The effect of switching from APS to CPS, while lower than the $LATE_{hc}$, remains positive and significant.

Our results directly relate to the existing literature on preschool impact. They are in line with the most positive results reported in Mozambique (Martinez et al., 2017a) and are in sharp contrast with the more disappointing results found in comparable studies in Cambodia, The Gambia , and Indonesia (Bouguen et al., 2018a, 2013; Blimpo et al., 2019; Brinkman et al., 2017). While implementation issues and failure to account for substitution patterns may explain some of these previous results, other studies, less concerned by the substitution issue, also raise doubts over the effectiveness of early childcare development programs (Ozler et al., 2018; Andrew et al., 2018). Our ITT results show that a properly implemented preschool provision, designed and conducted entirely by the Cambodian government, impacts the learning capacities of children. The effect of such a policy is particularly large on children who would have otherwise stayed at home. It means that a similar policy implemented in a context where no alternative childcare provision exists would prove to be a very effective education policy.

This article also relates to the literature on close substitute programs and on the identification of sub-LATEs (Kline and Walters, 2016; Heckman et al., 2000; Hull, 2018; Kirkeboen et al., 2016; Feller et al., 2016). We contribute to that literature by showing that the effect on children who would have stayed at home can be bounded. Our bounding strategy can be implemented in many contexts, is reliable, and is based on very plausible assumptions. Yet extracting bounds and implementing alternative identification strategies depend on one important condition: the experiment must be powered to detect effects on children who would have enrolled in any program (here, any preschool take-up). In other words, the program's take-up (here, the CPS take-up), which is typically used in power calculations (Duflo et al., 2008), would generally not provide enough detection power in presence of close substitutes. Including the substitution patterns in the design and power calculation – through pilot studies or a careful analysis of the available substitution offers – is critical for precisely isolating the sub-LATE parameters.

Appendix

5.A Tables

Table 5.1: Timetable

Period	**Activity**	**CPS construction**
03/2016	Begin CPS construction	0% completed
05/2016 - 07/2016	Baseline data collection	17% completed
12/2016	Monitoring survey (by phone)	72% completed
04/2017 - 06/2017	Follow-up data collection	86% completed

Note: Percentages refer to share of villages in the treatment group for which construction of a new CPS was completed at the day of data collection.

Table 5.2: Study Sample

	Total	Attrition rate	Treatment	Control	Differential attrition
Baseline May-July 2016					
Villages	178		120	58	
Households	4115		2839	1276	
Household members	22240		15347	6893	
children from 2 -4	4393		3058	1335	
Tested children	4316		3008	1308	
Midline April-June 2017					
Villages	178		120	58	
Households	3757		2578	1179	
Household members	20485		14080	6405	
children from 3-5	4018		2762	1256	
Tested children	3963		2721	1242	
Baseline & midline					
Villages	178		120	58	0.0%
Households	3718	13.9%	2572	1146	2.1%
Households members	20283	8.8%	14045	6238	-1.0%
children from 3-5	3973	9.6%	2751	1222	1.6%
Tested children	3857	10.6%	2671	1186	1.9%

The table provides the study universe in term of villages, households, eligible children, and tested children at baseline and at follow-up (1 year after baseline). The attrition and differential attrition columns give the respective overall attrition rate and the differential between treatment and control attrition.

Table 5.3: Treatment-Control Difference at Baseline – Household and Caregiver Data

	Baseline Sample			Baseline & Follow-up Sample		
	Obs.	C	T-C	Obs.	C	T-C
Household characteristics						
Household size	4115	5.402	0.004	3718	5.443	0.017
			(0.097)			(0.1)
Multidimensional poverty	4115	0.412	0.002	3718	0.393	0.004
			(0.032)			(0.031)
House is rented	3560	1.08	-0.002	3202	1.076	-0.007
			(0.015)			(0.015)
Income > $100	4115	0.452	-0.018	3718	0.476	-0.036
			(0.039)			(0.039)
No one completed prim. school	4115	0.221	0.014	3718	0.224	0.002
			(0.025)			(0.026)
Farming activity	4074	0.825	0.005	3716	0.838	0
			(0.029)			(0.029)
Caregivers characteristics						
Female	4391	0.89	0.019	3916	0.89	0.019
			(0.013)			(0.013)
Age	4391	40.777	-0.227	3916	40.669	-0.104
			(1.014)			(1.017)
# of years of education	4330	4.16	-0.216	3868	4.165	-0.173
			(0.239)			(0.25)
Biological parent	4333	0.596	-0.008	3866	0.602	-0.011
			(0.034)			(0.035)
Malnourished	4371	0.141	0.011	3897	0.141	0.009
			(0.015)			(0.016)
Ravenscore (cognitive test)	4344	0.05	-0.107	3872	0.048	-0.095
			(0.067)			(0.07)
Cognitive parenting score	4379	-0.006	0.017	3906	0.023	-0.01
			(0.056)			(0.058)
Negative parenting score	4380	0.002	0.059	3907	0.009	0.043
			(0.062)			(0.063)
Socio-emotional parenting score	4379	-0.008	-0.026	3906	0.022	-0.057
			(0.048)			(0.049)

Each line represents a regression of an outcome variable on treatment group indicators. The first panel looks at the data collected at baseline, while the second at the data collected at baseline among individuals present at follow-up. Estimates correct for heteroskedasticity and intra-village correlations.
* 10%, ** 5%, *** 1 % significance level

Table 5.4: Treatment-Control Difference at Baseline – Children Data

	Baseline Sample			Baseline & Midine Sample		
	Obs.	C	T-C	Obs.	C	T-C
Sample children Characteristics						
Age	4393	3.476	0.005	3918	3.485	0.002
			(0.03)			(0.032)
Female	4393	0.506	-0.022	3918	0.506	-0.02
			(0.017)			(0.018)
Child ill in the last month	4380	0.778	0.023	3907	0.782	0.019
			(0.018)			(0.018)
Complete vaccination	4381	0.548	-0.03	3908	0.554	-0.027
			(0.037)			(0.037)
Underweight	4313	0.302	0.012	3852	0.306	-0.001
			(0.02)			(0.02)
Stunting	4299	0.341	0.026	3841	0.335	0.022
			(0.019)			(0.02)
Sample children Score						
Emerging numeracy	4316	0	-0.065	3857	0.009	-0.066
			(0.046)			(0.045)
Language	4316	0	-0.05	3857	0	-0.034
			(0.051)			(0.049)
Executive function	4316	0	-0.004	3857	0.001	0.013
			(0.049)			(0.05)
Fine motor	4316	0	0.027	3857	0.005	0.033
			(0.052)			(0.054)
Gross motor	4316	0	-0.013	3857	-0.008	0.005
			(0.051)			(0.054)
Socioemotional	4303	0	-0.041	3846	0.005	-0.043
			(0.058)			(0.06)
Pre-program Preschool attendance						
Currently attending preschool	4380	0.153	0.066***	3907	0.152	0.072***
			(0.023)			(0.024)
Days in preschools	4379	35.957	10.968*	3906	35.758	12.697**
			(5.631)			(5.732)
Currently attending IPS or CPS	4380	0.123	0.061***	3907	0.122	0.069***
			(0.023)			(0.024)
Currently attending SPS	4380	0.03	0.005	3907	0.03	0.003
			(0.01)			(0.009)
Home based program	4375	0.104	0.031	3901	0.112	0.023
			(0.02)			(0.022)
Home visit	4375	0.017	0.003	3901	0.018	0.002
			(0.006)			(0.007)

Each line represents a regression of an outcome variable on treatment group indicators and province fixed effect (omitted). Estimates correct for heteroskedasticity and intra-village correlations.

* 10%, ** 5%, *** 1% significance level

Table 5.5: Baseline Comparison of Informal, Community, and State Preschool

	Obs.	IPS	CPS-IPS	SPS-IPS
General Characteristics				
Used only for preschool	267	0.526	0.232***	0.057
			(0.07)	(0.073)
Class-size	267	20.647	1.353	5.895***
			(1.324)	(2.1)
Preschool material				
Tables, 0/1	266	0.15	0.059	0.596***
			(0.061)	(0.061)
Chairs, 0/1	266	0.211	0.031	0.564***
			(0.065)	(0.061)
Books, 0/1	252	0.711	-0.033	0.058
			(0.073)	(0.066)
Pen, 0/1	256	0.539	0.058	0.219***
			(0.077)	(0.069)
Games, 0/1	259	0.577	-0.061	-0.025
			(0.077)	(0.075)
Blackboard, 0/1	263	0.71	0.032	0.133**
			(0.069)	(0.059)
Sum material, 0/6	267	2.827	0.125	1.395***
			(0.257)	(0.285)
Preschool problems				
Poor building, 0/1	267	0.075	-0.059**	0.022
			(0.028)	(0.042)
Low teachers wage, 0/1	267	0.18	-0.084*	-0.014
			(0.051)	(0.055)
Budget constraint, 0/1	267	0.241	-0.144***	-0.032
			(0.053)	(0.061)
Not enough spots, 0/1	267	0.714	-0.214***	-0.367***
			(0.075)	(0.069)
Not enough supplies, 0/1	267	0.737	-0.076	-0.167**
			(0.072)	(0.07)
Poor teacher quality, 0/1	267	0.06	0.02	-0.005
			(0.04)	(0.034)
Class held irregularly, 0/1	267	0.098	-0.065*	-0.042
			(0.034)	(0.038)
Sum problems, 0/10	267	2.526	-0.317**	-0.596***
			(0.15)	(0.16)
Teacher characteristics				
Any training, 0/1	267	0.955	-0.003	0.003
			(0.033)	(0.03)
Days of training	221	78.9	13.7	152.9***
			(22.733)	(37.474)
Is paid, 0/1	267	0.759	0.176***	0.185***
			(0.049)	(0.046)
Wage, USD	250	35.185	3.797	50.856***
			(2.362)	(8.178)

Baseline comparison between the three type of preschool types available in Cambodia (IPS, SPS and CPS), according to the village chief questionnaire. Based on the full sample of 267 schools at baseline.
* 10% significance level ** 5% significance level *** 1% significance level

Table 5.6: Follow-up Comparison, Informal, Community, and State preschools

	Obs.	IPS	CPS-IPS	SPS-IPS
General preschool characteristics				
Used for preschool only	339	0.627	0.357***	-0.209**
			(0.061)	(0.085)
Open since, days since 1960	279	19138	1532.4***	-2994.4***
			(283.661)	(897.24)
Preschool problems				
Poor building, 0/1	339	0.729	-0.517***	-0.464***
			(0.061)	(0.077)
Too many children, 0/1	339	0.407	-0.175***	-0.195**
			(0.066)	(0.078)
Not enough teacher, 0/1	339	0.237	0	-0.009
			(0.061)	(0.077)
Not enough training, 0/1	339	0.407	0.077	-0.309***
			(0.069)	(0.072)
Not enough tables & chairs, 0/1	339	0.678	-0.405***	-0.443***
			(0.068)	(0.084)
Not enough teaching material, 0/1	339	0.814	-0.101*	-0.172**
			(0.058)	(0.082)
No sanitary facility, 0/1	339	0.593	0.197***	-0.438***
			(0.067)	(0.08)
No clean water, 0/1	339	0.678	0.064	-0.316***
			(0.066)	(0.087)
Class held irregularly, 0/1	339	0.288	0.015	-0.132*
			(0.065)	(0.075)
Other, 0/1	339	0.051	0.021	-0.018
			(0.032)	(0.036)
Sum problems, 0/10	339	4.881	-0.825**	-2.497***
			(0.323)	(0.38)
Teacher characteristics				
Is paid, 0/1	339	0.966	0.006	-0.019
			(0.025)	(0.037)
Paid regularly	339	0.915	-0.009	0.033
			(0.039)	(0.045)
Wage, USD	326	44.5	0.22	132.5***
			(6.474)	(11.313)
# of working days	336	5.103	-0.024	0.219**
			(0.088)	(0.095)
# of teachers at school	338	1.034	-0.018	0.073
			(0.025)	(0.053)

Follow-up comparison between the three types of preschool types available in Cambodia, according to the village chief questionnaire (1 questionnaire per preschool). Based on the full sample of 339 schools.
* 10% significance level ** 5% significance level *** 1% significance level

Table 5.7: Reasons for Non-Enrollment in Preschool, Follow-up

	Reason Applies	Most Important
Afraid child gets hurt on way to school	76%	17%
Child too active / not enough supervision	56%	7%
No one there to bring and pick up child	56%	17%
School is too far away	45%	6%
Child refuses to go / cries / is afraid	44%	11%
Child does not speak well enough	41%	6%
Too young (no detailed reason)	24%	17%
Enrollment was turned down	22%	13%
School facility is not adequate	11%	0%
School construction is not yet finished	10%	1%
Teacher is not present / cancels too often	9%	1%
Child has long-term illness / disability	9%	2%
Teacher is not well qualified	7%	0%
Child does not need any preschool	6%	0%
Did not think about sending child to preschool	5%	0%
Other	4%	3%
School is too expensive	2%	0%
Child must help with household chores	1%	0%
Personal disputes with teacher	1%	0%
		100%
Observations	3333	3333

Table 5.7 provides the reasons given by caregivers at follow-up to explain why their 3-5 year old children is not enrolled in any preschool (asked only in places where a preschool exists). We first asked which reason applies, then among the reasons that apply what is the most important one.

Table 5.8: Principal Strata Names and Notations

		z=1		
		a	h	c
z=0	a	a-never-takers (ANT) $Y^{z,t}_{ant}$	a-defiers $\emptyset$	a-compliers (AC) $Y^{z,t}_{hc}$
	h	ha-defiers $\emptyset$	h-never-takers (HNT) $Y^{z,t}_{hnt}$	h-compliers (HC) $Y^{z,t}_{hc}$
	c	ca-defiers $\emptyset$	ch-defiers $\emptyset$	c-always-takers, $\emptyset$

Table 5.8 gives the name, acronym and the expected value notation of each principal strata i.e. for each value of Z and D.

Table 5.9: Non-parametric Descriptive Statistics of Principal Strata at Baseline

	Y^0_{hnt}	Y^0_{ant}	$Y^0_{hnt} - Y^0_{ant}$	Y^0_{hc}	Y^0_{ac}	$Y^{0,0}_{hc} - Y^0_{ac}$
Caregiver variables						
Female	0.911	0.905	0.006	0.75	0.906	-0.156*
	(0.011)	(0.016)	(0.019)	(0.094)	(0.022)	(0.093)
Age	39.183	38.738	0.445	43.893	43.224	0.669
	(0.723)	(1.009)	(1.002)	(5.784)	(1.587)	(5.331)
Years of education	3.884	4.532	-0.648***	4.128	4.445	-0.317
	(0.146)	(0.233)	(0.248)	(1.439)	(0.45)	(1.357)
Cognitive parenting	-0.091	0.096	-0.187***	-0.056	0.199	-0.254
	(0.034)	(0.052)	(0.056)	(0.339)	(0.097)	(0.32)
Negative parenting	0.144	0	0.143**	-0.347	-0.059	-0.289
	(0.047)	(0.057)	(0.062)	(0.386)	(0.106)	(0.357)
Socioemotional parenting	-0.076	0.079	-0.156**	0.072	0.134	-0.062
	(0.038)	(0.057)	(0.063)	(0.327)	(0.077)	(0.313)
Children variables						
Numeracy	-0.37	0.439	-0.809***	0.315	0.276	0.039
	(0.032)	(0.068)	(0.071)	(0.232)	(0.102)	(0.242)
Language	-0.372	0.539	-0.911***	0.252	0.217	0.034
	(0.024)	(0.075)	(0.072)	(0.188)	(0.12)	(0.182)
Executive functions	-0.36	0.638	-0.997***	0.05	0.222	-0.172
	(0.033)	(0.069)	(0.067)	(0.259)	(0.114)	(0.262)
Fine motor	-0.304	0.596	-0.9***	-0.101	0.216	-0.317
	(0.024)	(0.084)	(0.083)	(0.224)	(0.132)	(0.239)
Socio-emotional	-0.209	0.203	-0.412***	-0.034	0.254	-0.288
	(0.036)	(0.063)	(0.064)	(0.363)	(0.092)	(0.344)

Table 5.9 gives the estimation for $y^{0,0}_{hnt}$, $y^{0,0}_{ant}$ $y^{0,0}_{hc}$ and $y^{0,0}_{ac}$ at follow-up. $y^{0,0}_{hnt}$ and $y^{0,0}_{ant}$ are directly observed. Their comparison is given in column $y^{0,0}_{hnt} - y^{0,0}_{ant}$. $y^{0,0}_{hc}$ and $y^{0,0}_{ac}$ are non-parametrically estimated using equation 5.6. Their standard errors are calculated using the delta method. Column $y^{0,0}_{hc} - y^{0,0}_{ac}$ gives the comparison between h and a compliers in the control group at follow-up.
* 10%, ** 5%, *** 1% significance level

Table 5.10: Village Infrastructure at follow-up

	Obs.	C	T-C
Any preschool in village	178	0.81	0.123**
			(0.057)
Any CPS in village	178	0	0.858***
			(0.032)
Any alternative preschool in village (APS)	178	0.81	-0.552***
			(0.066)
$\cdots$ informal preschool in village (IPS)	178	0.655	-0.564***
			(0.068)
$\cdots$ formal preschool in village (FPS)	178	0.241	-0.058
			(0.067)

Table 5.10 presents village level regressions of the outcome variable in rows against the treatment variable. Estimates correct for heteroskedasticity. Column *T-C* gives the result of the regression without any control, Column *C* the average in the control, and Column *Obs.* the number of observations.
* 10%, ** 5%, *** 1 % significance level

Table 5.11: Enrollment and ITT Exposure

	Full sample			5-Year-olds		
	Obs.	C	T-C	Obs.	C	T-C
Enrollment rate						
Any school	4011	0.435	0.106***	1153	0.683	0.099**
			(0.036)			(0.043)
CPS	4011	0	0.389***	1153	0	0.473***
			(0.025)			(0.032)
Alternative preschool, APS	4011	0.435	-0.283***	1153	0.683	-0.374***
			(0.034)			(0.045)
⋯Informal preschool (IPS)	4011	0.284	-0.247***	1153	0.397	-0.32***
			(0.034)			(0.047)
⋯State preschool (SPS)	4011	0.112	-0.041*	1153	0.18	-0.052
			(0.024)			(0.04)
⋯Primary school	4011	0.04	0.004	1153	0.106	-0.002
			(0.009)			(0.022)
Months of Exposure						
Any school	4006	3.672	1.01**	1149	6.94	1.289*
			(0.401)			(0.69)
CPS	4006	0	3.421***	1149	0	5.205***
			(0.253)			(0.406)
Alternative preschool, APS	4006	3.672	-2.411***	1149	6.94	-3.916***
			(0.347)			(0.646)
⋯Informal preschool, IPS	4006	2.411	-2.088***	1149	4.193	-3.402***
			(0.32)			(0.578)
⋯State preschool, SPS	4006	0.994	-0.426*	1149	1.968	-0.744*
			(0.219)			(0.45)
⋯Primary school	4006	0.267	0.104	1149	0.779	0.23
			(0.074)			(0.24)

Table 5.11 gives the first stage for several measures of preschool participation, for the full sample and the sample of 5-years-olds. Estimates correct for heteroskedasticity and are clustered at the village level. Column *T-C* gives the result of the regression without any control, Column *C* the average in the control, and Column *Obs.* the number of observations.
* 10%, ** 5%, *** 1% significance level

Table 5.12: ITT Impacts on Children Performance

	Full sample					5 Year-olds				
	Obs.	C	(1)	(2)	(3)	Obs.	C	(1)	(2)	(3)
Cognitive Development (CD)										
Executive functions	3963	0	0.055	0.05*	0.036	1138	0.738	0.113**	0.074*	0.059
			(0.038)	(0.026)	(0.028)			(0.049)	(0.044)	(0.042)
Language	3963	0	0.055	0.046	0.051*	1138	0.783	0.141*	0.077	0.061
			(0.04)	(0.03)	(0.027)			(0.079)	(0.062)	(0.062)
Numeracy	3963	0	0.048	0.049*	0.035	1138	0.777	0.196**	0.169***	0.128**
			(0.038)	(0.029)	(0.029)			(0.077)	(0.064)	(0.064)
Fine motor	3963	0	0.077*	0.061**	0.062**	1138	0.832	0.156**	0.072	0.080
			(0.042)	(0.03)	(0.028)			(0.07)	(0.056)	(0.056)
Socio-emotional	3959	0	0.039	0.05	0.045	1138	0.369	0.102*	0.109**	0.091*
			(0.037)	(0.037)	(0.036)			(0.054)	(0.054)	(0.050)
CD index	3963	0	0.055*	0.051***	0.051***	1138	0.7	0.142***	0.1***	0.080**
			(0.03)	(0.02)	(0.019)			(0.048)	(0.038)	(0.036)
Physical Development (PD)										
Height for age	3934	-1.703	0.002	-0.003	0.010	1111	-1.724	0.044	0.024	0.005
			(0.035)	(0.035)	(0.022)			(0.063)	(0.062)	(0.035)
Weight for age	3934	-1.533	0.058*	0.056*	0.047**	1111	-1.65	0.092	0.08	0.018
			(0.033)	(0.033)	(0.020)			(0.066)	(0.066)	(0.032)
Gross Motor	3963	0	0.019	0.006	-0.004	1138	0.702	0.002	-0.036	-0.033
			(0.038)	(0.03)	(0.033)			(0.051)	(0.05)	(0.049)
PD index	3969	-1.068	0.022	0.019	0.025	1140	-0.847	0.025	0.023	0.013
			(0.026)	(0.026)	(0.017)			(0.051)	(0.046)	(0.025)

Table 5.12 gives the ITT treatment effect estimates for each outcome variable. Estimates correct for heteroskedasticity and are clustered at the village level. Column *Obs.* indicates the number of children who took the test at follow-up. Column C the performance of the children in the control villages, standardized using the control group (except for anthropometrics measures). Column *(1)* gives the results only with province fixed effect, *(2)* adds covariates and (3) implements a double LASSO estimation.
* 10%, ** 5%, *** 1% significance level

Table 5.13: $LATE_{hc}$ Bounds

	Bounds			Narrow bounds	
	Obs.	Lower	Higher	Lower	Higher
Any School Exposure (m)	4006	2.71***	9.537***	3.064***	8.875***
		(0.815)	(1.672)	(0.793)	(1.206)
Cognitive Development					
Executive functions	3959	0.126*	0.439*	0.137**	0.352**
		(0.065)	(0.224)	(0.063)	(0.143)
Language	3959	0.117	0.408	0.133*	0.391**
		(0.075)	(0.267)	(0.072)	(0.183)
Numeracy	3959	0.127*	0.441	0.143**	0.417**
		(0.072)	(0.271)	(0.069)	(0.183)
Fine motor	3959	0.152**	0.527*	0.162**	0.383*
		(0.075)	(0.271)	(0.075)	(0.211)
Socio-emotional	3959	0.127	0.442	0.143	0.395*
		(0.092)	(0.342)	(0.089)	(0.21)
Cognitive index	3959	0.13**	0.451**	0.144***	0.387***
		(0.051)	(0.195)	(0.05)	(0.132)
Physical Development					
Height for age	3959	-0.013	-0.045	-0.025	-0.096
		(0.086)	(0.298)	(0.085)	(0.217)
Weight for age	3959	0.136	0.47	0.115	0.222
		(0.084)	(0.317)	(0.083)	(0.221)
Gross motor	3959	0.02	0.068	0.004	-0.089
		(0.076)	(0.264)	(0.077)	(0.178)
Index Anthropometrics	3959	0.046	0.161	0.029	0.018
		(0.065)	(0.228)	(0.063)	(0.153)

Table 5.13 gives the bounds for the $LATE_{hc}$. The lower bound is the $LATE_{cps}$, the upper bound is the $LATE_{ps}$. $LATE_{cps}$ and $LATE_{ps}$ are estimated using province fixed effect, gender, age and baseline test scores as control variable ($\boldsymbol{W}$). Narrow bounds are estimated using $\boldsymbol{W}$ and $\boldsymbol{B}$ as control variables and instrument the endogeneous variable (CPS enrollment or any preschool enrollment) by Z, $\boldsymbol{B}$ and $\boldsymbol{B}$*Z. $\boldsymbol{B}$ includes a dummy for each quintile of $\hat{S}_{ac}(\boldsymbol{X})$ the predicted share of a-compliers. Standard errors are robust to heteroskedasticity and are clustered at the village level.
* 10%, ** 5%, *** 1% significance level

Table 5.14: Bounds and Conditional LATE

			Prov. FE & village char.		Prov. FE, village & hh char.		$S_{ac}(\boldsymbol{X})$ Quintiles	
	Lower bound	Upper bound	$LATE_{hc}$	$LATE_{ac}$	$LATE_{hc}$	$LATE_{ac}$	$LATE_{hc}$	$LATE_{ac}$
School exposure (months)	2.71***	9.537***	8.558***	0.446	8.606***	0.402	8.081***	0.697
	(0.815)	(1.672)	(0.802)	(0.534)	(0.743)	(0.522)	(1.417)	(0.727)
Overid. test p-value			0.686		0.877		0.846	
Executive functions	0.126*	0.439*	0.11	0.08	0.225*	0.039	0.265	0.077
	(0.065)	(0.224)	(0.126)	(0.075)	(0.115)	(0.073)	(0.173)	(0.108)
Overid. test p-value			0.128		0.05		0.776	
Language	0.117	0.408	0.129	0.071	0.198*	0.056	0.363	0.025
	(0.075)	(0.267)	(0.122)	(0.094)	(0.112)	(0.088)	(0.23)	(0.131)
Overid. test p-value			0.835		0.77		0.659	
Numeracy	0.127*	0.441	0.105	0.075	0.2*	0.052	0.386*	0.028
	(0.072)	(0.271)	(0.118)	(0.091)	(0.113)	(0.088)	(0.208)	(0.12)
Overid. test p-value			0.187		0.138		0.476	
Fine motor	0.152**	0.527*	0.055	0.172**	0.058	0.167**	0.244	0.123
	(0.075)	(0.271)	(0.163)	(0.083)	(0.152)	(0.08)	(0.26)	(0.115)
Overid. test p-value			0.844		0.949		0.716	
Socio-emotional	0.127	0.442	-0.009	0.172	0.118	0.132	0.337	0.051
	(0.092)	(0.342)	(0.17)	(0.118)	(0.162)	(0.121)	(0.253)	(0.164)
Overid. test p-value			0.478		0.176		0.25	
Children index	0.13**	0.451**	0.078	0.114*	0.16*	0.089	0.319**	0.061
	(0.051)	(0.195)	(0.101)	(0.061)	(0.121)	(0.058)	(0.247)	(0.082)
Overid. test p-value			0.327		0.126		0.586	
Observation	3959	3959	3959	3959	3959	3959	3959	3959
First-stage F	238.68	11.151	8.285	8.612	6.413	6.658	6.712	6.599

Table 5.14 first provides the bounds from Table 5.13 in the first two columns. Then, we provide the estimates of $LATE_{hc}$ and $LATE_{ac}$ based on the approach described in Section 5.3.5: we instrument the endogenous variables by a set of baseline variables, $\boldsymbol{X}$, Z and their interactions. $\boldsymbol{X}$ is composed of the province fixed effect, and village characteristics in column 3 and 4 to which we add households characteristics in column 5 and 6 (education level, raven test score, poverty dummy). In the last two columns, $\boldsymbol{X}$ only includes the $S_{ac}(\boldsymbol{X})$ quintiles. For each conditional LATE estimation, we provide the p-value of the over-identification test and the Sanderson and Windmeijer (2016) F statistics of the first stage regressions. Standard errors are robust to heteroskedasticity and clustered at the village level.
* 10%, ** 5%, *** 1% significance level

5.B Figures

Figure 5.1: State Preschool (SPS)

Note: State preschools are generally attached to a primary school and classes are given by a formal preschool teacher. State preschools have usually more experienced teachers and state preschool teachers are better paid than community teachers. Classes in State preschool last 3 hours, 5 days a week against 2 hours in CPS or IPS. Source: Own pictures.

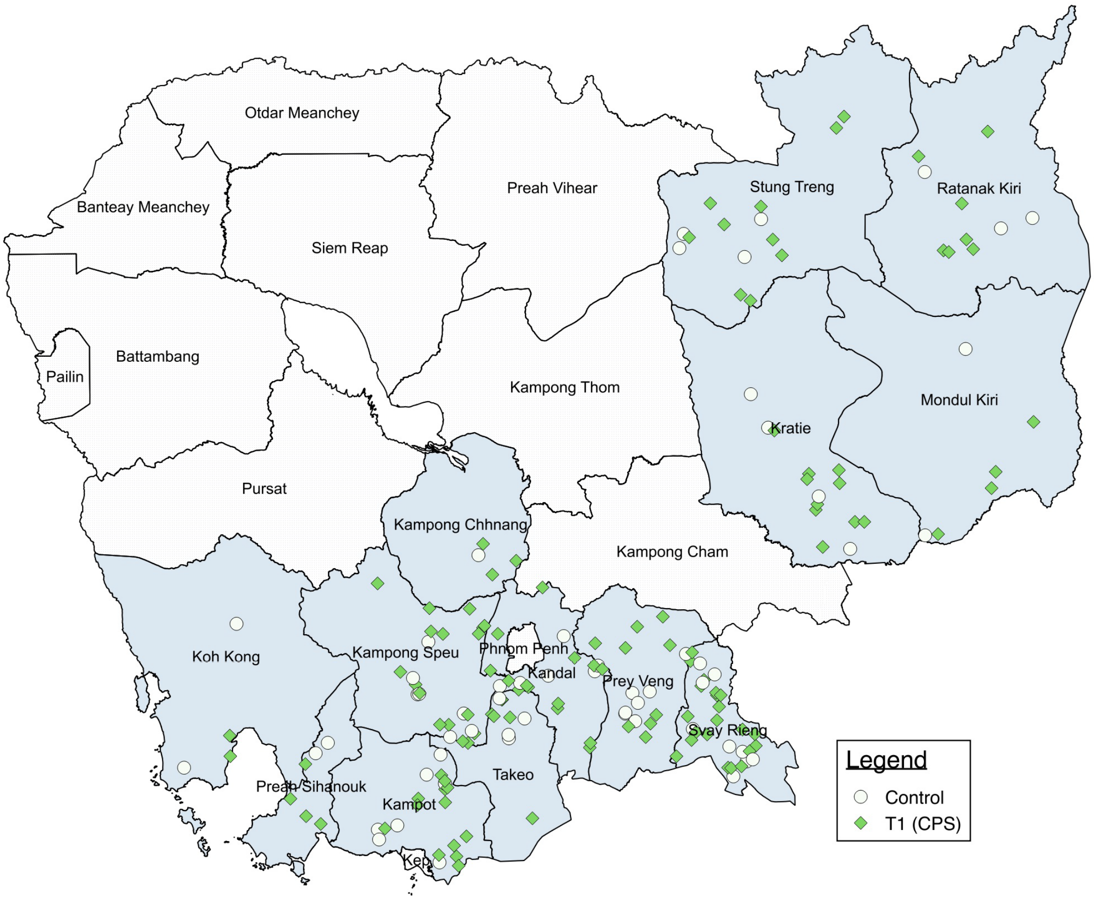
Figure 5.2: Informal Preschool (IPS)

Note: pictures of an informal preschool (IPS) classroom, usually given at the community teacher's house (here under her house). Source: Own pictures.

Figure 5.3: Community Preschool (CPS)

Note: Pictures of a community preschool (CPS). CPS were built under GPE II. CPS are all standard: they are usually much better equipped than informal preschools. Newly recruited teachers receive better training and usually higher wages. Class lasts for 2 hours each day. Source: Own pictures.

Figure 5.4: Location of Treatment and Control Villages in Cambodia



Note: Map shows sample villages by treatment status in the 13 provinces. Source: Own illustration.

Figure 5.5: Principal Strata Before and After CPS Construction

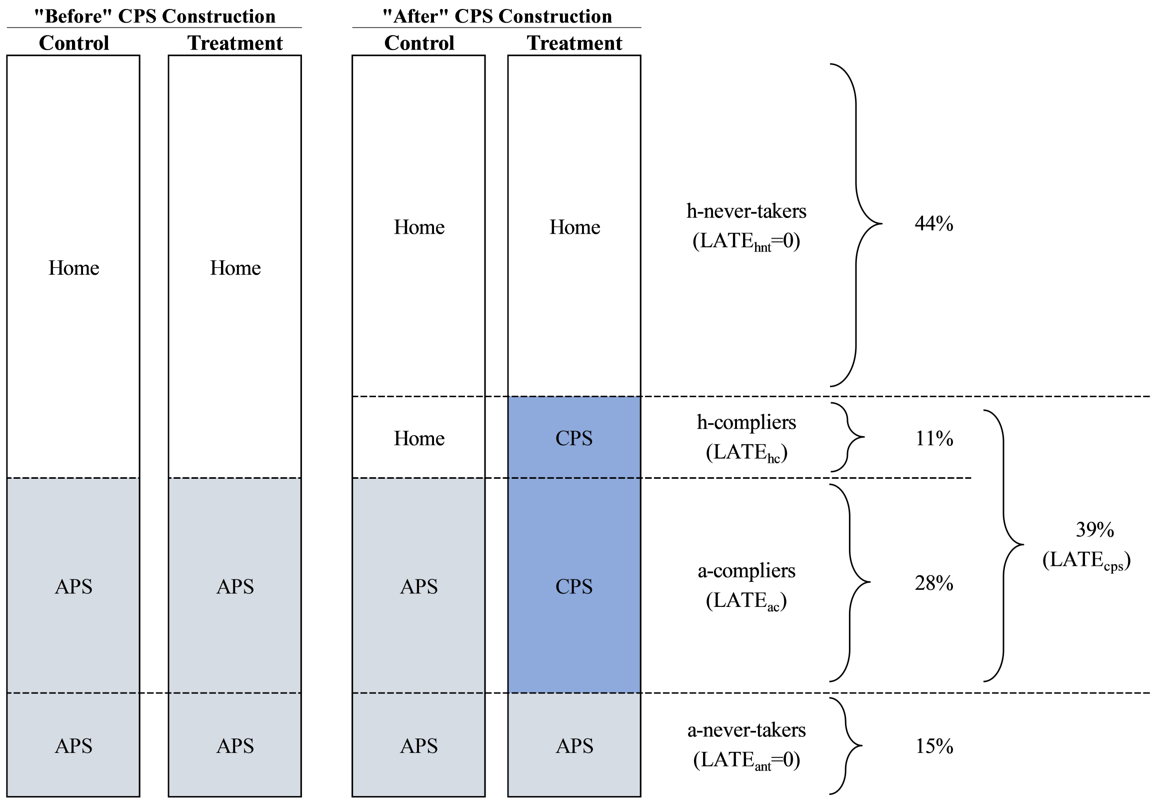


Figure 5.5 shows care arrangements ($D \in \{c, a, h\}$) of children in treatment and control groups. The left panel shows the counterfactual scenario in the absence of the program. The right panel shows the observed scenario at follow-up under implementation of the program. Randomization implies that the control group at follow-up is equivalent to the treatment group at follow-up in the absence of the program.

Figure 5.6: Principal Strata Notations

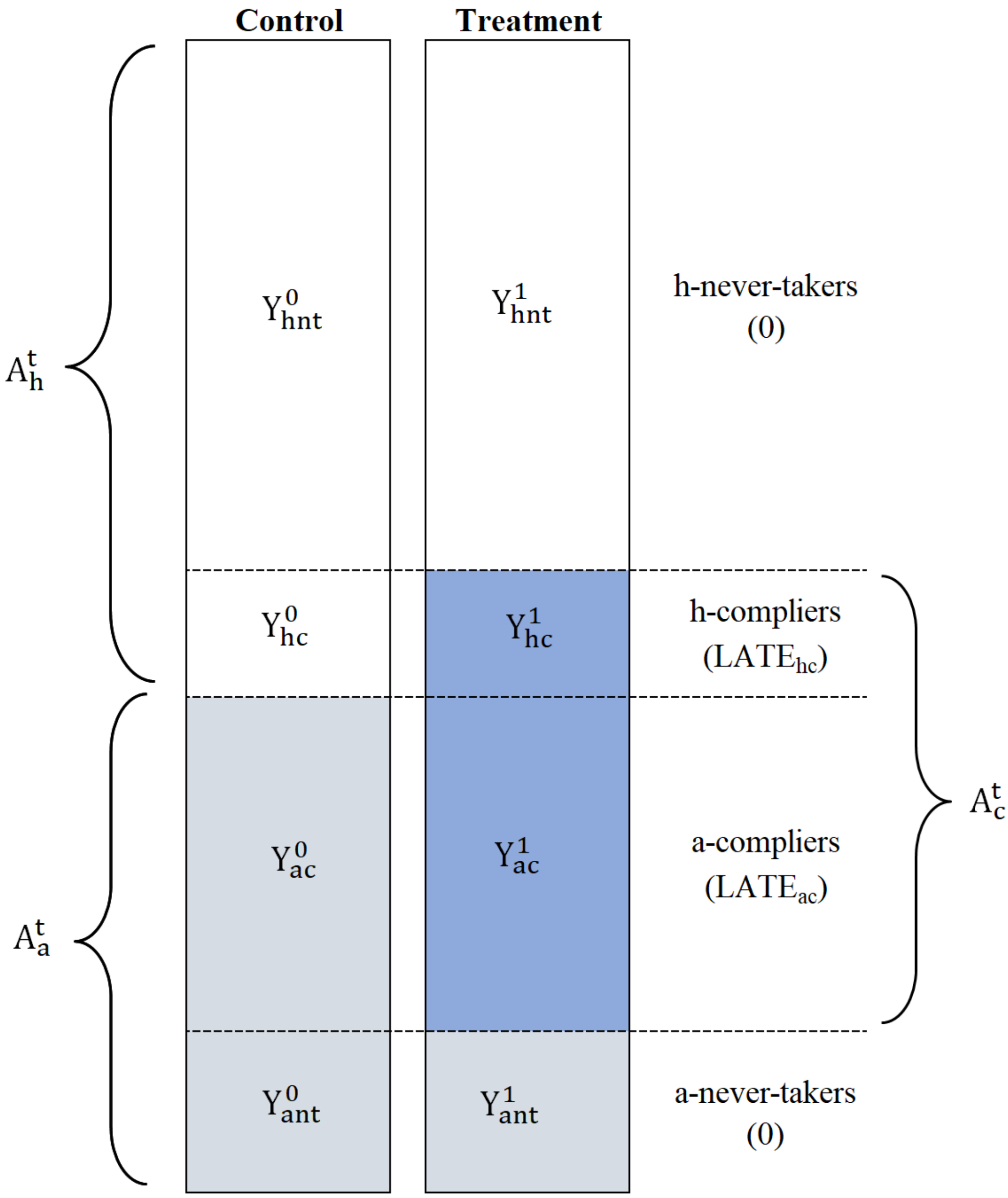


Figure 5.6 shows the notation for each principal strata in treatment and control group after the construction of CPS (Y^z_p) as well as the notation for the weighted average (A^z_d).

Figure 5.7: $LATE_{cps}$ by $\hat{S}_{ac}(\boldsymbol{X})$ Quintiles

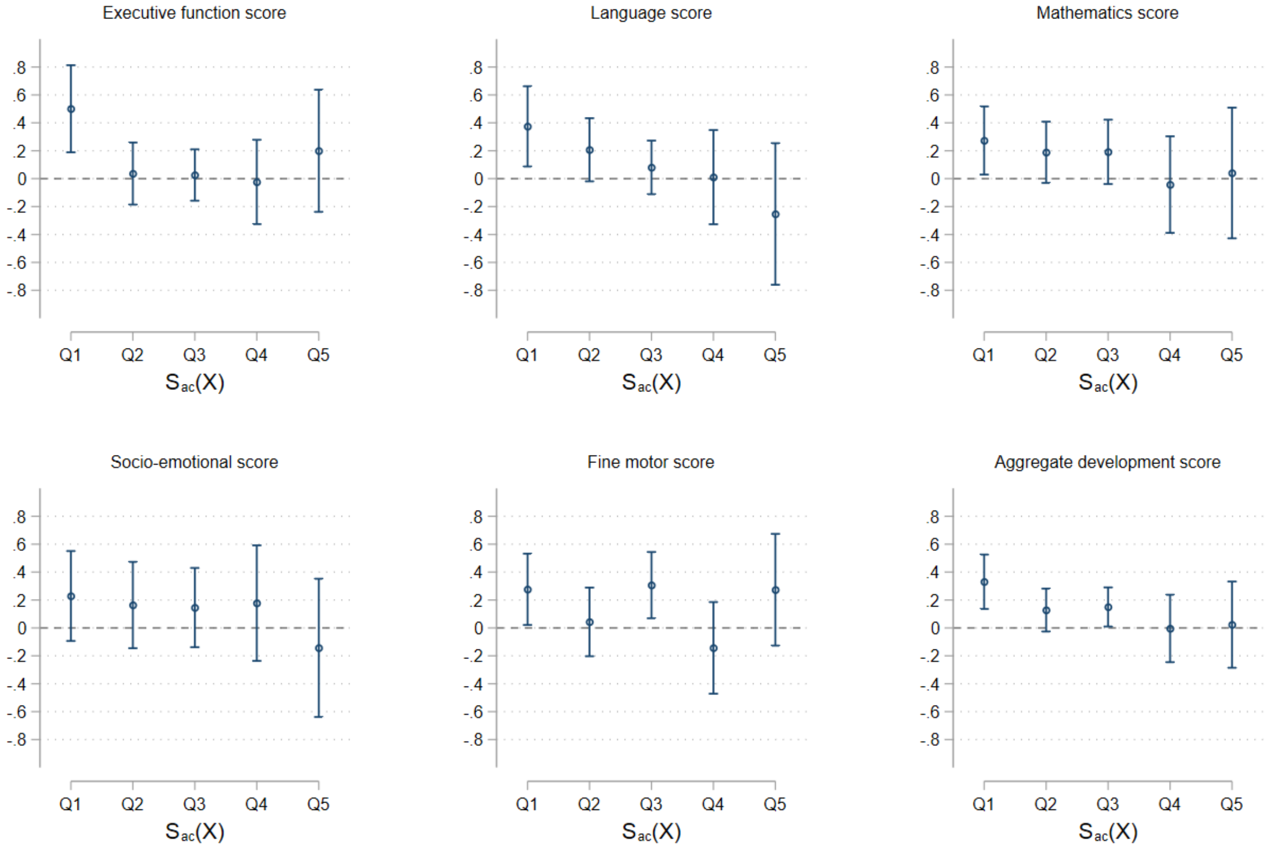


Figure 5.7 shows the treatment effect for the six child development test scores by quintile of $Sac(\boldsymbol{X})$ i.e. the quintiles of the predicted share of a-compliers. Under A3 A4 A5 and A6, the treatment effect should decline as the share of alternative compliers increases.

5.C Appendix

5.C.1 ITT assumptions

The identification of equation 5.1 first relies on the assumption A1 that Z is independent of D and Y. Tables 5.3 and 5.4 confirm that no imbalances on observable characteristics occur; hence, we consider the randomization as being successful and Z as independent of D and Y (Assumption A1).

Additionally, identification requires absence of spill-over effects across treatment and control group villages (Assumption A2[40]). While a few treatment group villages are in the vicinity of control group villages, we have no reason to believe that the construction of CPS had any impact on the education provision of children in the control group. For instance, no control group children attended a CPS at baseline or follow-up. Further, CPS teachers are almost always hired from the same village and, thus, their recruitment is not related to the availability of teachers in the control group.

5.C.2 $LATE_{cps}$ assumptions

In addition to assumptions A1 and A2, the identification of $LATE_{cps}$ relies on the first-stage assumption (A3) or non-zero average causal effect of Z on D (Angrist et al., 1996). We verify assumption A3 using the following first-stage regression:

$$\mathbb{1}_{\{D_i=c\}} = \alpha_0^c + \alpha_{FS}^c Z_v + \boldsymbol{W_i \alpha^c} + \mu_v^c + \epsilon_{iv}^c \tag{5.15}$$

where $\mathbb{1}_{\{D_i=c\}}$ takes 1 when the child is enrolled in CPS and 0 otherwise and A3 is respected when $\alpha_{FS} > 0$. Table 5.11 show that this is valid assumption in our case.

In the presence of close substitutes, we reformulate the exclusion restriction in the following way:

Assumption 4. - *Exclusion Restriction (A4)*

$$Y_i(c, 1) = Y_i(c, 0) \tag{A4.1}$$

$$Y_i(a, 1) = Y_i(a, 0) \tag{A4.2}$$

$$Y_i(h, 1) = Y_i(h, 0) \tag{A4.3}$$

As we show in the result sections, Case A4.1 can be ruled out by construction in our context since no control children attend a CPS. Assumption A4.2 and A4.3, are subject

[40] We do not use the traditional SUTVA assumption because, for the identification of ITT, the absence of spill-over (or general equilibrium effect) across treatment branches is sufficient. SUTVA, as described by Angrist et al. (1996), has larger implications that will be covered in A4.

to violations if the construction of a CPS affects the performance of the never-takers (a or h). CPS construction may reduce APS class-size, change APS peer composition, or make more salient to parents the importance of early educative investment, affecting the performance of both a and h never-takers. We have reasons to believe this is unlikely to be the case in our context.

First, in term of class-size, we have reasons to believe that this problem is unlikely. Indeed, as seen in Table 5.10, the construction of a CPS generally means that the IPS shut down: only 7 treatment villages kept their IPS when a CPS was constructed (6% of the treatment villages). This means that in 94% of the cases, the CPS did not have the indirect effect of reducing IPS class size: thus, IPSs are unlikely to have indirectly benefited from class size reduction.[41] Yet, since SPSs were not shut down when a CPS was constructed, SPSs are more likely to have been indirectly affected by the CPS construction. We look at this possibility in the last row of Table 5.10. Since we did not collect class sizes in SPSs, as a proxy, we use the average number of sampled children enrolled in SPS class per village. Since we did not sample all children in the village (but an average of 26 children), we divide this number by our average sample weight (here estimated at 53%). The last row of Table 5.10 indicates that, on average, the number of sampled children enrolled in SPS in treatment group is 1 unit lower than in the control group. The point estimate is not significant but would correspond, if taken at face value, to a class size reduction of about 1.83 children (0.97/0.53). Since SPS enrollment concerns about 8% of the sampled children (0.083), and since the impact of class size is reported in the literature to be maximum -3 pp per additional students (Bouguen et al., 2017), the indirect effect on class size reduction is estimated to be $1.83 * 0.083 * 0.03 = 0.4$ pp maximum. This would correspond to about 8% of the overall treatment effect (5.1% of a SD). Hence, reduction of class size in SPS is unlikely to significantly modify the magnitude of the treatment effect.

Second, peer composition may violate the exclusion restriction if, for instance, CPSs attract specific children, leaving SPS or the remaining IPS with more homogeneous or better/worst peers. As mentioned in the body of the text, since APSs are composed of better quality schools, we would expect the peer composition to have improved in APSs and, as a result, the a-never-takers to benefit from more favorable conditions. We test this possibility in Table 5.15 for SPS children, where we look at baseline balancing for the SPS children at follow-up. We do not find any significant difference between

[41] Note that we cannot test the IPS class-size because IPSs were closed down when a CPS opened . As a result IPSs are not comparable in treatment and control.

treatment and control in terms of baseline characteristics. This suggests that CPS constructions did not modify a-never-takers' composition.[42]

Lastly, the exclusion restriction may be violated if CPS construction modified the involvement of never-taker parents (h and/or a). While we report elsewhere (Berkes et al., 2019b) that the program positively impacts parents' perceptions and self-reported parenting practices, such effect does not constitute an A4 violation as long as it only affects the a- and h-compliers.[43] A4 would be violated, however, if the parenting effect expands to h-and a-never-takers.

We look at this possibility in Table 5.16, where we estimate the ITT effect on SPS never-takers children. Table 5.16 shows that parents do not report different perceived returns to education, that children do not perform better, but that parents do declare being more involved, on average, in their children's education in the treatment. Since parenting scores are self-reported, this could simply be a reporting bias: with the construction of a preschool in the village, all parents are more inclined to report positive parenting behaviors, while their actual parenting involvement might not have been significantly modified. Yet, as suggested in the body of the text, it could also be that all parents changed their behavior toward early education because of the construction. If that were the case, it would be a violation of A5. We should not overestimate the magnitude of the problem, however. First, the parenting results are driven by socio-emotional parenting – the dimension least correlated with cognitive performance, according to (Berkes et al., 2019), while cognitive parenting, the parenting measure with the highest predictive power, is of lower magnitude and non-significant (+11 pp). Second, even if we take the cognitive parenting at face value, the potential bias remains minimal. According to Berkes et al. (2019), children's performance increases by a maximum of 10 pp for each standard deviation increase of the cognitive parenting index. Hence, an 11 pp effect would translate into a $0.1*11 = 0.11\ pp$ effect on children's performance. Since this effect applies to only 8.3% of all children, the potential effect of the violation of the assumptions is infinitesimal ($0.083 * 0.011 = 0.09$pp compared with an overall effect of 5.1 % of a SD). Given the low magnitude the potential bias, we really do not believe it is a cause of concern for our experiment.

In all, the fact that we do not find any positive impacts for children in Table 5.16 is evidence that Assumption A4 is valid on the whole. Indeed, class-size, peer composition, and parental involvement are all forces that would bias upward the impact in Table 5.16 in case of violation of A4. With an overall ITT effect on SPS children esti-

[42] Again, the same cannot be done for IPS, as CPS construction forced many IPSs to close (see Table 5.10) and, therefore, the IPS treatment sample is a selected one.

[43] Remember that our experiment measures the overall effect of a preschool construction, including indirect effects on parental perception and involvement.

mated at -0.02 SD, we are confident that our experiment is not affected by a violation of the exclusion restriction.

While A4.1, 2 and 3 are necessary conditions for the $LATE_{cps}$ to be identified, they are not sufficient without an extended monotonicity assumption that takes close substitutes into account.

Assumption 5. - *Extended Monotonicity Assumption (A5)*
No child belongs to one of the following strata:

$$\textit{ch-defiers: } D_i(0) = c,\ D_i(1) = h \tag{A5.1}$$

$$\textit{ca-defiers: } D_i(0) = c,\ D_i(1) = a \tag{A5.2}$$

$$\textit{ah-defiers: } D_i(0) = a,\ D_i(1) = h \tag{A5.3}$$

$$\textit{ha-defiers: } D_i(0) = h,\ D_i(1) = a \tag{A5.4}$$

Cases A5.1 and A5.2 are both analogous to defiers in the traditional LATE framework.(Angrist et al., 1996) Since enrollment into CPS is zero in our control group, we can rule out these two cases and consider valid the traditional monotonicity assumption.

Yet, cases A5.3 and A5.4 are theoretically possible. A5.3 (respectively A5.4) corresponds to the situation, where the CPS construction would either decrease (resp. increase) APS attendance. While the existence of ah-defiers is very unlikely, as the construction of CPS is unlikely to reduce the overall demand for preschool, ha-defiers deserve more attention.[44] If, for instance, CPS construction positively modifies the perception of preschool in general and entices some parents to enroll their children in APS instead of CPS (because of shorter distance or because the CPS have no additional capacities), then A5 would be violated. Since CPS cater to a maximum of 25 children, excess demand for CPS may result in higher APS attendance. Relatedly, if APS are already at capacity when the CPS opens, children switching from APS to CPS would make room for ha-defiers. This situation would again violate A5.[45]

However, in our context, the presence of ha-defiers is unlikely. First, as pointed out in Table 5.7, when asked about the reasons why their children are not enrolled at preschool, few parents stated that it was because the preschool was already full. While some parents stated that enrollment was turned down, based on observed class sizes and qualitative interviews with teachers, we interpret this as lack of self-sufficiency

[44]We treat the *ha-defiers* under our extended monotonicity assumption, while Feller et al. (2016) treats it as a sub-assumption called "irrelevant alternatives".

[45]Kline and Walters (2016) discuss this issue for Head Start, where assignment to program preschools could make rationed slots in non-program preschools available to non-treated children. We refer to the same issue as a violation of the extended monotonicity assumption.

Table 5.15: Baseline Description of Children Enrolled in SPS at Follow-up

	Obs.	C	T-C
Household Characteristics			
Household size	336	5.279	0.323
			(0.211)
Multidimensional poverty index	336	0.314	0.022
			(0.082)
Farmer	330	0.866	-0.121*
			(0.065)
No one > 5 years of education	336	0.171	0.022
			(0.052)
Caregivers characteristics			
Raven score (cognitive test)	328	0.139	0.025
			(0.147)
Cognitive parenting score	329	0.26	-0.197
			(0.154)
Negative parenting score	329	0.004	0.086
			(0.134)
Socio-emotional parenting score	329	0.194	-0.077
			(0.11)
Children characteristics			
Early Numeracy	322	0.456	-0.177
			(0.116)
Language	322	0.537	-0.135
			(0.14)
Executive functions	322	0.504	0.003
			(0.168)
Fine motor	322	0.343	0.065
			(0.138)
Socio-emotional	322	0.346	0.129
			(0.131)
Gross motor	321	0.305	-0.101
			(0.139)

Table 5.15 presents children level regressions of the outcome variable in line against the treatment variable. Estimates correct for heteroskedasticity and are clustered at the village level.
* 10%, ** 5%, *** 1 significance level

Table 5.16: ITT estimate on Children Enrolled in SPS at Follow-up

	Obs.	C	T-C
Cognitive Development (CD)			
Executive functions	332	0.544	0.11
			(0.071)
Language	332	0.664	-0.037
			(0.097)
Numeracy	332	0.677	-0.065
			(0.093)
Fine motor	332	0.739	-0.144
			(0.093)
Socio-emotional	332	0.344	0.027
			(0.094)
CD index	332	0.594	-0.022
			(0.051)
Parenting Score (PS)			
Negative parenting	336	-0.04	-0.109
			(0.107)
Socioemotional parenting	336	-0.129	0.216*
			(0.122)
Cognitive parenting	336	0.154	0.114
			(0.147)
PS Index	336	0.022	0.146*
			(0.088)
Parental Perception			
Optimal preschool age	336	3.821	-0.091
			(0.084)
Optimal Primary school age	336	5.829	-0.06
			(0.065)
Perceived Income no school	336	106.183	-9.848
			(7.538)
Perceived Income Prim. School	336	144.429	6.804
			(11.1)
Perceived Income Sec. School	336	227.777	12.6
			(21.687)

Table 5.16 presents children level regressions of the outcome variable in line against the treatment variable. Estimates correct for heteroskedasticity and are clustered at the village level.
* 10%, ** 5%, *** 1 significance level

and emotional maturity to go to preschool rather than capacity constraints. Second – and perhaps more importantly – in the vast majority of cases, the construction of a CPS caused the IPS to shut down: only 7 IPSs remained open after the 103 CPS were constructed, for a total of only sixty-six children enrolled in an IPS following the construction of a CPS. In the vast majority of cases, children staying at home simply could not have enrolled in an IPS because the IPS no longer existed. This applies to IPSs, not SPSs. SPSs remained open regardless of the treatment status of the village. Therefore, we could be in the presence of ha-defiers if, for instance, the children switching from SPS to CPS could be replaced by children who would have otherwise stayed at home. Again, we do not believe this is likely: SPSs provide a much better education environment – in terms of teacher quality, equipment, and even peers than CPS; see Tables 5.5 and 5.6. Thus, parents lacked a reason to remove their children from an SPS and to enroll them into a CPS and, as a result, CPS did not create room in SPSs for ha-defiers.

CHAPTER 6

The Effects of Quality on Child Development in Cambodian Preschools[1]

6.1 Introduction

The importance of early learning opportunities for the development of children and their later life outcomes is well documented and understood (e.g. Currie and Almond (2011); Heckman (2006); Moffitt et al. (2011); Ruhm and Waldfogel (2012)). High quality early childhood education and care (ECEC) is acknowledged to be an important tool to mitigate inequalities in child development due to different home environments (e.g. Waldfogel (2015); Havnes and Mogstad (2015)). While global pre-primary education enrollment rates have almost doubled in the past two decades, compromises in the quality of ECEC programs are often made in order to provide access faster or to more children. Achieving a high quality of ECEC is important as inadequate programs may fall short of the promises made for such early investments in human capital or even be detrimental to early development (Bouguen et al., 2018b; Bernal et al., 2019; Fort et al., 2020). While the case for high-quality ECEC is easily made, it is less clear which quality dimensions matter the most for child development and how to measure them.

This paper uses data from rural Cambodia to study if children who received higher quality preschool education, mainly two to three hours of class per day for children at the age of three to five, show better executive functioning, language and early numeracy skills and less socio-emotional problems. These outcomes are important skills in the dynamic process of lifelong learning (Kautz et al., 2014). A particular focus of the analysis lays on the validation of the novel quality measure and the distinct role the

[1]I thank Yasmine Bekkouche whose support during the data collection and analysis has been essential for the writing of this paper. I am also grateful to C. Katharina Spieß, Adrien Bouguen, Jan Marcus and Ludovica Gambaro for their valuable comments and the World Bank, in particularly Deon Filmer, Tsuyoshi Fukao and Simeth Beng, for providing the data.

quality aspects play for child development. The children in this sample are between 4 and 6 years old at the time of the last follow-up and attended preschool for an average of 9 months. I use a rich set of control variables including child outcomes before preschool attendance in a value-added model. While the preschool quality measures are based on the Measure of Early Learning Environment (MELE) scale, the child outcomes are based on the Measure of Development and Early Learning (MODEL) component of MELQO. Both scales are part of the Measuring Early Learning Quality and Outcomes (MELQO) initiative (UNESCO, 2017; Raikes et al., 2019). In addition, the caregiver administered version of the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 2001; Woerner et al., 2004) is used to measure children's socio-emotional problems.

The data for this study is based on a sample of rural Cambodian villages which was collected to measure effects of a large-scale community preschool construction program (Berkes et al., 2019,b). Data for other, more and less formalized, preschools was collected as well which allows to observe quality differences within and between preschool types. Overall, data on the quality of 329 preschools and about 3,000 preschoolers plus 2,500 non-enrolled children is available. While the data is not representative for Cambodia, it covers a particularly disadvantaged sample of children with a stunting rate of about 36%. The case is particularly interesting to study because it likely resembles the situation of the millions of children world-wide which have not yet benefited from the global rise in preschool access and who are particularly likely to suffer from preschool quality deficits once they have the opportunity to be enrolled.

There are four main findings. First, the five measured quality aspects, which are supposed to have an independent influence on child development (teacher characteristics, equipment, classroom setting, curriculum content and pedagogical approaches, teacher-child interactions) capture more than five latent constructs. I conduct a factor analysis to identify potential important dimensions within the quality aspects. The results show that even the five aspects consist of two to three latent constructs and combining them into single indices, such as using the first principal component for each aspect, omits substantial variation in preschool quality. For example, the measure of equipment is subdivided into a dimension capturing the most rudimentary equipment (tables, chairs and a blackboard), a dimension for learning equipment and a dimension capturing the preschool facilities such as electricity and water access, which vary relatively independently within the studied sample.

Second, the descriptive comparisons of preschool quality show marked differences in quality between and within preschool types. Despite uniform quality standards for formal community and state preschools, preschools with untrained teachers, a lack of basic equipment or a low quality of pedagogical approaches in the observed class exist for all preschool types. The statistics also show that relatively low-cost community preschools can compete with more sophisticated state preschools in only some aspects of quality.

Third, the results of the value-added analysis show that several associations between preschool quality and child outcomes would remain undiscovered when looking only at one-dimensional measures per quality aspect, such as a first principal component of individual items. Based on the more detailed indices from factor analyses I show that children with better preschool equipment show better cognitive development and less socio-emotional problems. In addition, better teacher-child interactions are associated with significantly less socio-emotional problems. There is also an effect of teacher salary on cognitive outcomes and an effect of preschool teacher training on language development of the children. The effect of equipment is entirely driven by the factor measuring the most rudimentary equipment, such as tables, chairs and a blackboard. Eliminating these most pressing quality deficiencies might be a very cost-effective measure to improve child development.

Fourth, the effects of equipment are particularly pronounced for the middle two quartiles of household wealth and baseline test scores. The finding is consistent with a mechanism where the most and least disadvantaged children do not depend on equipment, either because they learn independently of equipment, or they learn nothing at all. This could be the case if they bring too few skills to preschool and are not able to participate in the class in any way. Or if the teacher is focusing on the least disadvantaged group in an environment where targeting all children in the class is not possible due to few personal resources. Such a pattern would be consistent with earlier findings from another study based on the same data set where only children from the highest wealth quartile were able to benefit from the construction of community preschools in Cambodia (Berkes et al., 2019b). In contrast to the effects by equipment, the effects of teacher-child interactions on socio-emotional development (less socio-emotional problems) are particularly pronounced for the lowest and highest quartile.

Causal interpretation of the results depends on the assumptions that preschool quality is uncorrelated with unobserved confounders after conditioning on control variables. As the estimations are based on the sample of children which were enrolled at preschool at least once, a potential threat to identification is that parents observe the preschool quality and base the decision to enroll their child at preschool on this observation. Using the sample of all children, enrolled at preschool or not, I look at associations between preschool enrollment and quality. Consistent with previous results showing asymmetries in information about preschool quality among parents, e.g. from Germany (Camehl et al., 2018) or the US (Mocan, 2007; Bassok et al., 2017; Herbst, 2018), the results indicate that no strong selection is taking place. While easily observable aspects of structural quality are entirely unrelated to preschool enrollment, a very weak relationship between quality measures and preschool enrollment is observed.

Different approaches have been taken to study the effectiveness of preschool programs or their components so far. The role of teachers for student learning is widely agreed to be large (Rivkin et al., 2005; Hanushek and Rivkin, 2012; Chetty et al., 2014; Jackson et al., 2014) and improvements of teacher quality are often seen as key and main challenge for realizing the promises of education (World Bank, 2018). Yet, the same research shows that readily observable characteristics of teachers explain very little of teacher value added effects. This makes it difficult to use these findings to guide the improvement of large-scale preschool programs as effective levers for better learning outcomes are not precisely identified. The findings in this paper confirm that teacher characteristics explain little in process quality, yet show that a minimum standard of preschool teacher training - as least 5 weeks - is important for process quality and the lack of basic equipment has clear effects on child development.

Instead of relating gains in test scores with an overall teacher value-added effect, another research branch has focused on determining teacher characteristics or other measures of quality that are particularly crucial for child development. A common categorization is to distinguish measures of structural quality (such as teacher education or class size) and process quality (such as the quality of teacher-child interactions) of which the second seems particularly predictive of child development (e.g. Araujo et al. (2016, 2019)). However, teacher training programs targeting the improvement of process quality in low-income countries often show mixed effects and positive impacts often faded out quickly (Yoshikawa et al., 2015; Ozler et al., 2018; Wolf et al., 2019). More positive results were found in Colombian preschools where pedagogical training to improve the learning environment did improve children's cognitive development while improvements in structural quality alone even had negative effects (Andrew et al., 2019). While the approach to study the effect of specific teacher training or other quality improvement programs in randomized controlled trials has its merits, a disadvantage of these studies is that they are usually only comparing two different quality modes between which multiple dimensions of quality are simultaneously shifted, which makes it difficult to identify the role of individual dimensions. Variations in quality within a program often remain unexploited, partly because quality measures are not very detailed or not based on observation of ongoing classes. This paper fills an important research gap by showing that even under uniform quality standards that were set by the state for the public preschool system, negative outliers in terms of quality exist and associations with child development suggest that this affects the development of children.

Finally, this paper adds to the existing literature on the validity of different preschool quality measures suitable for a low-income country context. While CLASS (Hamre et al., 2007; Pianta et al., 2012) is one of the most prominent tools used to measure daycare quality, it has not been verified in this context and likely does not distinguish finely enough for low quality levels. Similarly to other tools such as the Stallings instrument (Stallings, 1977; Stallings and Mohlman, 1988; Stallings et al., 2014), or the ECERS (Perlman et al., 2004; Gordon et al., 2013), the CLASS also does also not capture subject specific teacher performance. While the Early Child-hood Environment Rating Scale (ECERS) captures seven focus areas it does not capture subject specific quality. Due to its alignment with the US context the ECERS-R, as previously observed in Indonesia (Brinkman et al., 2017), would also likely required substantial adaptations to fit to the Cambodian context. The MELE tool used for this study provides a promising alternative for the low-income country context, measures the quality of instructions separately for different subjects and is designed to be adapted to national standards.

The remainder of this paper is structured as follows. The next section describes the institutional context of preschools in Cambodia. The sample of the study is described in Section 6.3. Details of the data, particularly the scoring of preschool quality data is part of Section 6.4. The main results, based on the empirical framework in Section 6.5, are discussed in Section 6.6. Section 6.7 concludes.

6.2 Background

Despite two decades of robust economic growth, Cambodia remains one of the poorest countries in Southeast Asia, with a GDP per capita of $1,643 in 2019 ($4,571 in PPP terms). While the country has lower preschool enrollment rates than its neighbors Thailand and Vietnam, the Cambodian early childhood education and care (ECED) sector has seen large improvements in enrollment rates in the past decade.[2] According to official figures, the percentage of primary school grade 1 pupils with some early childhood care and education experience was 65% in 2018 and is projected to increase to 74.5% by 2023. Preschool enrollment rates ranged from 12.6% at age 3, 34% at age 4 to 58% at age 5 in 2018 (MOEYS, 2019a). Substantial amounts have been invested in recent years into the training of preschool teaching staff, preschool construction and preschool improvements MoEYS (2014); MOEYS (2019a).

[2]Source: Preprimary school enrollment data from UNESCO Institute for Statistics (2010–2019).

6.2.1 Preschools in Cambodia

Two distinct types of public preschools co-exist in Cambodia, state preschools and community preschools. While state preschools are under the administration of the Ministry of Education, Youth and Sports, community preschools are administered by the local commune. These two types of preschools form the majority of preschools, particularly in rural areas.[3] In total, 4,301 state preschools exist in the country of which 3,783 are located in rural areas (MOEYS, 2019b). Among the 2,970 community preschools in the country, 2,873 are located in the rural areas. While 509 private preschools exist in Cambodia, only 99 of them are based in rural areas. Thus, community and state preschools, some of them run jointly with NGOs, are what is usually available to young children in rural Cambodia. The Cambodian government reports the percentage of pre-primary schools with access to basic drinking water, basic sanitation facilities and basic hand-washing facilities to be 30.5% in 2018, projected to increase to 55.3% in 2023 (MOEYS, 2019a). I assess the public preschool types in rural Cambodia using own data in Section 6.6.1 and confirm that a large number of preschools does not meet basic quality criteria such as having tables and chairs for the teacher and all students.

In the remainder of the paper, it will be distinguished between state preschools (SPS), formal community preschools with a uniform quality standard for equipment at time of implementation (CPS) and informal community preschools lacking such a standard (IPS).[4]

SPS are financed directly by the Ministry of Education, Youth and Sports. The concept involves providing a relatively high level of quality using large, well equipped preschool facilities with preschool teachers that were trained for two years in a preschool training center in Phnom Penh. To implement this in a cost-efficient way, state preschools are often established on preexisting primary school grounds, sharing sanitary facilities and playgrounds. Teachers of state preschools usually receive a compensation of about $250 per month (183% of the 2019 GDP per capita) by the government. Classes are held on five days a week for three hours a day.

To scale up the reach and capacity of the public preschool system in Cambodia under a scarcity of funds and trained staff, community preschools were established. Community preschool teachers are typically trained for five weeks and are compensated by the local commune. They receive roughly $50 per month (37% of the 2019 GDP

[3] Cambodia has the lowest urbanization rates of Southeast Asia with just 24% of the population living in urban areas in 2019 according to World Bank data.

[4] The terms CPS and IPS are not the formal and usual nomenclature in Cambodia. Both types of community preschools would be considered informal, while SPS are often also referred to as formal preschools. In this paper, the terms CPS and IPS are used to distinguish between the distinct levels of quality that exist in practice.

per capita) but salaries have started to improve in the recent years. For a long time, community preschools were lacking uniform equipment and building standards. Classes by informal community preschools (IPS) are often held in a teachers home, a community hall or a pagoda. Even basic equipment like chairs or tables was often no longer available by the time this study was conducted. The quality of IPS varies tremendously and drop out rates of teachers were high since they had to rely on other sources of income as well.

To increase preschool access and to improve the unsatisfactory quality of IPS, the Cambodian government, with financial support by the Global Partnership for Education (GPE) established new community preschools which were established with an own building and additional materials (CPS). Most of the newly established CPS replaced previously existing IPS. While the CPS are still administered by the commune and CPS teachers receive the same training as IPS teachers and payments remain low, a uniform quality standard guaranteed through centralized procurement of goods and construction of CPS preschool building. The equipment included tables, chairs, and a blackboard among other things. Responsibility for the teacher recruitment and training, the curriculum and the ongoing monitoring of the CPS, including monitoring the regular payment of teacher salaries by the commune is held by the central government. CPS teachers are trained how to structure a preschool lesson, including pedagogical strategies, curriculum content and testing. The training also includes basic knowledge of child development, child rights and parental education. Before and after the training, CPS teachers participate in a written examination.

As part of an impact evaluation that studied the effects of the CPS introduction on enrollment and child development (Berkes et al., 2019b), an extensive costing analysis was conducted.[5] Using the ingredient method, the annual costs of CPS per child were estimated to be \$256. With annual costs of \$426 per child, the estimated costs of SPS are much higher. Costs of IPS were not assessed due to lack of a uniform quality standard.

The common absence of a proper building and delays in IPS teacher payments made the IPS model unsustainable. As a previous study has shown (Bouguen et al., 2013), many schools that were planned during an ealier construction program never opened or were only operating for a short period of time. CPS are considered a promising compromise by the Cambodian government, between the high costs of SPS and the operational problems with the much cheaper IPS model. Whether the CPS proves similarly effective to SPS however depends on whether the aspects that were improved for the CPS model are crucial for child development.

[5]See Berkes et al. (2019a) for the full discussion of cost estimates.

6.3 Sample

The sample of this study was originally drawn to measure the effects of community preschools in a randomized controlled trial. All 305 sample villages are situated in the south and northeast parts of Cambodia. Eligibility criteria for villages to participate in the study were expressed demand for a new community preschool, a high poverty rate, and a large number of children aged zero to five years. Before baseline, villages were randomly assigned to the control group or one of the three treatment groups, which all included the construction of a new community preschool and additional demand-side interventions to strength the demand: a door-to-door initiative in which parents were informed about the new preschool and a home-based program, in which parents received additional information about parenting and the importance of preschool. Effects of the randomized intervention are reported in Berkes et al. (2019).[6]

In this study, the sample is used to exploit variations in preschool quality and study its links to child development. The variation in quality that is exploited in the analysis comes from quality variations within preschool types, while the variation between preschool types will be controlled for by using preschool type fixed effects. As the preschools established as part of the randomized controlled trial were all CPS, the randomized treatment is factored out by using preschool type fixed effects.

The analysis builds on three waves of data, collected annually in the late spring and early summer of 2016-2018. Up to 26 eligible households per village with children at the age between 25 and 59 months were sampled in the first data collection (baseline).[7] This makes the sample representative for households in the village that have children at preschool age at the time of the follow-ups. National representativeness of the sample is not given as the villages for this study were selected based on their need for a functioning preschool and a higher poverty rate. This captures a very relevant group of children that are still lacking access to preschool education. The program for which the sample was originally drawn also focused on the southern and eastern provinces only. In total, 7,053 eligible households were selected during the baseline data collection. A total of 207 CPS, 59 IPS and 63 SPS were identified within the sample village boundaries. IPS and SPS might not be representative of IPS and SPS in Cambodia as the villages were selected based on the need for an additional preschool. Yet, as preschools were usually established as part of previous development programs and established using similar concepts, the preschools outside the sample in rural Cambodia are likely characterized by similar quality and potentials for improvement.

[6] See also Berkes et al. (2019) for additional sample and data descriptions.

[7] The sample was collected using a similar sampling strategy as the Expanded Programme on Immunization of the World Health Organization (Henderson and Sundaresan, 1982).

6.4 Data

In each selected household, a household survey, caregiver survey, and one assessment per eligible child were conducted.[8] At the village level, interviews were conducted with village heads and preschool teachers. In addition, the endline data collection was complemented by a classroom observation survey conducted at all preschools within the sample villages to measure process and structural quality of preschools.

The household survey includes information about family structure, household wealth, and other socioeconomic background characteristics. I construct a dwelling quality index and household assets index using Multiple Correspondence analysis (MCA). Each variable is coded using the sign of the coordinate of the first MCA dimension. I then used the standardized version of each coded variable, and averaged them to create a wealth index from which wealth quartiles are derived. The preschool enrollment history and the Strengths & Difficulties Questionnaire are part of a caregiver survey. The caregiver survey also includes a short version of the Raven's Progressive Matrices Test to measures caregiver's nonverbal reasoning abilities (Carpenter et al., 1990; Raven, 2000). Lastly, the caregiver survey includes a detailed measure of self-reported parenting practices. In 25 questions, the caregiver reports about practices that can be attributed to three dimensions of parenting. Whether parents engage in activities that foster the development of cognitive skills of the child, e.g. by playing games or reading stories, is captured by the domain "cognitive parenting". Providing emotional support and responsiveness, e.g. by comforting, encouraging or giving compliments to the child is captured by the domain "emotional parenting". More "negative parenting" means parents are more likely to to discipline their child and use harsh punitive approaches.

6.4.1 Preschool quality data

The preschool quality measures used in this study are primarily based on the MELE, the Measure of Early Learning Environments scale (MELE-A), developed as part of the Measuring Early Learning Quality & Outcomes (MELQO) initiative. The MELQO initiative was initiated by UNESCO, UNICEF, World Bank and the Brookings Institution to pair commonly articulated global concepts for child development and ECE quality in a core measurement tool with the possibility for adaptation to local quality standards in a (UNESCO, 2017; Raikes et al., 2019). The basic toolkit is intended for

[8]The caregiver is defined as the direct relative (parent, grandparent, aunt/uncle, or adult sibling) who takes care of the child most of the time. In most cases, the caregiver is a biological parent (60.4% at baseline). In the provinces of Kampong Speu, Kandal, Prey Veng, Svay Rieng and Takeo, the caregiver is often a grandparent of the child. These are provinces with relatively high levels of manufacturing industry; mothers who work in factories only occasionally get to spend time with their child during a work day.

adaptation to low- and middle-income countries where access to quality ECEC facilities is particularly important and data on quality often not available (Raikes et al., 2017). The MELQO consists of two parts, the MELE to measure preschool quality and the MODEL (Measure of Development and Early Learning) to measure child outcomes. A study looking at the validity and reliability of the MELE for a sub-saharan country has shown good psychometric properties but has highlighted the importance of aligning the MELE with cultural context and national standards (Raikes et al., 2020). In Colombia, changes in domains, constructs and items were required to fit the MELE to national standards (Ponguta et al., 2019).

While other tools to measure the quality of teacher-child interactions, such as the CLASS (Hamre et al., 2007; Pianta et al., 2012) or the Stallings instrument (Stallings, 1977; Stallings and Mohlman, 1988; Stallings et al., 2014) have been used in a preschool context before (Araujo et al., 2016), they were rarely used in a context of relatively low preschool quality. The CLASS would have had to be adapted extensively to measure the quality differences present in the sample of preschools. Similarly, the ECERS (Perlman et al., 2004; Gordon et al., 2013) had to be adapted when used in Indonesia (Brinkman et al., 2017). However, in contrast to MELE these scales do not provide subject specific observations, such as the pedagogical approaches to teach math subjects specifically. To guide policy recommendations this level of detail a crucial prerequisite for the tool used in this study. The CLASS also requires very extensive training of classroom observers. While the MELE also requires training it is a bit more “hands on” than the CLASS. The Stallings instrument would have been feasible to implement but it does also not provide the level of detail desired for this analysis. Therefore, the MELE provides a good middle way with much detail and breadth in its items.

6.4.1.1 Measuring preschool quality

For this study, several steps have been conducted with the objective to obtain a version of the MELE that is well adapted to the preschools in the studied context, that is culturally appropriate and that is also feasible to use. In a first step, the core MELE has been compared to the national curriculum to identify sections that should be removed or added. For example, items such as having a learning corner were removed as they are neither part of the concept nor present in Cambodian preschools. Numerous examples were modified to fit the curriculum. In a second step, preschools were visited by the research team to test the draft versions and conduct focus-group discussion with preschool principals and teachers to get their feedback. During this process items were separated into an observation tool and a teacher survey, as some items were difficult to observe without the help of a teacher (such as documenting learning materials which are sometimes stored in lockers). In a third step, the tools were tested in 10 preschools for feasibility together. This led to several adaptations, such as separating observations for the first 60 minutes of the class and the remaining duration.[9] Due to the need to continuously jump back and forth in the list of items, tablet-based data collection turned out to be impractical and a paper-based tool was used for this part of the survey. A summary sheet was added to the questionnaire that allowed enumerators to easily count different interactions, e.g. the frequency of negative interactions or the minutes the children spent without supervision. The video material was used afterwards to discuss situations where classroom observers differed in their assessments. The video material was also used for illustration purposes during the final training.

6.4.1.2 Scoring preschool quality data

Using the teacher survey and classroom observation tools the preschools are rated in three aspects of structural quality (*teacher characteristics, equipment, classroom setting*) and two aspects of process quality (curriculum and pedagogy, teacher-child interactions) ex-ante of data collection. Principal component analysis and factor analysis is used to test if the 5 aspects in MELE can be summarized in 5 indices, or if independent constructs of quality within each aspect justify the use of more indices for this sample of preschools in rural Cambodia.

In a first step, all individual items of each aspect are used to run principal component analyses. The first principal component of each aspect is used to generate a summary score. For example, the first principal component of *equipment* is strongly correlated with items such as tables and chairs.[10] Yet, it is not correlated with the presence of a toilet facility. This is mainly driven by the fact that the presence of these items is not correlated with each other in the sampl, but does not warrant the exclusion of the presence of a toilet facility from the analysis. I therefore conduct a factor analysis to test if the aspects consist of more than one construct or latent variables, which should enter the empirical analysis separately.

To find out how the individual items within the 5 aspects of process and structural quality could be grouped by distinct constructs, an exploratory factor analysis is conducted. To confirm compact patterns of correlations between the items of each quality aspect, the Kaiser-Meyer-Olkin statistic (Kaiser, 1974) is used as a test of sampling adequacy. The test statistic varies between 0.6 (mediocre; 0.60-0.69) and 0.77 (middling;

[9]Videos were recorded of the class. Yet, due to technical limitations, only the first 60 minutes were recorded. The classroom observers made separate assessments for the first 60 minutes and the remaining duration to allow the rated recordings to be used for inter-rater reliability assessments.

[10]See Tables 6.6 - 6.11 for a list of individual items and correlations. More details are discussed in Section 6.6.1.

0.70-0.79) for the five aspects. The Bartlett's test for sphericity rejects zero covariance with between the items with a p-value smaller than 0.0001 for all five aspects. Using parallel analysis a large number of factors per aspect is obtained, i.e. up to 17 factors with eigenvalue greater than zero are obtained. Using a procedure similar to Jones et al. (2017) and promax oblique rotation, the number of factors is reduced by rejecting factors with only one item with a loading greater than 0.3. Once I arrive at a number of factors where each factor has at least two items with loadings greater than 0.3 I stop. This procedure yields three constructs for the aspects teacher characteristics, equipment, classroom setting and curriculum and pedagogy and two dimensions for the aspect teacher child interactions. The individual items for each construct are shown in Table 6.18 and 6.19. The subsequent analysis is conducted with both the scores based on principal component analysis and the construct scores based on factor analysis to see where differences matter.

6.4.2 Child development data

6.4.2.1 Measuring child development

To measure child development the MODEL part of MELQO was used as starting point for a locally adapted tool (See UNESCO (2017) for a description of the measures-of-development process and Raikes et al. (2019) for evidence of validity). The final, approximately 45-minute, child assessment comprehensively covers the cognitive domains executive function, language development and early numeracy. In addition, fine motor development and, at baseline and midline only, gross motor development was assessed. Anthropometric measures (height and weight) were assessed at all waves.[11]

To increase the sensitivity and breadth of the child assessments, additional tests were added to the MODEL toolkit. To add a measure of cognitive flexibility, the Dimensional Card Sort test (Zelazo, 2006) was added. To measure receptive vocabulary skills a Khmer test based on picture recognition derived from the Test de Vocabulario en Imagenes Peabody (TVIP), a version of the Peabody Picture Vocabulary Test adapted for Spanish-speaking populations that was normed to low-income populations in Mexico and Puerto Rico was used. (Dunn et al., 1986; Dunn and Dunn, 2007). Additionally, a test for knowledge of reading concepts based on a tool used by the Cambodian Ministry of Education's Early Child Development Department and a sustained attention test was used. The caregiver-administered Strengths and Difficulties Questionnaire was used to measure socio-emotional problems of children.

[11] An in-depth discussion of midline child tests, scoring methods, cultural adaptations, pretesting procedures, and questions about parenting practices can be found in Berkes et al. (2019)

During the test development and adaptation phase, adequate fit to the local context was prioritized over consistency for purposes such as international comparisons. The goal was to generate a well-functioning test for Cambodia for which a large share of tests in the domains language and early numeracy relate directly to the Cambodian preschool curriculum. The executive function test are only indirectly related.

6.4.2.2 Scoring child development data

To construct the composite scores language, early numeracy and executive function, first all individual tests are standardized. This guarantees that all the tests contribute equal variance to the respective composite score. Most tests are scored by counting the number of correct answers. As long as a child participated in some of the tests, a score of zero was assigned for tests where the child was unable to pass the practice trial stage. The standardized test scores of each domain were then summed into composite scores and standardized again for better interpretability. In general, the higher a cognitive score implies better cognitive development, while a higher socio-emotional problem score implies more socio-emotional problems.

Finally, a cognitive development index is prepared by using the first principal component of the executive function, language development and early numeracy scores.[12]

6.4.3 Matching children to preschools

Children are matched to preschools based on an elimination process, which is described in the following. As preschool identifiers in the child-level data proved to be erroneous, this method is used to avoid that children are matched to a preschool they did not attend. Of a total of 6,961 children for which enrollment histories and child tests are available, 4,626 were enrolled at a preschool at some point before the endline data collection in 2018. Of these, 4,258 children were enrolled only at preschools inside the village boundaries. The other 368 children cannot enter the estimation as no preschool quality data was collected for preschools outside the sample villages. In a next step, children which were enrolled at a preschool type which exists more than once inside the village boundaries are removed from the sample. When a child is enrolled at a CPS but a second CPS is available in that village, matching cannot be conducted without error. 3,372 children remain after this step. In a last step, children for which the corresponding preschool quality data is missing had to be dropped. When a preschool quality observation could not be arranged, e.g. if the preschool teacher stopped teaching due to illness or other personal reasons, no class was observed. Therefore, I ended up

[12]The individual tests, their distributions, and the scoring methods are summarized in the appendix of Berkes et al. (2019b).

with a sample of 2,968 children for which preschool attendance is clearly identified and quality data is available. If a child has attended more than one preschool, it is matched to the preschool that was attended for the longest time.

Due to these restrictions, the sample cannot be considered representative for the regions initially sampled for this study. However, the sample has been drawn on the basis of demand for a community preschool which makes it likely to be negatively selected in terms of other village infrastructure characteristics. Due to the removal of villages with multiple preschools of one type, the sample likely becomes further negatively selected. This makes the sample interesting in its own way, as it captures some of the most disadvantaged children of Cambodia in terms of preschool infrastructure (besides children with no access to any preschool at all). Table 6.1 shows summary statistics of all children (columns 1-3) and the children obtained after the matching procedure(columns 4-6).

6.5 Empirical Framework

To capture the associations between preschool quality and child development, a value-added model of the following form is used

$$Y_t = aY_{t-2} + bX + \mathbf{c}Q + dE + \mu + \rho + \epsilon, \tag{6.1}$$

where Y_t and Y_{t-2} denote outcomes at the endline and baseline. X includes a set of control variables, i.e. child age, age squared and child gender. Regional variation in child development are captured using 13 province fixed effects ρ. The quality measures, either defined by the first principal component or the construct scores based on factor analysis are captured by Q. These two different measures of quality are used below to better capture the role of multidimensionality within quality aspects. E linearly captures the months the child was enrolled at the preschool. Due to a high collinearity between preschool type and some quality variables (such as teacher salary), preschool type fixed effects μ are added to the regression. Yet, the sensitivity analyses show that this does not matter substantially for the results. Standard errors ϵ are clustered at the village level to take the sampling procedure into account.

6.6 Empirical Analysis

6.6.1 Observed preschool quality

To assess differences in preschool quality by preschool type, the teacher survey and classroom observation tool are used. The analysis distinguishes between the different preschool types due to their different underlying concepts. While IPS and CPS are similar in teacher recruitment and training procedures, only CPS benefit from a standardized building. In contrast, SPS classes are usually taught on primary school premises. The SPS teachers are recruited and trained through a different and, officially, more extensive process than in community preschools. To assess the status of preschools mean differences between preschool types in summary scores based on principal component analysis and factor analysis (Table 6.2) are presented.[13]

Overall, the summary measures in Table 6.2 confirm that there are significant differences in quality between preschool types. While CPS and IPS teachers do not differ significantly in overall teacher characteristics and the observed teacher-child interactions, the improvements in equipment and classroom setting for the CPS over IPS can be confirmed. Further, the quality of math teaching is higher in CPS than IPS. In terms of basic equipment (tables, chairs, blackboard) and learning equipment, the CPS is even superior to the SPS. The is arguably because CPS are newer and have just recently been equipped. However, the CPS is inferior to IPS and SPS in terms of water access, sanitary facilities and electricity access; a result that doesn't show when looking only at first principal component. Yet, SPS clearly show a more skilled selection of teachers which seems to translate in a broader curriculum taught, better pedagogical approaches and better teacher-child interactions.[14]

This picture is confirmed by unpacking these comparisons into the individual items in Tables 6.12 and 6.13. SPS teachers have more school education, perform better at the nonverbal reasoning test and are by far the most likely to have had more than 8 weeks of preschool teacher training, which usually means two years of training. Yet, SPS teachers also often had no or almost no preschool specific teacher training at all, as these are often primary or secondary school teachers. Overall, their higher qualification goes along with a much higher monthly payment of 250 USD on average, in contrast of 61-67 USD in CPS and IPS. In terms of within-classroom equipment CPS tend to be best equipped, likely due to their recent establishment and centralized procurement of goods by the government. However, outside of the classroom, SPS are clearly better

[13] In addition we provide distributions of summary scores (Figures 6.1 and 6.2), correlations of the summary scores (Table 6.6) and individual items (Tables 6.7–6.11), and mean differences in individual items and individual items (Tables 6.12 and 6.13) in the Appendix.

[14] The summary scores also show a bell-shaped distribution for all preschool types (Figure 6.1).

equipped with more common access to a playground, water source and toilet facility. The comparison of the classroom setting confirms that SPS classes tend to last for 3 hours a day, 1 hour more than in CPS and IPS. Yet, the average number of enrolled and attending children per classroom is larger for SPS than in community preschools.

Table 6.13 also shows differences in the individual process quality items. The higher qualifications of SPS teachers translate into better teaching. SPS teachers tend to cover more subjects and teach them in a more playful way, e.g. asking open-ended questions. SPS teachers are also more likely to give feedback and encouragement to students. It is important to note, that this might also be due to the longer class duration.

Looking at correlations of summary scores, Table 6.6 shows that both the five aspects captured by first principal components, as well as the underlying dimensions are only weakly correlated and therefore seem to capture independent constructs. Yet, the correlations of the individual items with their respective summary score show that the first principal component that is used as a summary score does not capture all the variables well. For example, the classroom setting score does give much weight to the length of class. Likewise, the equipment summary score does not capture features such as a toilet or a functional water source. Hence, the factor analyses should show if associations between these items and child development exist, that can not be uncovered by looking at associations by first principal components alone.

Finally, the relationship between teacher characteristics and process quality dimensions is presented in in Table 6.16. Female teacher are more likely to use art materials in class. Teacher training of 5 and more weeks predicts the quality of literacy and expressive language exercises and training of more than 8 weeks also predicts the quality of math exercises. A consistent positive and mostly statistically positive association between nonverbal reasoning skills and process quality shows as well. Preschool teachers with a primary or secondary teacher training are significantly less likely to use art and writing materials in class, yet, more likely to shows enjoyment or no negative interactions. The findings imply that ensuring a consistent quality of teachers that have at least primary school education, 8 weeks of training and that perform well in a simple intelligence test might help to avoid negative outliers in terms of process quality.

6.6.2 The value-added of preschool quality on cognitive development

Table 6.3 shows the results from a step-wise regression of child development scores on preschool quality indicators and controls as shown in equation 6.1. A step-wise approach is taken to investigate the role of process quality in particular and to test for the importance of multi-dimensional quality measures - even within a particular domain. No association between teacher characteristics or classroom setting and the cognitive development index is found based on the principal component measures of structural quality (column 1). However, a higher equipment score is associated with significantly higher cognitive development. Adding the measures of quality for curriculum & pedagogy and teacher-child interactions does not change the parameters for structural quality indicators. Hence, the estimates for structural quality would not suffer from substantial omitted variables bias due to unobserved process quality if process quality were unobserved. However, quality measures based on just the first principal component in an effort to reduce dimensionality of quality aspects is likely omitting important variation as the factor analysis showed. While there is a negative association between the curriculum and pedagogy aspect and cognitive development based on the principal component, the association vanishes when the more detailed factors are used.

In columns 3-4 associations between the factor analysis based measures and cognitive development are explored. The results show that teacher characteristics should be treated as multidimensional as the salary measure is significantly associated with the cognitive development index. Yet, teacher age, school education and non-verbal reasoning skills and training do not seem to matter for child score value added. The associations between equipment and child outcomes can only be observed for the most rudimentary items: chairs and tables for the teacher and children and a black board. No association is found for other learning equipment or the preschools access to electricity and water. Results change only marginally when process quality measures are added to the equation for the cognitive development index outcome and no association between process quality factors and cognitive development in found.

For a closer investigation of the link between teacher salary and training and cognitive development, results for individual cognitive outcomes, executive function, language development and early numeracy, are shown in Table 6.14. Mainly, the effects of teacher characteristics are driven primarily by the language outcome, somewhat by the early numeracy outcome and not by the executive function outcome. This finding is in line with results in Berkes et al. (2019b), where effects were particularly found for the domains which includes tests that are closely linked to the curriculum. The preschool curriculum training focuses particularly on the teaching of early language, literacy an numeracy skills. The value-added by basic equipment is driven by all outcomes similarly.

6.6.3 The value-added of preschool quality on socio-emotional development

Similar findings as for the cognitive development of children can be found for the associations between quality measurements and socio-emotional problems in Table 6.3.

A higher equipment score is associated with a significantly lower occurrence of socio-emotional problems (columns 5 and 6).

No associations between process quality scores based on principal components and socio-emotional problems can be observed (column 7). However, a different picture emerges for the role of teacher-child interactions for socio-emotional development when looking at individual factors (column 8). While teachers being more able to address children needs, e.g. by encouraging children more or giving better feedback, cause children to show less socio-emotional problems, the opposite is found for the teachers attitude. Teachers which tend to enjoy teaching more and show less negative attitudes towards the children tend to cause more socio-emotional problems if the regression is interpreted in a causal way. One possible explanation for this finding is that students tend to show more socio-emotional problems if the teacher is not disciplining them, which indicates revesred causality.

6.6.4 Subgroup analysis for the value-added of preschool quality on child outcomes

Finally, effects for subgroups are analyzed by running regressions separately by quartiles of the baseline test score and baseline wealth score. While the association are similar for both subgroup classifications and tend to be weakest for the highest wealth quartile - perhaps because they are able to learn at preschool independently of the observed variation in teacher quality if weaker teachers only focus on them - no clear pattern emerges overall.

6.6.5 Selection of teachers and students

The value-added results are based on the sample of children enrolled at preschool. One potential threat to the identification of causal effects of preschool quality is if the selection of the sample itself depends on preschool quality. As preschool enrollment is strongly correlated with age and the development of a child (see supplementary material in Berkes et al. (2019b)), it might also be correlated with the preschool quality. For example, a parent might not send their child to school if the child requires a degree of supervision that cannot be provided at the available preschool. This parents might be more likely to enroll their children if the teacher is particularly able to address children's needs. In such a case, a downward bias of preschool quality effects would be likely. On the contrary, parents might also decide not to send their child to preschool when they perceive a teacher's quality as too low and think their child is better off at home. In such a case an upward bias of estimates would be likely.

To test if parents are able to observe the teachers quality and react to it the relationship between preschool enrollment and quality is estimated by focusing on villages with at least one observed preschool. In case more than one preschool is available at the village, the average of preschool quality variables is taken. Further, it is controlled for preschool fixed effects in a flexible way by controlling for the number of preschools of each type within the village. In addition, it is controlled for the same variables as in the main regressions, as only selection after conditioning on control variables poses a threat to identification. The results in Table 6.17 show that there is only a weak link between preschool enrollment and quality. Children are more likely to ever be enrolled (column 1) at preschool if teachers are more likely to use art and writing materials, yet, less likely if they are more likely to teach math as a subject. Possibly this is because parents deem these subject to be more or less appropriate to be taught at preschool. As shown by the F-test in Table 6.17, no jointly significant effect of quality indicators on preschool enrollment at the extensive margin can be found. As quality might not only affect enrollment at the extensive margin but change the duration of enrollment, e.g. by allowing earlier enrollment due to better caretaking capacities, it is also tested for a relationship between quality and months of enrollment in column 2. The results are similar but a significant positive relationship can also be found for the quality of literacy and expressive language teaching which makes the quality indicators jointly significant at the 5% level. While a small relationship between preschool quality and enrollment cannot fully be ruled out, the relationship seems to be small in size and insignificant for the quality indicators for which effects on child development are observed in the main results

6.7 Conclusion

This study provides detailed insights into associations between preschool quality and child outcomes in a low-income country context. Consistent with other studies, the analysis confirms that teacher characteristics such as age or education explain little in value added in cognition and socio-emotional test scores. The detailed measure include 14 different quality dimensions based on a factor analysis, which indicated that structural and process quality can not be summarized in a small number of indices without the loss of crucial information. While associations between quality dimensions are low overall, a strong association between the lack of basic equipment and child outcomes is found. Further, teacher-child interactions are significantly associated with socio-emotional development of children.

The results show that even in experimental studies, strong but relatively rare quality deficiencies remain often undiscovered and detailed measures are required to uncover them and measure their effects. To maintain and further improve the quality of the current preschool infrastructure, active monitoring and ongoing quality assurance is required to ensure optimal effects on child development.

While the findings have shown that associations between process quality and child development can be low, the results do not imply that process quality has a low relevance for child development in general. It is possible, and likely, that an improvement in process quality levels that goes beyond the ones observed in this study would make a significant difference to learning outcomes. As ECEC care programs in Cambodia are further improved, additional research should study the importance and interactions of different process quality dimensions in this context.

6.A Tables

Table 6.1: Summary statistics

	All children			Preschool enrolled children		
	(1)	(2)	(3)	(4)	(5)	(6)
	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Age (yrs)	6953	5.446	.851	2975	5.44	.799
Female	6953	.49	.5	2975	.508	.5
Length/height-for-age Z-score	6813	-1.653	1.05	2930	-1.637	1.035
Months enrolled: CPS	6953	3.8	5.465	2975	6.545	5.861
Months enrolled: IPS	6953	1.346	3.862	2975	1.404	3.969
Months enrolled: SPS	6953	1.027	2.959	2975	1.258	3.303
Executive function score	6899	-.012	1.023	2956	.071	.956
Language score	6899	.016	1.01	2956	.038	.906
Mathematics score	6899	-.007	1.01	2956	.041	.933
SDQ: overall problems	6952	.007	.998	2975	-.007	.993
Household size	6953	6.014	2.137	2975	5.998	2.131
Caregiver female	6950	.905	.293	2975	.908	.29
Caregiver age	6950	40.22	14.54	2975	41.369	14.581
Caregiver years of schooling	6859	3.498	3.243	2927	3.538	3.225
Caregiver Ravenscore	6888	.007	1	2944	.003	1.004
Wealth score	6952	-.042	1.009	2975	.005	.96

Summary statistics on sample of children with baseline test scores and preschool enrollment history. Columns 1-3 show all children. Columns 4-6 only show children which were enrolled at least once at preschool (main estimation sample).

Table 6.2: Structural and process quality by type of preschool and treatment status

	Method (1)	CPS (2)	IPS (3)	SPS (4)	IPS-CPS (5)	SPS-CPS (6)	N (7)
Structural quality							
Teacher characteristics	PCA	-0.34	-0.31	1.40	0.029	1.743***	326
Factor 1: age, education, non-verbal reasoning	FA	-0.16	-0.18	0.69	-0.025	0.848***	326
Factor 2: salary	FA	-0.33	-0.25	1.40	0.078	1.728***	326
Factor 3: training	FA	-0.20	-0.21	0.87	-0.009	1.067***	326
Classroom setting	PCA	-0.11	-0.50	0.81	-0.395***	0.921***	326
Factor 1: children enrolled/present, noise	FA	-0.04	-0.29	0.40	-0.247**	0.434***	326
Factor 2: instruction time, break time	FA	-0.25	-0.28	1.08	-0.032	1.329***	326
Factor 3: curriculum, tracking	FA	0.09	-0.34	0.02	-0.434***	-0.067	326
Equipment	PCA	0.43	-1.23	-0.26	-1.655***	-0.688***	325
Factor 1: tables, chairs, blackboard	FA	0.49	-1.31	-0.38	-1.802***	-0.873***	325
Factor 2: learning equipment	FA	0.19	-0.61	-0.06	-0.805***	-0.251**	325
Factor 3: water, electricity	FA	-0.21	0.07	0.63	0.281**	0.838***	325
Process quality							
Curriculum & pedagogy	PCA	-0.01	-0.28	0.29	-0.274**	0.300**	327
Factor 1: art and writing materials used	FA	-0.04	-0.12	0.25	-0.078	0.292***	327
Factor 2: math quality and math toys used	FA	0.05	-0.20	0.02	-0.256***	-0.030	327
Factor 3: literacy/expressive language quality	FA	-0.05	-0.15	0.30	-0.099	0.347***	327
Teacher-child interaction	PCA	-0.07	-0.09	0.31	-0.013	0.389***	313
Factor 1: addressing children's needs	FA	-0.07	-0.07	0.29	0.007	0.363***	313
Factor 2: enjoyment, negative attitude	FA	-0.01	-0.07	0.08	-0.060	0.092	313

Table shows preschool quality summary scores by type of preschool. Summary scores are calculated using either the first principal component of the individual variables (PCA) or using factor analysis (FA). See Section 6.4.1 for details. Columns 2—4 show averages by type of preschool. Columns 5 and 6 show differences between types of preschools. Differences are based on a regression of summary scores on binary preschool type variables using robust standard errors. Observations with missing values are dropped from the summary scores. For details about individual variables of summary scores see Table 6.12 and 6.13. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 6.3: Value-added of preschool quality on child development

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Cognitive development index				SDQ: overall problems			
Teacher characteristics	0.003	0.007			0.027	0.027		
	(0.026)	(0.026)			(0.029)	(0.030)		
Factor 1: age, education, non-verbal reasoning			-0.027	-0.020			0.026	0.010
			(0.022)	(0.021)			(0.027)	(0.027)
Factor 2: salary			0.098***	0.091**			0.018	0.033
			(0.036)	(0.036)			(0.049)	(0.048)
Factor 3: training			0.042	0.039			-0.021	-0.013
			(0.026)	(0.028)			(0.035)	(0.036)
Classroom setting	0.022	0.028*			0.019	0.011		
	(0.016)	(0.016)			(0.026)	(0.026)		
Factor 1: children enrolled/present, noise			0.025	0.023			0.008	0.011
			(0.024)	(0.024)			(0.036)	(0.034)
Factor 2: instruction time, break time			-0.035	-0.024			-0.003	-0.009
			(0.032)	(0.032)			(0.035)	(0.036)
Factor 3: curriculum, tracking			0.020	0.024			0.035	0.018
			(0.032)	(0.032)			(0.039)	(0.038)
Equipment	0.040*	0.060***			-0.075**	-0.085***		
	(0.021)	(0.022)			(0.029)	(0.032)		
Factor 1: tables, chairs, blackboard			0.074**	0.083***			-0.075*	-0.096**
			(0.031)	(0.031)			(0.040)	(0.038)
Factor 2: learning equipment			-0.008	0.006			-0.021	-0.012
			(0.027)	(0.028)			(0.034)	(0.037)
Factor 3: water, electricity			0.019	0.013			-0.040	-0.048
			(0.026)	(0.025)			(0.030)	(0.031)
Curriculum and pedagogy		-0.043**				0.034		
		(0.018)				(0.026)		
Factor 1: art and writing materials used				-0.039				0.030
				(0.028)				(0.036)
Factor 2: math quality and math toys used				-0.017				0.017
				(0.026)				(0.040)
Factor 3: literacy/expressive language quality				0.004				0.024
				(0.024)				(0.031)
Teacher-child interactions		0.004				-0.031		
		(0.018)				(0.022)		
Factor 1: addressing children's needs				0.006				-0.103***
				(0.024)				(0.029)
Factor 2: enjoyment, negative attitude				-0.009				0.063**
				(0.020)				(0.026)
Observations	2967	2967	2967	2967	2986	2986	2986	2986
Adjusted R^2	0.580	0.581	0.582	0.583	0.102	0.103	0.102	0.104
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Enrolment duration	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Preschool type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

This table shows step-wise regressions of endline cognitive development index and socio-emotional problems index on preschool quality indicators. Columns 1-2 and 5-6 show regressions on preschool quality indicators based on the first principal component. Columns 3-4 and 7-8 show regressions on preschool quality indicators based on factor analyses. To deal with missing information in control variables these cases are set to a constant value and binary variables indicating missing values are added. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6.4: Value-added of preschool quality on child development by baseline score quartile

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Cognitive development index				SDQ: overall problems			
	Baseline score quartile				Baeline score quartile			
	1	2	3	4	1	2	3	4
Teacher characteristics								
Factor 1: age, education, non-verbal reasoning	-0.033	-0.017	0.047*	-0.037	0.013	0.043	-0.052	-0.002
	(0.028)	(0.031)	(0.027)	(0.049)	(0.051)	(0.049)	(0.053)	(0.051)
Factor 2: salary	0.023	0.066	0.122**	0.097	0.057	-0.050	0.102	-0.039
	(0.045)	(0.064)	(0.056)	(0.079)	(0.102)	(0.096)	(0.083)	(0.125)
Factor 3: training	0.016	0.075*	0.033	0.062	0.053	-0.081	0.014	-0.016
	(0.035)	(0.040)	(0.040)	(0.055)	(0.069)	(0.056)	(0.064)	(0.084)
Classroom setting								
Factor 1: children enrolled/present, noise	0.018	0.029	0.002	0.006	0.022	0.023	-0.034	0.045
	(0.032)	(0.039)	(0.039)	(0.049)	(0.062)	(0.058)	(0.066)	(0.065)
Factor 2: instruction time, break time	-0.068*	0.003	-0.001	-0.024	-0.012	0.045	-0.023	-0.091
	(0.037)	(0.047)	(0.051)	(0.058)	(0.077)	(0.073)	(0.061)	(0.078)
Factor 3: curriculum, tracking	0.041	-0.012	0.040	0.047	-0.025	-0.089	0.110	0.081
	(0.038)	(0.044)	(0.052)	(0.067)	(0.075)	(0.082)	(0.077)	(0.080)
Equipment								
Factor 1: tables, chairs, blackboard	0.062	0.138***	0.079	0.099	-0.119*	-0.131	-0.138	0.016
	(0.045)	(0.046)	(0.049)	(0.062)	(0.065)	(0.081)	(0.086)	(0.062)
Factor 2: learning equipment	-0.041	-0.049	0.070*	0.033	-0.011	-0.034	-0.006	0.032
	(0.035)	(0.045)	(0.041)	(0.062)	(0.069)	(0.065)	(0.071)	(0.069)
Factor 3: water, electricity	0.010	0.061*	0.002	0.017	-0.062	-0.092	-0.022	-0.021
	(0.028)	(0.034)	(0.039)	(0.054)	(0.055)	(0.056)	(0.053)	(0.057)
Curriculum and pedagogy								
Factor 1: art and writing materials used	0.020	-0.083**	-0.052	-0.059	0.105	-0.022	0.049	-0.012
	(0.036)	(0.040)	(0.040)	(0.058)	(0.065)	(0.066)	(0.067)	(0.082)
Factor 2: math quality and math toys used	-0.070*	0.060	-0.046	0.008	0.107	0.059	-0.142*	0.029
	(0.036)	(0.039)	(0.040)	(0.062)	(0.080)	(0.077)	(0.076)	(0.080)
Factor 3: literacy/expressive language quality	0.043	-0.036	0.015	-0.037	-0.018	0.125**	0.072	-0.101
	(0.027)	(0.033)	(0.035)	(0.053)	(0.063)	(0.058)	(0.053)	(0.062)
Teacher-child interactions								
Factor 1: addressing children's needs	0.007	-0.003	-0.032	0.023	-0.186***	-0.027	-0.152***	-0.075
	(0.028)	(0.034)	(0.038)	(0.055)	(0.046)	(0.060)	(0.055)	(0.059)
Factor 2: enjoyment, negative attitude	-0.049*	0.022	-0.017	-0.015	0.071	0.059	0.080	0.062
	(0.026)	(0.028)	(0.029)	(0.044)	(0.049)	(0.044)	(0.053)	(0.055)
Observations	727	732	732	731	846	776	724	621
Adjusted R^2	0.319	0.263	0.284	0.406	0.085	0.051	0.053	0.016
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Enrolment duration	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Preschool type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

This table shows regressions of endline cognitive development index and socio-emotional problems index on preschool quality indicators separate by quartile of baseline cognitive development index (left four colums) and baseline socio-emotional problems score (right four columns). To deal with missing information in control variables these cases are set to a constant value and binary variables indicating missing values are added. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6.5: Value-added of preschool quality on child development by wealth quartile

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Cognitive development index				SDQ: overall problems			
	Wealth quartile				Wealth quartile			
	1	2	3	4	1	2	3	4
Teacher characteristics								
Factor 1: age, education, non-verbal reasoning	-0.021	-0.019	0.033	-0.056	0.082	0.034	-0.087*	-0.005
	(0.028)	(0.029)	(0.034)	(0.040)	(0.052)	(0.052)	(0.049)	(0.057)
Factor 2: salary	0.036	0.124**	0.091	0.044	-0.035	0.049	0.098	0.036
	(0.069)	(0.056)	(0.062)	(0.066)	(0.115)	(0.096)	(0.079)	(0.090)
Factor 3: training	0.112**	0.062	0.008	0.011	-0.022	-0.031	0.004	0.068
	(0.049)	(0.044)	(0.044)	(0.066)	(0.072)	(0.070)	(0.079)	(0.073)
Classroom setting								
Factor 1: children enrolled/present, noise	-0.003	0.047	0.018	0.052	-0.126*	-0.033	0.067	0.109*
	(0.039)	(0.041)	(0.038)	(0.036)	(0.074)	(0.066)	(0.063)	(0.057)
Factor 2: instruction time, break time	-0.020	-0.087*	-0.054	0.040	0.028	-0.050	-0.042	0.028
	(0.049)	(0.047)	(0.052)	(0.061)	(0.075)	(0.082)	(0.075)	(0.079)
Factor 3: curriculum, tracking	-0.015	0.034	0.043	0.007	0.149*	0.043	-0.094	-0.083
	(0.046)	(0.043)	(0.059)	(0.057)	(0.087)	(0.074)	(0.076)	(0.081)
Equipment								
Factor 1: tables, chairs, blackboard	0.083	0.112***	0.078	0.055	-0.105	-0.122	-0.131*	-0.079
	(0.065)	(0.040)	(0.047)	(0.048)	(0.104)	(0.079)	(0.069)	(0.074)
Factor 2: learning equipment	0.033	0.032	-0.028	0.018	-0.054	0.005	0.074	-0.003
	(0.037)	(0.041)	(0.050)	(0.055)	(0.087)	(0.073)	(0.075)	(0.074)
Factor 3: water, electricity	-0.013	-0.009	0.041	0.031	-0.099	-0.048	-0.063	-0.019
	(0.037)	(0.035)	(0.040)	(0.044)	(0.074)	(0.054)	(0.058)	(0.056)
Curriculum and pedagogy								
Factor 1: art and writing materials used	-0.017	-0.054	0.019	-0.083	-0.040	0.188***	0.027	-0.021
	(0.039)	(0.043)	(0.044)	(0.053)	(0.071)	(0.067)	(0.067)	(0.068)
Factor 2: math quality and math toys used	0.032	-0.026	-0.086*	0.009	0.020	-0.115*	0.022	0.089
	(0.045)	(0.043)	(0.050)	(0.044)	(0.082)	(0.068)	(0.078)	(0.085)
Factor 3: literacy/expressive language quality	0.008	0.030	-0.025	-0.006	0.068	-0.051	0.043	0.071
	(0.035)	(0.033)	(0.038)	(0.052)	(0.065)	(0.056)	(0.058)	(0.067)
Teacher-child interactions								
Factor 1: addressing children's needs	-0.022	-0.010	0.018	0.050	-0.135**	-0.028	-0.145**	-0.121**
	(0.038)	(0.033)	(0.044)	(0.047)	(0.059)	(0.057)	(0.062)	(0.055)
Factor 2: enjoyment, negative attitude	-0.000	-0.012	-0.021	-0.019	0.111	0.034	0.065	0.106**
	(0.035)	(0.027)	(0.033)	(0.044)	(0.079)	(0.046)	(0.051)	(0.053)
Observations	741	737	739	739	744	744	744	743
Adjusted R^2	0.542	0.577	0.600	0.592	0.105	0.090	0.082	0.103
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Enrolment duration	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Preschool type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

This table shows regressions of endline cognitive development index and socio-emotional problems index on preschool quality indicators separate by wealth quartile. To deal with missing information in control variables these cases are set to a constant value and binary variables indicating missing values are added. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

6.B Appendix

6.B.1 Correlations

Table 6.6: Correlation table: summary indices

		1	2	3	4	5	6
1	Teacher characteristics (PC1)	1					
2	Teacher characteristics 1: age, education, non-verbal reasoning	0.86	1				
3	Teacher characteristics 2: salary	0.75	0.5	1			
4	Teacher characteristics 3: training	0.6	0.25	0.28	1		
5	Classroom setting (PC1)	0.36	0.17	0.39	0.37	1	
6	Classroom setting 1: children enrolled/present, noise	0.2	0.03	0.25	0.26	0.91	1
7	Classroom setting 2: instruction time, break time	0.61	0.4	0.66	0.42	0.61	0.35
8	Classroom setting 3: curriculum, tracking	0.03	-0.04	-0.02	0.18	0.68	0.56
9	Equipment (PC1)	0.02	0.06	-0.08	0	0.18	0.16
10	Equipment 1: tables, chairs, blackboard	-0.1	-0.03	-0.17	-0.08	0.07	0.09
11	Equipment 2: learning equipment	0.09	0.09	0	0.06	0.24	0.18
12	Equipment 3: water, electricity	0.37	0.24	0.4	0.24	0.22	0.13
13	Curriculum & pedagogy (PC1)	0.2	0.19	0.15	0.13	0.21	0.15
14	Curriculum & pedagogy 1: art and writing materials used	0.19	0.18	0.15	0.11	0.16	0.09
15	Curriculum & pedagogy 2: math quality and math toys used	0.13	0.15	0.05	0.04	0.12	0.1
16	Curriculum & pedagogy 3: literacy/expressive language quality	0.19	0.06	0.18	0.24	0.31	0.24
17	Teacher-child interaction (PC1)	0.12	0.04	0.07	0.21	0.15	0.1
18	Teacher-child interaction 1: addressing children's needs	0.12	0.03	0.09	0.22	0.17	0.12
19	Teacher-child interaction 2: enjoyment, negative attitude	0.05	0.03	0	0.09	0.06	0.04

		7	8	9	10	11	12
7	Factor 2: instruction time, break time	1					
8	Factor 3: curriculum, tracking	0.07	1				
9	Equipment (PC1)	0.04	0.24	1			
10	Factor 1: tables, chairs, blackboard	-0.09	0.17	0.82	1		
11	Factor 2: learning equipment	0.13	0.24	0.86	0.43	1	
12	Factor 3: water, electricity	0.36	0.02	0.17	-0.18	0.35	1
13	Curriculum & pedagogy (PC1)	0.29	0.07	0.41	0.18	0.48	0.19
14	Factor 1: art and writing materials used	0.28	0.01	0.26	0.11	0.29	0.14
15	Factor 2: math quality and math toys used	0.15	0.04	0.45	0.22	0.52	0.12
16	Factor 3: literacy/expressive language quality	0.27	0.2	0.2	0.04	0.28	0.2
17	Teacher-child interaction (PC1)	0.09	0.28	0.08	0.03	0.08	-0.01
18	Factor 1: addressing children's needs	0.12	0.26	0.11	0.03	0.12	0.02
19	Factor 2: enjoyment, negative attitude	-0.04	0.23	-0.01	0.03	-0.05	-0.08

		13	14	15	16	17	18
13	Curriculum & pedagogy (PC1)	1					
14	Factor 1: art and writing materials used	0.84	1				
15	Factor 2: math quality and math toys used	0.89	0.59	1			
16	Factor 3: literacy/expressive language quality	0.49	0.2	0.29	1		
17	Teacher-child interaction (PC1)	0.1	0.07	0.04	0.16	1	
18	Factor 1: addressing children's needs	0.18	0.12	0.11	0.22	0.94	1
19	Factor 2: enjoyment, negative attitude	-0.12	-0.07	-0.14	-0.04	0.74	0.47

Table shows pairwise correlations between preschool quality indices for the 5 categories teacher characteristics, classroom setting, equipment, curriculum & pedagogy, and teacher child interaction. The first variable of each category is the first-principal component (PC1) of the individual variables. The variables based on factor analysis (Factor 1-2/3) are built as explained in Section 6.4.1.

Table 6.7: Correlation table: Teacher characteristics

		1	**2**	**3**	**4**	**5**	**6**
1	Female	1					
2	Age	0	1				
3	Potential experience (years)	0	0.33	1			
4	Highest school grade	-0.09	-0.55	-0.08	1		
5	Teacher weeks of training	0	-0.21	0.05	0.29	1	
6	Practical preschool training	-0.06	-0.12	0.13	0.04	0.43	1
7	Primary or secondary training	-0.1	-0.07	-0.09	0.28	-0.03	-0.11
8	Nonverbal reasoning (Raven's)	-0.09	-0.42	-0.04	0.48	0.14	-0.01
9	Monthly salary (USD)	-0.08	-0.2	0.05	0.45	0.54	0.23
10	Teacher fully paid, regularly	0.06	-0.08	0.01	0.19	0.1	0.04
	Teacher characteristics (PC1)	-0.17	-0.65	-0.14	0.8	0.59	0.32
	Factor 1: age, education, non-verbal reasoning	-0.08	-0.86	-0.34	0.85	0.31	0.05
	Factor 2: salary	-0.14	-0.21	-0.03	0.58	0.27	-0.01
	Factor 3: training	-0.03	-0.19	0.23	0.28	0.9	0.68
		7	**8**	**9**	**10**		
7	Primary or secondary training	1					
8	Nonverbal reasoning (Raven's)	0.12	1				
9	Monthly salary (USD)	0.5	0.2	1			
10	Teacher fully paid, regularly	0.12	0.08	0.23	1		
	Teacher characteristics (PC1)	0.42	0.58	0.75	0.32		
	Factor 1: age, education, non-verbal reasoning	0.25	0.66	0.4	0.17		
	Factor 2: salary	0.8	0.29	0.88	0.29		
	Factor 3: training	-0.11	0.12	0.63	0.14		

Table shows pairwise correlations between preschool quality indicators of the category teacher characteristics. Four summary variables are listed below (see Section 6.4.1 for further details).

Table 6.8: Correlation table: Equipment

		1	**2**	**3**	**4**	**5**	**6**
1	Table and chair for teacher	1					
2	Storage for teacher	0.61	1				
3	Tables and chairs for children	0.49	0.53	1			
4	Children's tables and chairs appropriately sized	0.34	0.4	0.52	1		
5	Blackboard/whiteboard and markers/chalks	0.21	0.19	0.2	0.08	1	
6	Electricity access	-0.13	-0.08	-0.13	-0.11	0.04	1
7	Field, playground, or school yard	0.04	-0.05	-0.04	-0.08	0.02	0.06
8	Equipment for gross motor activities on school yard	0.1	0.13	0.04	-0.02	0	0.04
9	First aid kit	0.19	0.14	0.07	-0.02	0.06	0.1
10	Functional water source	-0.08	-0.12	-0.17	-0.17	0.1	0.33
11	Functional drinking water source	0.14	0.08	0.04	0	0.05	0.19
12	Hand washing facility	0.18	0.16	0.1	-0.03	0.01	0.12
13	Toilet facility	-0.04	-0.17	-0.15	-0.13	0.06	0.25
14	Writing utensils	0.18	0.14	0.21	0.18	0.07	-0.1
15	Art materials	0.19	0.2	0.19	0.08	0.09	-0.12
16	Fantasy play materials	0.24	0.3	0.22	0.13	0.01	-0.02
17	Educational toys/math materials	0.29	0.28	0.18	0.14	0.06	-0.04
	Equipment (PC1)	0.68	0.7	0.62	0.44	0.27	0
	Factor 1: tables, chairs, blackboard	0.77	0.83	0.83	0.64	0.3	-0.13
	Factor 2: learning equipment	0.4	0.42	0.27	0.1	0.08	0.08
	Factor 3: water, electricity	0.03	-0.11	-0.17	-0.23	0.11	0.47
		7	**8**	**9**	**10**	**11**	**12**
7	Field, playground, or school yard	1					
8	Equipment for gross motor activities on school yard	0.36	1				
9	First aid kit	0.13	0.26	1			
10	Functional water source	0.17	0.12	0.15	1		
11	Functional drinking water source	0.16	0.09	0.16	0.46	1	
12	Hand washing facility	0.21	0.28	0.34	0.22	0.17	1
13	Toilet facility	0.15	0.08	0.13	0.43	0.18	0.24
14	Writing utensils	0.22	0.16	0.17	-0.03	0.11	0.13
15	Art materials	0.19	0.11	0.16	0	0.07	0.2
16	Fantasy play materials	0.07	0.22	0.23	-0.02	0.08	0.23
17	Educational toys/math materials	0.11	0.25	0.33	0.08	0.18	0.26
	Equipment (PC1)	0.21	0.36	0.43	0.07	0.26	0.43
	Factor 1: tables, chairs, blackboard	-0.05	0.05	0.13	-0.18	0.1	0.13
	Factor 2: learning equipment	0.37	0.52	0.54	0.21	0.26	0.55
	Factor 3: water, electricity	0.33	0.25	0.33	0.9	0.57	0.42
		13	**14**	**15**	**16**	**17**	
13	Toilet facility	1					
14	Writing utensils	0	1				
15	Art materials	0.01	0.67	1			
16	Fantasy play materials	-0.1	0.34	0.4	1		
17	Educational toys/math materials	0.05	0.51	0.58	0.53	1	
	Equipment (PC1)	0	0.49	0.51	0.55	0.65	
	Factor 1: tables, chairs, blackboard	-0.19	0.3	0.27	0.34	0.35	
	Factor 2: learning equipment	0.12	0.47	0.56	0.6	0.78	
	Factor 3: water, electricity	0.63	0.04	0.03	-0.01	0.16	

Table shows pairwise correlations between preschool quality indicators of the category equipment. Four summary variables are listed below (see Section 6.4.1 for further details).

Table 6.9: Correlation table: Classroom setting

		1	2	3	4	5	6
1	Length of class (minutes)	1					
2	Length of breaks (minutes)	0.58	1				
3	Number of children enrolled in this class	0.24	0.11	1			
4	Children present	0.33	0.17	0.61	1		
5	Num. of non-teachers adults in classroom	-0.09	0.02	0.02	-0.08	1	
6	Teacher follows curriculum to teach class	-0.06	0.06	0.13	0.17	0.1	1
7	Teacher documents children's development regularly	0.16	0.13	0.17	0.22	-0.01	0.42
8	Teacher documents children's attendance	0.11	0.34	0.11	0.15	0.05	0.38
9	Class is interrupted at least once	0.1	0.17	-0.01	0.09	-0.08	0.13
10	Presence of disturbing noise [1-3]	-0.02	-0.1	0.18	0.17	0	0.08
	Classroom setting (PC1)	0.63	0.42	0.7	0.78	-0.09	0.3
	Factor 1: children enrolled/present, noise	0.41	0.16	0.83	0.89	-0.07	0.25
	Factor 2: instruction time, break time	0.88	0.86	0.24	0.35	-0.09	-0.19
	Factor 3: curriculum, tracking	0.17	-0.07	0.36	0.45	0.07	0.73
		7	**8**	**9**	**10**		
7	Teacher documents children's development regularly	1					
8	Teacher documents children's attendance	0.36	1				
9	Class is interrupted at least once	0.2	0.29	1			
10	Presence of disturbing noise [1-3]	0.04	0.02	0.25	1		
	Classroom setting (PC1)	0.5	0.39	0.2	0.23		
	Factor 1: children enrolled/present, noise	0.33	0.21	0.22	0.37		
	Factor 2: instruction time, break time	0.11	0.14	0	-0.18		
	Factor 3: curriculum, tracking	0.79	0.53	0.06	0.08		

Table shows pairwise correlations between preschool quality indicators of the category classroom setting. Four summary variables are listed below (see Section 6.4.1 for further details).

Table 6.10: Correlation table: Curriculum & pedagogy

		1	**2**	**3**	**4**	**5**	**6**
1	Quality: Mathematics	1					
2	Quality: Literacy	-0.11	1				
3	Quality: Exressive language	-0.01	0.36	1			
4	Quality: Storybook	0.08	0.11	0.2	1		
5	Quality: General knowledge	0.12	0.08	0.18	0.12	1	
6	Quality: Fine motor skills	0.09	0	0.14	0.05	0.17	1
7	Quality: Gross motor skills	0.11	0.1	0.2	0	0.03	0.11
8	Educational toys/math materials used by children	0.24	-0.06	0	0.1	0.08	0.13
9	Fantasy play materials used by children	0.05	-0.08	-0.05	0.1	0.05	0
10	Art materials used by children	0.08	0.02	0.06	-0.01	0.03	0.21
11	Writing utensils used by children	0.11	0.1	0.04	-0.01	0.07	0.09
	Curriculum & pedagogy (PC1)	0.37	0.15	0.3	0.22	0.32	0.42
	Factor 1: art and writing materials used	0.16	0.08	0.07	-0.05	0.05	0.24
	Factor 2: math quality and math toys used	0.53	-0.11	0.16	0.33	0.37	0.38
	Factor 3: literacy/expressive language quality	0.04	0.66	0.83	0.38	0.4	0.31
		7	**8**	**9**	**10**	**11**	
7	Quality: Gross motor skills	1					
8	Educational toys/math materials used by children	0.08	1				
9	Fantasy play materials used by children	0.04	0.35	1			
10	Art materials used by children	0.14	0.37	0.24	1		
11	Writing utensils used by children	0.05	0.3	0.21	0.61	1	
	Curriculum & pedagogy (PC1)	0.33	0.61	0.38	0.7	0.63	
	Factor 1: art and writing materials used	0.19	0.45	0.27	0.89	0.82	
	Factor 2: math quality and math toys used	0.23	0.75	0.49	0.48	0.35	
	Factor 3: literacy/expressive language quality	0.35	0.04	-0.08	0.16	0.16	

Table shows pairwise correlations between preschool quality indicators of the category curriculum & pedagogy. Four summary variables are listed below (see Section 6.4.1 for further details). Variables with “Quality:” are sorted from 0 (subject not taught) to 4 (subject taught with highest quality).

Table 6.11: Correlation table: Teacher-child interactions

		1	**2**	**3**	**4**	**5**	**6**
1	The teacher enjoyed teaching [1-3]	1					
2	The teacher showed no negative attitudes [0-2]	0.53	1				
3	Quality of the disciplinary strategies used by the teacher [0-4]	0.43	0.25	1			
4	Less than 5 occurences of negative interactions	0.29	0.14	0.12	1		
5	More than 8 occurences of encouragements	0.17	0.08	0.17	-0.05	1	
6	Children wait more than 5 minutes without any specific activity	-0.06	-0.08	-0.06	-0.03	-0.13	1
7	The teacher correct student's work and give feedbacks [0-3]	0.13	-0.05	0.22	0.02	0.18	-0.08
8	Children ever left without supervision	-0.09	-0.05	-0.04	-0.07	-0.1	0.2
9	Quality of the engagement of children [0-4]	0.26	0.11	0.36	0	0.25	-0.17
10	Teacher's awareness of children's individual needs [0-3]	0.2	0.02	0.41	0.07	0.32	-0.12
11	Teacher's behavior with respect to gender equality [0-4]	0.13	0.08	0.28	0.07	0.22	-0.07
	Teacher-child interaction (PC1)	0.63	0.41	0.71	0.24	0.5	-0.26
	Teacher-child interaction 1: addressing children's needs	0.4	0.13	0.67	0.08	0.53	-0.24
	Teacher-child interaction 2: enjoyment, negative attitude	0.92	0.76	0.59	0.37	0.19	-0.13

		7	**8**	**9**	**10**	**11**
7	The teacher correct student's work and give feedbacks [0-3]	1				
8	Children ever left without supervision	-0.08	1			
9	Quality of the engagement of children [0-4]	0.25	0	1		
10	Teacher's awareness of children's individual needs [0-3]	0.42	-0.03	0.26	1	
11	Teacher's behavior with respect to gender equality [0-4]	0.18	-0.14	0.27	0.31	1
	Teacher-child interaction (PC1)	0.48	-0.19	0.61	0.65	0.53
	Teacher-child interaction 1: addressing children's needs	0.58	-0.15	0.61	0.79	0.56
	Teacher-child interaction 2: enjoyment, negative attitude	0.08	-0.12	0.35	0.23	0.22

Table shows pairwise correlations between preschool quality indicators of the category teacher-child interactions. Three summary variables are listed below (see Section 6.4.1 for further details). Discrete variables are sorted from 0 (lowest) to highest.

6.B.2 Preschool quality data

Table 6.12: Individual structural quality items by type of preschool and treatment status

	CPS (1)	IPS (2)	SPS (3)	IPS-CPS (4)	SPS-CPS (5)	N (6)
Teacher characteristics						
Female	0.94	0.97	0.84	0.03	-0.10*	327
Age	40.48	40.92	34.05	0.43	-6.44***	327
Potential experience (years)	6.07	6.46	5.97	0.38	-0.10	327
Completed grade 6	0.80	0.78	0.98	-0.03	0.18***	327
Completed grade 9	0.44	0.39	0.86	-0.05	0.42***	327
Completed grade 12	0.13	0.14	0.68	0.00	0.55***	327
1-4 weeks of teacher training	0.13	0.32	0.27	0.19***	0.14**	327
5-8 weeks of teacher training	0.70	0.46	0.03	-0.24***	-0.67***	327
More than 8 weeks of teacher training	0.05	0.07	0.41	0.01	0.36***	327
Practical preschool training	0.21	0.20	0.41	-0.01	0.20***	327
Primary or secondary training	0.02	0.05	0.44	0.03	0.42***	327
Nonverbal reasoning (Raven's)	-0.09	-0.04	0.41	0.05	0.50***	326
Monthly salary (USD)	60.60	67.32	250.22	6.72	189.62***	327
Teacher fully paid, regularly	0.75	0.85	0.95	0.10*	0.21***	327
Teacher fully paid, irregularly	0.22	0.15	0.03	-0.07	-0.19***	327
Equipment						
Table and chair for teacher	0.99	0.34	0.87	-0.65***	-0.12***	327
Storage for teacher	0.97	0.19	0.52	-0.78***	-0.44***	327
Tables and chairs for children	0.95	0.32	0.67	-0.63***	-0.28***	327
Children's tables and chairs appropriately sized	0.78	0.29	0.34	-0.48***	-0.44***	325
Blackboard/whiteboard and markers/chalks	0.95	0.83	0.97	-0.12**	0.02	327
Electricity access	0.07	0.25	0.24	0.19***	0.17***	327
Field, playground, or school yard	0.62	0.61	0.83	-0.01	0.21***	327
Equipment for gross motor activities on school yard	0.38	0.32	0.30	-0.05	-0.07	327
First aid kit	0.31	0.12	0.33	-0.19***	0.02	327
Functional water source	0.46	0.64	0.81	0.18**	0.35***	327
Functional drinking water source	0.61	0.54	0.67	-0.07	0.05	327
Hand washing facility	0.53	0.37	0.57	-0.16**	0.04	327
Toilet facility	0.26	0.42	0.90	0.16**	0.64***	327
Writing utensils	0.94	0.80	0.89	-0.14***	-0.05	327
Art materials	0.91	0.71	0.81	-0.20***	-0.10*	327
Fantasy play materials	0.65	0.29	0.25	-0.37***	-0.40***	327
Educational toys/math materials	0.77	0.44	0.67	-0.33***	-0.10	327
Classroom setting						
Length of breaks (minutes)	22.44	21.07	44.22	-1.38	21.78***	327
Number of children enrolled in this class	24.80	22.56	27.57	-2.25*	2.77**	327
Children present	17.60	14.98	21.21	-2.62**	3.61***	327
Num. of non-teachers adults in classroom	0.03	0.03	0.00	0.00	-0.03**	327
Teacher follows curriculum to teach class	0.68	0.46	0.49	-0.22***	-0.19***	327
Teacher documents children's development regularly	0.38	0.19	0.37	-0.19***	-0.01	327
Teacher documents children's attendance	0.87	0.68	0.86	-0.20***	-0.02	327
Class is interrupted at least once	0.58	0.58	0.62	-0.00	0.04	327
Presence of disturbing noise [1-3]	1.36	1.47	1.27	0.12	-0.09	327
Number of schools	207	59	63			

Columns 1—3 show averages by type of preschool. Columns 4-5 show differences between types of preschools. Differences are based on regressions of dependent variables on binary preschool type variables using robust standard errors. Variables with a *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 6.13: Individual process quality items by type of preschool and treatment status

	CPS (1)	IPS (2)	SPS (3)	IPS-CPS (4)	SPS-CPS (5)	N (6)
Curriculum content and pedagogy						
Quality: Mathematics	2.16	1.97	2.06	-0.19	-0.10	327
Quality: Literacy	1.91	1.88	3.02	-0.03	1.11***	327
Quality: Exressive language	2.08	1.81	2.08	-0.27	-0.00	327
Quality: Storybook	1.90	1.39	1.71	-0.51*	-0.18	327
Quality: General knowledge	3.20	3.49	3.49	0.29	0.29	327
Quality: Fine motor skills	0.79	0.81	1.32	0.03	0.53***	327
Quality: Gross motor skills	1.06	0.90	1.38	-0.17	0.32**	327
Educational toys/math materials used by children	0.42	0.25	0.46	-0.17**	0.04	327
Fantasy play materials used by children	0.40	0.19	0.08	-0.21***	-0.32***	327
Art materials used by children	0.55	0.53	0.70	-0.03	0.15**	327
Writing utensils used by children	0.61	0.63	0.79	0.01	0.18***	327
Teacher-child interactions						
The teacher enjoyed teaching [1-3]	2.58	2.49	2.60	-0.08	0.03	327
The teacher showed no negative attitudes [0-2]	1.94	1.93	1.94	-0.00	-0.00	327
Quality of the disciplinary strategies used by the teacher [0-4]	2.95	3.02	3.13	0.07	0.18	327
Less than 5 occurences of negative interactions	0.56	0.49	0.48	-0.07	-0.07	314
More than 8 occurences of encouragements	0.43	0.44	0.60	0.01	0.18**	324
Children wait more than 5 minutes without any specific activity	0.15	0.10	0.10	-0.05	-0.06	327
The teacher correct student's work and give feedbacks [0-3]	1.86	2.00	2.32	0.14	0.46***	327
Children ever left without supervision	0.15	0.20	0.24	0.06	0.09	327
Quality of the engagement of children [0-4]	2.57	2.76	3.25	0.19	0.68***	327
Teacher's awareness of children's individual needs [0-3]	1.41	1.36	1.44	-0.05	0.03	327
Teacher's behavior with respect to gender equality [0-4]	3.27	3.19	3.35	-0.09	0.07	326
Number of schools	207	59	63			

Columns 1—3 show averages by type of preschool. Columns 4-5 show differences between types of preschools. Differences are based on regressions of dependent variables on binary preschool type variables using robust standard errors. Variables with "Quality:" are sorted from 0 (subject not taught) to 4 (subject taught with highest quality). Discrete variables are sorted from 0 (lowest) to highest. Variables with a *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

6.B.3 Additional analyses

Table 6.14: Results for cognition subscales

	(1) Executive function	(2) Language	(3) Early numeracy
Teacher characteristics			
Factor 1: age, education, non-verbal reasoning	0.002	-0.016	-0.017
	(0.023)	(0.023)	(0.021)
Factor 2: salary	0.018	0.092**	0.072*
	(0.039)	(0.038)	(0.038)
Factor 3: training	0.021	0.060**	0.001
	(0.025)	(0.029)	(0.030)
Classroom setting			
Factor 1: children enrolled/present, noise	0.030	0.005	0.044
	(0.024)	(0.026)	(0.028)
Factor 2: instruction time, break time	-0.025	-0.009	-0.051
	(0.033)	(0.034)	(0.034)
Factor 3: curriculum, tracking	-0.010	0.028	0.045
	(0.032)	(0.036)	(0.034)
Equipment			
Factor 1: tables, chairs, blackboard	0.092***	0.068**	0.072**
	(0.034)	(0.032)	(0.032)
Factor 2: learning equipment	0.027	0.029	-0.011
	(0.027)	(0.029)	(0.030)
Factor 3: water, electricity	-0.010	-0.002	0.031
	(0.025)	(0.028)	(0.025)
Curriculum and pedagogy			
Factor 1: art and writing materials used	-0.024	-0.039	-0.044
	(0.031)	(0.029)	(0.032)
Factor 2: math quality and math toys used	-0.040	-0.032	0.023
	(0.033)	(0.028)	(0.029)
Factor 3: literacy/expressive language quality	0.020	0.011	0.003
	(0.024)	(0.027)	(0.024)
Teacher-child interactions			
Factor 1: addressing children's needs	0.014	-0.010	0.013
	(0.025)	(0.028)	(0.024)
Factor 2: enjoyment, negative attitude	0.020	-0.020	0.004
	(0.019)	(0.022)	(0.021)
Observations	2967	2967	2967
Adjusted R^2	0.399	0.500	0.521
Province FE	Yes	Yes	Yes
Enrolment duration	Yes	Yes	Yes
Preschool type FE	Yes	Yes	Yes

This table shows regressions of the endline cognitive development index subscales on preschool quality indicators. Tables shows regressions on preschool quality indicators based on factor analyses. To deal with missing information in control variables these cases are set to a constant value and binary variables indicating missing values are added. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6.15: Results for SDQ subscales

	(1) Emotional	(2) Conduct	(3) Hyperactivity	(4) Peer	(5) Prosocial
Teacher characteristics					
Factor 1: age, education, non-verbal reasoning	0.007	0.013	0.015	-0.011	0.055*
	(0.029)	(0.025)	(0.025)	(0.026)	(0.030)
Factor 2: salary	0.032	0.038	-0.036	0.052	0.006
	(0.062)	(0.049)	(0.044)	(0.056)	(0.048)
Factor 3: training	-0.014	-0.035	-0.016	0.038	0.027
	(0.040)	(0.035)	(0.036)	(0.041)	(0.033)
Classroom setting					
Factor 1: children enrolled/present, noise	0.009	0.021	0.017	-0.023	0.059*
	(0.032)	(0.032)	(0.035)	(0.036)	(0.032)
Factor 2: instruction time, break time	-0.015	0.019	-0.022	0.001	-0.004
	(0.036)	(0.038)	(0.035)	(0.037)	(0.039)
Factor 3: curriculum, tracking	0.031	-0.014	0.023	-0.002	-0.052
	(0.042)	(0.036)	(0.042)	(0.038)	(0.040)
Equipment					
Factor 1: tables, chairs, blackboard	-0.114***	-0.072**	-0.041	0.004	-0.061
	(0.035)	(0.036)	(0.040)	(0.050)	(0.045)
Factor 2: learning equipment	0.009	-0.001	-0.030	-0.009	0.033
	(0.036)	(0.033)	(0.037)	(0.036)	(0.037)
Factor 3: water, electricity	-0.032	-0.037	-0.003	-0.051*	-0.027
	(0.028)	(0.028)	(0.033)	(0.029)	(0.030)
Curriculum and pedagogy					
Factor 1: art and writing materials used	0.028	0.030	-0.034	0.053	0.013
	(0.035)	(0.035)	(0.035)	(0.033)	(0.037)
Factor 2: math quality and math toys used	-0.031	-0.001	0.050	0.028	-0.047
	(0.042)	(0.035)	(0.040)	(0.036)	(0.042)
Factor 3: literacy/expressive language quality	0.006	0.024	-0.003	0.037	0.006
	(0.032)	(0.030)	(0.032)	(0.028)	(0.032)
Teacher-child interactions					
Factor 1: addressing children's needs	-0.022	-0.098***	-0.063**	-0.079***	0.030
	(0.027)	(0.028)	(0.029)	(0.029)	(0.031)
Factor 2: enjoyment, negative attitude	0.049*	0.045	0.027	0.032	0.003
	(0.026)	(0.029)	(0.023)	(0.025)	(0.034)
Observations	2986	2986	2986	2986	2986
Adjusted R^2	0.040	0.096	0.079	0.022	0.067
Province FE	Yes	Yes	Yes	Yes	Yes
Enrolment duration	Yes	Yes	Yes	Yes	Yes
Preschool type FE	Yes	Yes	Yes	Yes	Yes

This table shows regressions of the endline SDQ scocio-emotional problems subscales (emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems) and prosocial behavior on preschool quality indicators. Tables shows regressions on preschool quality indicators based on factor analyses. To deal with missing information in control variables these cases are set to a constant value and binary variables indicating missing values are added. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6.16: Predicting process quality using teacher characteristics

	Curriculum and pedagogy			Teacher-child interactions	
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2
Female	0.287*	0.110	0.138	0.158	-0.103
	(1.71)	(0.62)	(1.15)	(0.89)	(-0.63)
Age	0.00223	0.00502	0.00668	0.00350	-0.00611
	(0.47)	(1.15)	(1.58)	(0.75)	(-1.14)
Potential experience (years)	0.00677	-0.00157	-0.000570	0.00258	-0.000274
	(1.04)	(-0.21)	(-0.11)	(0.33)	(-0.04)
Completed grade 6	0.215*	0.310***	0.00790	0.0715	0.0228
	(1.73)	(2.63)	(0.07)	(0.54)	(0.15)
Completed grade 9	0.0561	0.124	0.0567	-0.0161	-0.191
	(0.52)	(1.28)	(0.65)	(-0.14)	(-1.60)
Completed grade 12	-0.0855	-0.102	0.0978	0.138	-0.00290
	(-0.65)	(-0.80)	(0.82)	(1.04)	(-0.02)
1-4 weeks of teacher training	-0.165	-0.0883	-0.0262	0.220	-0.0202
	(-1.08)	(-0.65)	(-0.21)	(1.49)	(-0.10)
5-8 weeks of teacher training	-0.0452	0.0782	0.485***	0.166	0.0307
	(-0.32)	(0.59)	(4.28)	(1.27)	(0.17)
More than 8 weeks of teacher training	-0.0575	0.305*	0.471***	0.278*	0.209
	(-0.32)	(1.81)	(3.08)	(1.66)	(0.90)
Practical preschool training	-0.118	-0.0678	-0.109	-0.0313	-0.0692
	(-1.09)	(-0.59)	(-1.20)	(-0.24)	(-0.48)
Primary or secondary training	-0.352**	-0.222	0.0721	0.165	0.331*
	(-2.20)	(-1.59)	(0.57)	(1.24)	(1.88)
Nonverbal reasoning (Raven's)	0.116**	0.0793*	0.0799*	0.0742	0.111**
	(2.37)	(1.88)	(1.84)	(1.48)	(2.01)
Monthly salary (USD)	-0.000164	-0.000511	0.000994	0.000478	-0.0000384
	(-0.20)	(-0.67)	(1.35)	(0.63)	(-0.04)
Teacher fully paid, regularly	-0.142	-0.229	0.272	0.493*	0.445
	(-0.43)	(-0.95)	(1.12)	(1.72)	(0.93)
Teacher fully paid, irregularly	-0.0371	-0.0662	0.402	0.534*	0.394
	(-0.11)	(-0.26)	(1.62)	(1.77)	(0.82)
Other work for pay	-0.0496	0.0544	0.209*	0.00854	-0.0323
	(-0.37)	(0.39)	(1.67)	(0.05)	(-0.17)
Length of class (minutes)	0.00457***	0.00245	0.00614***	0.00296	0.000473
	(2.71)	(1.47)	(3.77)	(1.38)	(0.20)
Observations	325	325	325	311	311
Adjusted R^2	0.227	0.187	0.288	0.347	0.159
Province FE	Yes	Yes	Yes	Yes	Yes

This table shows regressions of preschool quality indicators based on factor analysis on teacher characteristics. The curriculum and pedagogy factors capture aspects of art and writing materials used (Factor 1), math quality and math toys used (Factor 2, and literacy/expressive language quality (Factor3). The teacher-child interaction factors capture teachers addressing children's needs (Factor 1) and teacher's enjoyment and negative attitudes (Factor2). Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6.17: Determinants of preschool enrollment

	(1) Ever enrolled at preschool	(2) Total months at preschool
Teacher characteristics		
Factor 1: age, education, non-verbal reasoning	-0.010	0.082
	(0.015)	(0.200)
Factor 2: salary	0.026	-0.041
	(0.020)	(0.323)
Factor 3: training	-0.000	0.084
	(0.016)	(0.233)
Classroom setting		
Factor 1: children enrolled/present, noise	0.025	0.027
	(0.016)	(0.232)
Factor 2: instruction time, break time	-0.023	-0.288
	(0.021)	(0.278)
Factor 3: curriculum, tracking	-0.007	-0.093
	(0.021)	(0.333)
Equipment		
Factor 1: tables, chairs, blackboard	-0.013	-0.282
	(0.021)	(0.277)
Factor 2: learning equipment	-0.007	0.042
	(0.020)	(0.268)
Factor 3: water, electricity	0.020	0.155
	(0.016)	(0.224)
Curriculum and pedagogy		
Factor 1: art and writing materials used	0.038**	0.604***
	(0.019)	(0.227)
Factor 2: math quality and math toys used	-0.046**	-0.740***
	(0.021)	(0.284)
Factor 3: literacy/expressive language quality	0.027	0.623***
	(0.017)	(0.221)
Teacher-child interactions		
Factor 1: addressing children's needs	-0.013	-0.219
	(0.016)	(0.212)
Factor 2: enjoyment, negative attitude	-0.019	-0.234
	(0.013)	(0.152)
F-test (p-value)	0.116	0.040
Observations	5468	5468

This table shows regressions of preschool enrollment and months enrolled at preschool at the time of endline on preschool quality indicators based on factor analysis. Control variables are the number of preschools of each type per village, distance to the closest preschool, child age, child age squared, gender, height-for-age at baseline, household size, caregiver gender, caregiver age, caregiver years of schooling, caregiver non-verbal reasoning score, wealth quantile dummies and baseline test scores. The F-test tests the joint significance of all preschool quality indicators shown. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

6.B.4 Factor analysis

Table 6.18: Factor analysis with structural quality items

		Factor 1	**Factor 2**	**Factor 3**	**PC1**
Teacher characteristics					
1	Female	-0.02	-0.11	-0.01	-0.10
2	Age	**-0.77**	0.12	-0.05	**-0.39**
3	Potential experience (years)	**-0.35**	0.08	0.23	-0.08
4	Highest school grade	**0.61**	0.24	0.10	**0.48**
5	Teacher weeks of training	0.12	0.09	**0.70**	**0.36**
6	Practical preschool training	-0.02	-0.08	**0.57**	0.19
7	Primary or secondary training	-0.03	**0.70**	-0.18	0.25
8	Nonverbal reasoning (Raven's)	**0.54**	0.04	0.01	**0.35**
9	Monthly salary (USD)	0.03	**0.67**	**0.42**	**0.45**
10	Teacher fully paid, regularly	0.06	0.21	0.08	0.19
Equipment					
1	Table and chair for teacher	**0.68**	0.09	0.10	**0.37**
2	Storage for teacher	**0.70**	0.14	-0.03	**0.38**
3	Tables and chairs for children	**0.75**	-0.01	-0.04	**0.33**
4	Children's tables and chairs appropriately sized	**0.60**	-0.09	-0.09	0.24
5	Blackboard/whiteboard and markers/chalks	**0.32**	-0.09	0.17	0.14
6	Electricity access	-0.04	-0.04	**0.42**	0.00
7	Field, playground, or school yard	-0.12	**0.30**	0.17	0.12
8	Equipment for gross motor activities on school yard	-0.10	**0.46**	0.07	0.20
9	First aid kit	0.00	**0.42**	0.16	0.23
10	Functional water source	-0.02	-0.05	**0.79**	0.04
11	Functional drinking water source	0.16	0.02	**0.51**	0.14
12	Hand washing facility	0.02	**0.39**	0.25	0.23
13	Toilet facility	-0.08	-0.04	**0.55**	0.00
14	Writing utensils	0.13	**0.38**	-0.06	0.26
15	Art materials	0.07	**0.49**	-0.11	0.27
16	Fantasy play materials	0.11	**0.52**	-0.15	0.30
17	Educational toys/math materials	0.09	**0.66**	-0.05	**0.35**
Classroom setting					
1	Length of class (minutes)	0.18	**0.66**	0.02	**0.42**
2	Length of breaks (minutes)	0.01	**0.68**	-0.08	0.28
3	Number of children enrolled in this class	**0.66**	0.05	0.01	**0.46**
4	Children present	**0.67**	0.13	0.07	**0.52**
5	Num. of non-teachers adults in classroom	-0.08	-0.06	0.08	-0.06
6	Teacher follows curriculum to teach class	0.05	-0.18	**0.50**	0.20
7	Teacher documents children's development regularly	0.06	0.06	**0.53**	**0.33**
8	Teacher documents children's attendance	0.01	0.09	**0.37**	0.26
9	Class is interrupted at least once	0.20	-0.04	-0.03	0.13
10	Presence of disturbing noise [1-3]	**0.38**	-0.22	-0.07	0.15

Table shows the pattern matrix for exploratory factor analysis using the iterated principal factor method with promax oblique rotation of items (Factor 1-2/3) and principal component analysis (PC),, separately for the aspects teacher quality, equipment and classroom setting. Factor loadings ≥ 0.3 in bold. See Section 6.4.1.2 for additional details about factor analyses.

Table 6.19: Factor analysis with process quality items

		Factor 1	Factor 2	Factor 3	PC1
Curriculum and pedagogy					
1	Quality: Mathematics	-0.01	**0.39**	-0.06	0.26
2	Quality: Literacy	0.08	-0.22	**0.51**	0.10
3	Quality: Exressive language	-0.01	-0.02	**0.59**	0.21
4	Quality: Storybook	-0.15	0.23	0.23	0.15
5	Quality: General knowledge	-0.07	0.23	0.23	0.22
6	Quality: Fine motor skills	0.10	0.19	0.17	0.29
7	Quality: Gross motor skills	0.10	0.08	0.22	0.23
8	Educational toys/math materials used by children	0.18	**0.48**	-0.10	**0.42**
9	Fantasy play materials used by children	0.10	**0.34**	-0.15	0.27
10	Art materials used by children	**0.64**	0.10	0.01	**0.49**
11	Writing utensils used by children	**0.61**	0.01	0.04	**0.43**
Teacher-child interactions					
1	The teacher enjoyed teaching [1-3]	0.09	**0.73**		**0.38**
2	The teacher showed negative attitudes [0-2.5]	-0.12	**0.67**		0.24
3	Quality of the disciplinary strategies used by the teacher [0-4]	**0.45**	**0.33**		**0.43**
4	More than 4 occurences of negative interactions	-0.04	**0.32**		0.14
5	More than 8 occurences of encouragements	**0.44**	0.00		0.30
6	Children wait more than 5 minutes without any specific activity	-0.19	-0.04		-0.16
7	The teacher correct student's work and give feedbacks [0-3]	**0.53**	-0.12		0.29
8	Children ever left without supervision	-0.10	-0.06		-0.11
9	Quality of the engagement of children [0-4]	**0.47**	0.12		**0.36**
10	Teacher's awareness of children's individual needs [0-3]	**0.68**	-0.05		**0.39**
11	Teacher's behavior with respect to gender equality [0-4]	**0.46**	0.03		**0.32**

Table shows the pattern matrix for exploratory factor analysis using the iterated principal factor method with promax oblique rotation of items (Factor 1-3) and principal component analysis (PC), separately for the aspects curriculum and pedagogy and teacher-child interactions. Factor loadings ≥ 0.3 in bold. Variables with "Quality:" are sorted from 0 (subject not taught) to 4 (subject taught with highest quality). Discrete variables are sorted from 0 (lowest) to highest. See Section 6.4.1.2 for additional details about factor analyses.

Figure 6.1: Preschool quality histograms for indices based on first principal component and factor analysis

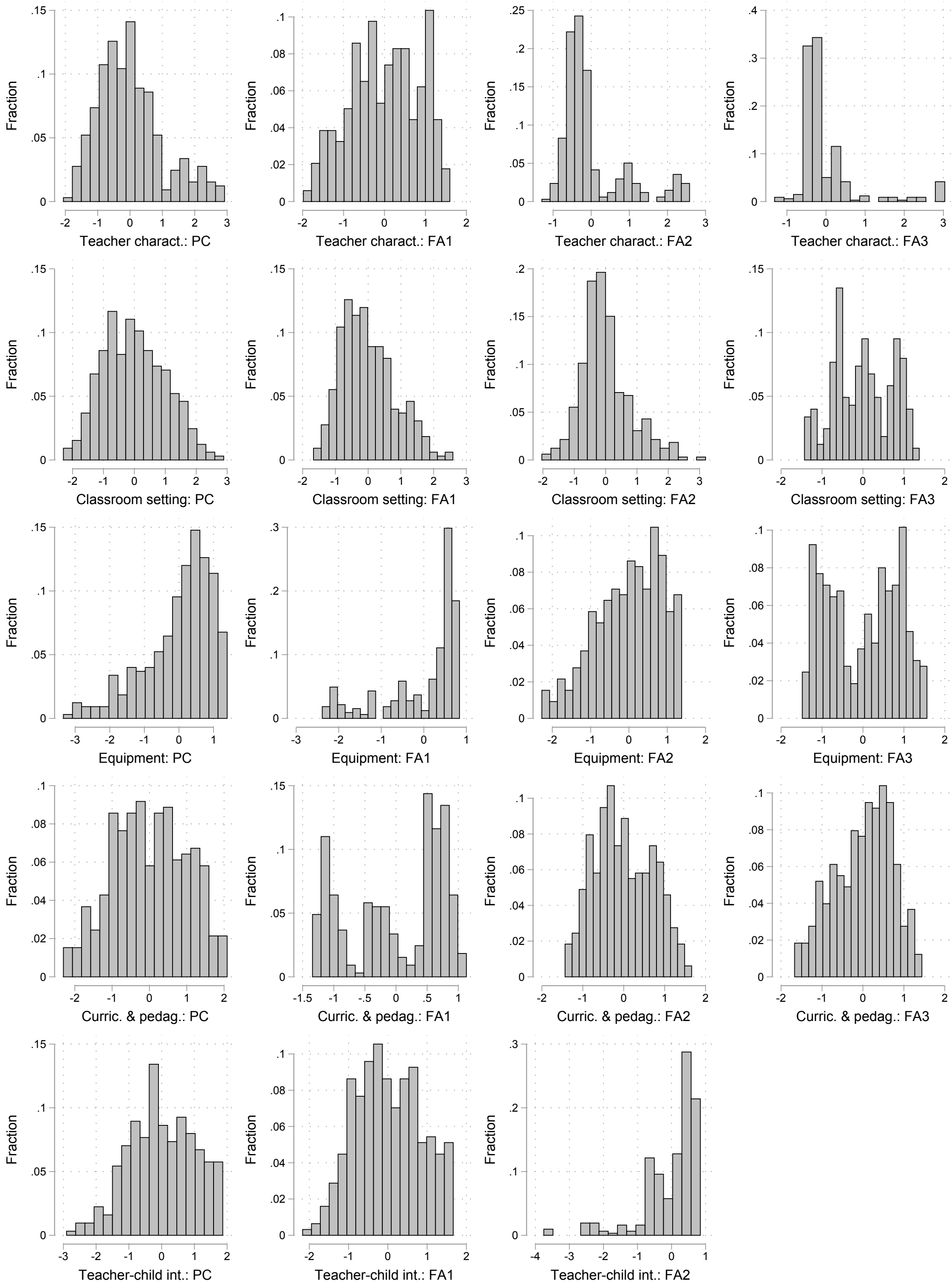


Figure shows distribution of preschool quality indices based on first principal component and factor analysis.

Figure 6.2: Preschool quality histograms by type of preschool

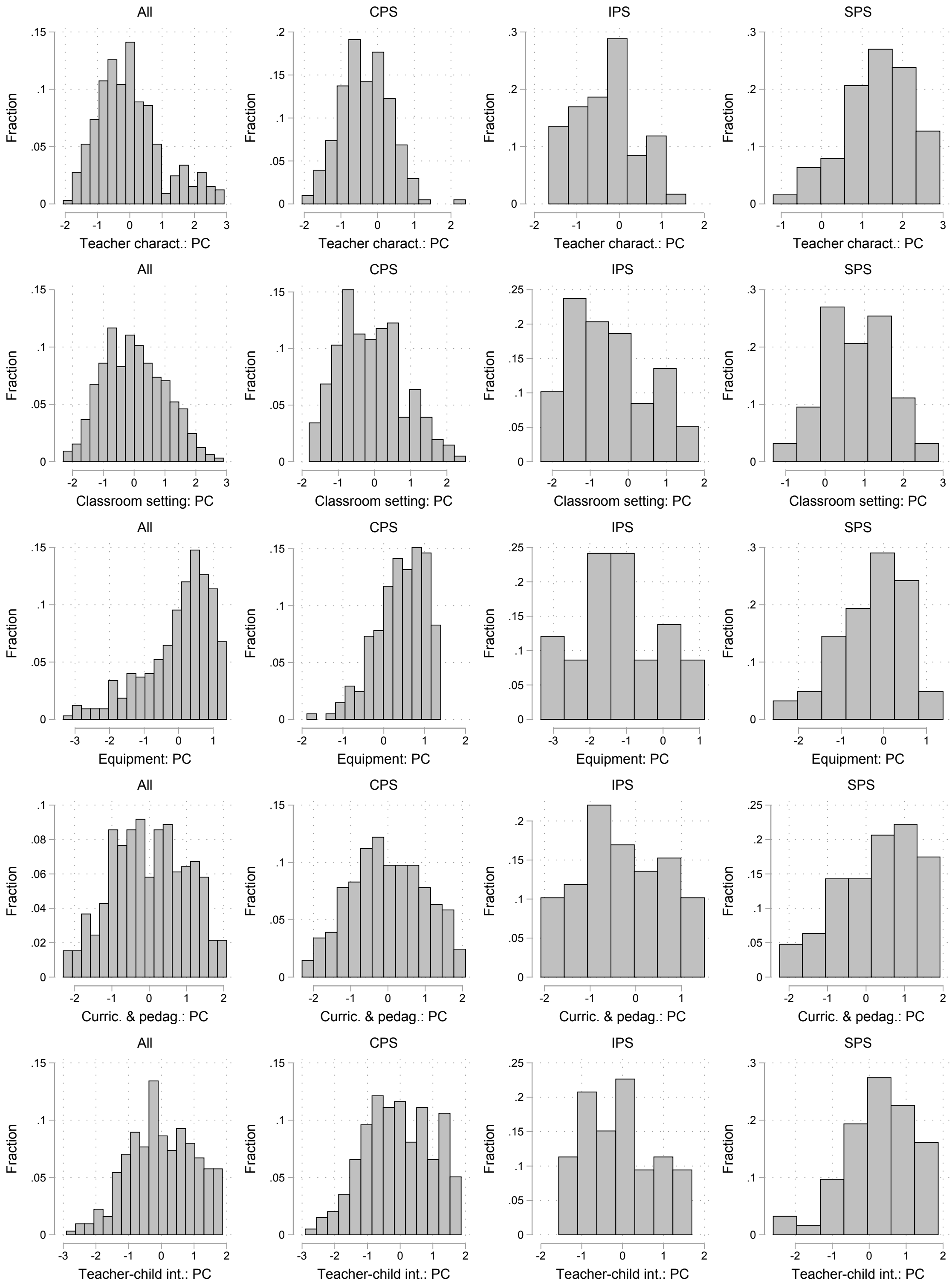


Figure shows distribution of preschool quality indices based on first principal component by preschool type.

CHAPTER 7

Joint Roles of Parenting and Nutritional Status for Child Development: Evidence from Rural Cambodia[1]

7.1 Introduction

Early childhood development sets the course for lifelong learning, health, and well-being (Black et al., 2017). Despite the potential human capital that can be built through investments in early childhood development (e.g. Heckman, 2006), recent estimates suggest that 250 million children living in low- and middle-income countries are at risk of not fulfilling their developmental potential (Black et al., 2017). Family poverty is associated with several conditions that negatively affect child development, as documented in an extensive body of research. In sum, children who experience early stunting, poor health, or lack of educational opportunities are less likely to achieve their developmental potential than other children (Britto et al., 2017), with a common underlying negative influence of family poverty. Family poverty has been shown to have a wide range of adverse effects on child development in many countries, with gradients in child development due to family income evident as early as age 4 months and increasing in size throughout the course of childhood (Paxson and Schady, 2007; Naudeau et al., 2011; Fernald et al., 2012; Schady et al., 2015; World Bank, 2018; Hamadani et al., 2014).

[1]This chapter is joint work with Abbie Raikes (University of Nebraska Medical Center), Adrien Bouguen (University of Santa Clara) and Deon Filmer (World Bank). The authors thank Tsuyoshi Fukao and members of the Early Child Development Department of the Royal Government of Cambodia's Ministry of Education for input and assistance. Useful comments from Sophie Naudeau and Magdalena Bendini on an earlier version of this paper are gratefully acknowledged. This study benefited from funding from the World Bank's Strategic Impact Evaluation Fund (TF0A0326) and the Ministry of Education, Youth and Sport of Cambodia. The authors are responsible for any errors. The findings, interpretations, and conclusions expressed in this paper are those of the authors and do not necessarily represent the views of the World Bank, its executive directors, or the governments it represents.

The mechanisms accounting for the association between family wealth and child development remain imperfectly described, yet clarifying these mechanisms is central to building effective interventions. The association between family wealth and child development has been hypothesized to include a range of factors, including health and nutrition as well as environmental stimulation, such as parenting quality and the degree of stimulation and emotional support in home environments (e.g. McLoyd, 1998; Black et al., 2017). This study focuses on the contributions of parenting and nutrition to child development in the preschool years.

7.1.1 Undernutrition, parenting, and child development

7.1.1.1 Undernutrition.

Undernutrition, evidenced by shortened stature or stunting within the first years of life, poses a formidable risk to child development in low-income countries. A large body of work demonstrates strong associations between early nutritional status and child development (Miller et al., 2016; Sudfeld et al., 2015; Prado and Dewey, 2014). Children with poor early nutrition, including inadequate caloric intake and the lack of micronutrients, are at risk of failing to reach expected linear growth milestones, leading to height stunting, long-term cognitive and motor deficits, and eventually worse economic outcomes (e.g. Galasso et al., 2016; Sudfeld et al., 2015).

Although the overall negative association between stunting and child development is well documented, the extent to which stunting influences long-term child development is dependent on several factors, specifically the timing and severity of early nutritional deficits and the degree of environmental stimulation (Prado and Dewey, 2014). Despite the formidable associations between early stunting and later deficits in learning (e.g. Sudfeld et al., 2015), the associations between stunting and later poor outcomes can be mistaken as casual (e.g. Leroy and Frongillo, 2019), nor it is not yet clear how stunting exerts its strong and negative impact on child development or whether the degree of stimulation in children's environments may mitigate or exacerbate this association (Black et al., 2019). Environmental stimulation and nutritional status may work in an additive manner, meaning that poor quality environments and nutritional status may contribute unique effects to child development. Conversely, they may interact, that is, stimulating environments may counteract the impacts of poor nutrition, or the impact of stimulating environments may only be evident for children with good nutritional status.

7.1.1.2 Parenting.

Parents play a central role in providing young children with stimulating environments. In interactions with young children, parents engage in multiple forms of teaching and support, by which several elements of child growth and development are facilitated (e.g. Shonkoff et al., 2000). Three types of parenting have been identified as important predictors of children's well-being: cognitive stimulation, sensitivity and responsiveness, and emotional warmth (Bornstein and Putnick, 2012; Kuklina et al., 2006). Cognitively stimulating parenting refers to parents' stimulation of learning through use of language, introduction of new games and ideas, as well as exposure to high-quality learning environments. Emotionally responsive parenting, including warmth and sensitivity, refers to parents' ability to respond sensitively to children, verbally and physically (Bornstein and Putnick, 2012).

Parenting quality, or the extent to which parents engage in stimulating, warm, and responsive interactions with young children, has strong, long-term, and consistent impacts on child development, as evidenced through impact evaluations of parenting programs and observational studies (Britto et al., 2017; Jeong et al., 2017; Bradley and Corwyn, 2002). Despite evidence on the overall effects of parenting, some reviews have emphasized the need for more documentation of parenting program impacts on behavior and emotional difficulties (e.g. Mejia et al., 2012) as well as variance in the size of the association between parenting and child development across countries (Jeong et al., 2017). Parenting quality has been shown to be a central mediator of the relationship between family poverty and negative child outcomes (Hamadani et al., 2014; Blair and Raver, 2012). Parenting quality is lower among families living in poverty, due to the stress of living without adequate economic resources and the concomitant characteristics of poverty, such as low levels of parental education, which in turn contribute to the negative impacts of poverty on children (McLoyd, 1998; Bornstein and Bradley, 2014). Thus, a key hypothesis is that parenting quality accounts for a portion of the variance in the association between family poverty and child development. As evidence toward this hypothesis, Hamadani et al. (2014) report that wealth gradients in child development increased sixfold from ages 7 to 64 months, and a large portion of the wealth gradient that emerged in cognitive development over the first five years of life in children in Bangladesh was accounted for by differences in parent cognitive stimulation, parental education, and linear growth.

Existing research suggests some degree of differentiation between domains of parenting: parents who are emotionally responsive are not necessarily cognitively stimulating; moreover, emotionally responsive parenting is especially important for socioemotional development, while cognitively stimulating parenting promotes young children's learning and development. *Negative parenting* or reliance on harsh, punitive physical and verbal parenting, relates to disruptions in socioemotional development and delayed cognitive development (Chang et al., 2003; Hughes and Ensor, 2006). Maternal education, specifically mothers' executive function, is related to the development of children's executive function (Cuevas et al., 2014) and associations between family resources, including maternal education, and overall parenting quality have also been reported (e.g. Lugo-Gil and Tamis-LeMonda, 2008). Using a very brief self-report survey questionnaire of parenting practices that did not include observations of actual parenting behavior, a tendency for parents to engage in more emotionally supportive parenting than cognitively stimulating parenting has been documented across countries (Bornstein and Putnick, 2012). However, it is also important to note that parenting is deeply culturally embedded, reflecting locally held values, priorities, and traditions, and dynamic. Thus, conceptualizations of parenting from other countries may not apply when taken to a new context (Selin, 2013). Replication of these findings is necessary to determine if and how previously described constructs of parenting apply across countries.

Research focus on parenting, nutrition, and child development is also warranted because the relationships between parenting, child nutrition, and child development are multi-factorial and complex. Several studies have examined the hypothesis that undernutrition arises in part from poor quality parenting, particularly mediated through lack of maternal responsiveness. The results have been inconclusive. Maternal depression and other common mental disorders in the first two years after a child's birth have been associated with child underweight and stunting (Nguyen et al., 2014; Surkan et al., 2011). This effect may be mediated through lack of engaged and stimulating parenting, which in turn influences child feeding and growth, but the sources of this association are not yet clear. A recent systematic review of interventions focused on improving responsive feeding yielded few studies documenting positive impacts of responsive feeding on child undernutrition (Bentley et al., 2011), but the review also found that parental positive vocalizations during feeding could increase child acceptance of food. In sum, although the literature suggests that maternal mental disorders and undernutrition may be associated, and that specific parenting behaviors during feeding could be critical for addressing undernutrition, a general association between parenting quality defined broadly and undernutrition is not consistently documented in existing studies. Instead, the studies suggest that parenting and undernutrition may interact with one another or exert independent effects on child development.

There is empirical justification for hypothesizing that there are independent and interaction effects between environmental stimulation, including parental stimulation, stunting, and child development. In support of the hypothesis that nutrition and parenting exert unique influences on child development, even when children have experienced early stunting, sensitivity to environmental stimulation is consistent throughout the first five years of life (Black et al., 2019). There is also evidence of an interaction effect between parenting and nutrition. Cognitive stimulation, as delivered through parenting, may ameliorate some of the effects of poor early nutrition on child development (Paxson and Schady, 2007; Grantham-McGregor et al., 1991), and stimulating home environments may mitigate the association between stunting and child development for infants and toddlers (Black et al., 2019; Nguyen et al., 2017), but a similar effect was not noted for preschoolers (Black et al., 2019). Thus, the role of environmental stimulation in ameliorating the negative impacts of stunting may vary by child age. For example, in a large-scale sample in Peru, Cueto et al. (2016) found that more nourished children benefitted more from preschool, also suggesting that impacts of environmental stimulation on cognitive development in stunted children may change as children grow older.

However, gaps in the research remain. Few studies have examined how early nutritional status interacts with later environmental stimulation to influence child development (Black et al., 2015; Prado and Dewey, 2014). Yet, this hypothesis is consistent with a burgeoning body of work arguing for the importance of examining the interactions between biological and environmental factors on child development (e.g. Belsky et al., 2007; Ellis et al., 2011). Although interventions may be most effective—and needed—among children facing the most substantial risks, it is also possible that children with fewer skills are also less likely to benefit from the stimulation in their environments, leading to gaps that persist and grow over time, even when environmental resources are enhanced. Little work to date has examined this interaction in the context of environmental conditions that characterize child development in low- and middle-income countries. If more nourished children are more responsive to environmental stimulation, more work may be needed to design effective interventions for children who are undernourished.

The present study examines the emergence of wealth gradients in child development and the joint roles of nutritional status and parenting in explaining the gaps between high- and low-wealth children. We had four hypotheses. First, we expected to find wealth gradients in child development based on family wealth in a low-income country, and that parenting and nutritional status would account for some portion of the wealth gradients in child development. Second, we expected to find that there are different types of parenting, such as cognitively stimulating and emotionally responsive. Third, we hypothesized that different types of parenting would have unique implications for child development, and that all types of parenting quality would show associations with family wealth and child development. We did not expect to find that parenting quality would be directly related to child nutritional status. Fourth, we hypothesized that parenting and nutritional status would interact in associations with child development, and that parenting quality would be differentially related to child development depending on nutritional status. In sum, we hypothesized that family assets would exert a strong and consistent impact on child development, but that parenting quality and nutritional status could account for a portion of the variance attributable to family wealth, as demonstrated by a reduction in the impact of family assets on child development when accounting for parenting and nutritional status.

7.2 Methods

7.2.1 Participants

We investigated our hypotheses using data from a study of children in rural Cambodia, using a unique panel data set of 6,508 children (ages 2-4 in 2016) and their families, collected in 2016 and 2017. Participants were recruited in 2016 as part of a large-scale study of early childhood development. The panel study was designed to investigate the effects of a preschool construction program on primary school readiness.[2] Preschools were randomly assigned among 305 rural villages located in 13 provinces in Cambodia.[3] The eligibility criteria for villages to participate in the program were a high poverty rate and a high number of children between ages 0 and 5 years. Thus, the sample was not meant to be representative of Cambodian households. Instead, it covered households with children at preschool age in rural villages of the 13 provinces.

During the baseline data collection exercise in 2016, a survey firm[4] used an adapted version of the EPI walk to select up to 26 eligible households in each of the 305 vil-

[2]The study was approved by the institutional review board of the Paris School of Economics and the National Ethics Committee for Health Research by the Cambodian Ministry of Health. Our main outcomes for this study, collected in 2017, stem from a first follow-up on the randomized trial intervention. Between the baseline and the first follow-up, community preschools were constructed in some of the sample villages. An awareness campaign was conducted to stimulate enrollment in the new preschools. Further, a home-based program, in which trained core-mothers address a range of topics around parenting, was implemented. These interventions were randomized between villages and still at an early stage during the time of the first follow-up. We therefore ignore the interventions in this analysis. We show in the sensitivity analyses (Appendix A3) that our regression estimates are robust to the inclusion of village fixed effects that would absorb potential effects of the randomized treatment.

[3]The 13 provinces are located in the south and northeast of Cambodia: Kampong Chhnang, Kampong Speu, Kampot, Kandal, Koh Kong, Kratie, Mondulkiri, Preah Sihanouk, Prey Veng, Ratanakiri, Sung Treng, Svay Rieng, and Takeo.

[4]Angkor Research and Consulting Ltd., under the close supervision of the Principal Investigators.

lages.[5] To be eligible, households must have had at least one child between ages 24 and 59 months. A total of 7,053 households were invited to participate in the study, of which 6,934 completed the entire baseline interview (98.3%). Approximately 11 months after the baseline data collection, the same survey firm completed interviews with 6,457 (91.5%) of the baseline households.[6] This study relies on the follow-up sample, permitting control for baseline test scores in regressions and observations of outcomes during an age period that is critical for the development of language and early numeracy skills.

The sample for this study includes all households that participated in the baseline and follow-up, and for which the caregiver interview and child assessment were completed. In total, these are 6,290 households, with 6,718 tested children age (M=3.92) 2 ($n = 176$), 3 ($n = 2,191$), 4 ($n = 2,396$), 5 ($n = 1,921$), and 6 ($n = 34$). Hence, the majority of children at follow-up were at the target age for preschools in Cambodia (ages 3 to 5 years). The average number of households per village in the follow-up sample is 21.2 and the average number of eligible children per household is 1.07. The major ethnic groups in the sample are Khmer (91.04%), Cham/Khmer Islam (2.10%), Pnong (1.98%), Lao (1.31%), and Tumpoung (1.06%). The languages spoken in the households include Khmer (94.66%), followed by Lao (1.53%), Pnong (1.22%), Cham (0.92%), and Tumpoung (0.89%). Trained local translators joined the caregiver interviews and child assessments when the respondent did not speak Khmer.

7.2.2 Measures

We developed and adapted the battery of child tests in cooperation with the researchers of the Measuring Early Learning Quality and Outcomes project (MELQO) (UNESCO, 2017) and the local survey firm. The MELQO tools were designed to provide a starting point for national-level adaptation of global measures of child development (see UNESCO (2017), for an overview) and have demonstrated adequate validity (Raikes et al., 2019).

Household survey. The household survey included questions on family structure and socioeconomic background. We constructed a dwelling quality index and household assets index.[7] Both indexes were then standardized and summed to create a wealth in-

[5] EPI refers to the Expanded Programme on Immunization of the World Health Organization (see e.g., Henderson and Sundaresan, 1982).

[6] The attrition rate of 8.5% can almost entirely be explained by seasonal and permanent relocation of households. The study did not follow up households that moved beyond the boundaries of the sample villages.

[7] The direction of each categorical variable for dwelling quality (quality of the floor, quality of the roof, electricity, toilet. . .) was first assessed using a Multiple Correspondence analysis (MCA). Then,

dex that was used to describe quintiles in further analyses. For easier interpretation of the regression estimate, we constructed a binary measurement of household wealth and define low living standard as belonging to the lowest two quintiles of the wealth index. Household size ($M = 4.48, SD = 1.68$) and the binary measure of household wealth ($M = 0.40$) were used as measures of socioeconomic background at the household level.

Caregiver survey. The educational background of the caregiver is measured by years of schooling ($M = 3.48, SD = 3.15$) ; a binary indicator indicating completion of primary school (six years) ($M = .266$) was used in the regression analyses for better interpretability. Caregiver non-verbal reasoning was measured using a set of 12 Raven's Progressive Matrices (Carpenter et al., 1990; Raven, 2000). This test was also administered at baseline. When administered to the same caregiver again 11 months later, non-verbal reasoning scores showed a correlation of ($r_{sp} = 0.46$).

To account for the multidimensionality of parent-child interactions (Bornstein and Putnick, 2012), we developed a questionnaire that covers multiple aspects of parenting and is suitable for assessing the domains of parenting in the studied context. During baseline pretests, we combined questions from parenting questionnaires used in projects in Cambodia (Bouguen et al., 2018a), Indonesia (Brinkman et al., 2017), and the Lao People's Democratic Republic (Brinkman et al., 2016). Questions that were not well understood by parents were adapted during baseline pretests. After baseline, we assessed the reliability of our measures using exploratory factor analysis and Cronbach's alpha. Although the exploratory factor analysis already indicated a three-factor solution, questions for some domains were underrepresented at baseline and some intraclass correlations were low. We added additional questions related to parental warmth and lovingness, as well as overburdened or distressed parenting (later called socioemotional and negative parenting) before follow-up. We also decided to change the scale from a subjective (never/almost never, rarely, sometimes, regularly, always/almost always) to a more objective scale (never or almost never, a few times in the month, about once a week, several times a week, every day or almost every day), since we were worried about a nonstandard measurement error caused by different perceptions (e.g., reading to a child "regularly"). The updated instrument showed better reliability in the pretests and main data collection of the follow-up.[8]

each variable was coded using the sign of the coordinate of the first MCA dimension. We then took the average of the standardized version of each coded variable. The Cronbach alpha is 49%. For the household assets, we simply standardized each asset and took the standardized version of the sum. The Cronbach alpha is 65%.

[8]We report the full list of questions in Appendix 7.C.1.

In the final parenting questionnaire at follow-up, caregivers answered 25 items related to regular activities, parent-child bonding, punitive behavior, and child-rearing methods. Exploratory factor analyses, further described in the analyses plan and results section, were used to identify unique dimensions of parenting quality within the parenting scale. Parenting composite scores were constructed for each factor retained, by summing all the items loaded on one factor and standardizing the sum obtained with the sample mean and standard deviation.

Child survey and assessment. We defined child age as the number of days between birth and follow-up, divided by 365.25. Stunted growth was defined as a height-for-age z-score (HAZ) two standard deviations below the mean, using the growth standards of the World Health Organization (2006) ($M = .380, r_{sp} = .706$) (see Appendix 7.C.3 for sensitivity analyses using cutoff points other than -2 SD). The child assessment includes different tests for the domains of language development, early numeracy, and executive function. Most of the child tests are based on the MELQO toolkit (see UNESCO (2017), for a description of the measures of development process, and Raikes et al. (2019), for evidence of validity).

Tests were added to the MELQO items to increase the sensitivity and breadth of the child assessment. The additional tests included the following: the Dimensional Change Card Sort test (Zelazo, 2006); a receptive vocabulary test based on picture recognition derived from the Test de Vocabulario en Imagenes Peabody (TVIP), which is a version of the Peabody Picture Vocabulary Test adapted for Spanish-speaking populations and was normed to low-income populations in Mexico and Puerto Rico (Dunn et al., 1986; Dunn and Dunn, 2007); a test for knowledge of reading concepts (based on a monitoring tool used by the Cambodian Ministry of Education's Early Child Development Department); and a sustained attention test. Children's socioemotional development was measured using the caregiver-reported Strengths and Difficulties Questionnaire. Since the overall purpose was to generate a well-functioning test for Cambodia (as opposed to maintaining consistency for other purposes, such as international comparisons), adaptations were prioritized to ensure adequate fit to the local context.[9]

Before constructing the composite scores of child test domains, individual tests were first scored and standardized, thus ensuring that all the tests contributed equal variance to the composite score. For almost all the tests, scoring was done by assigning one point for each correct response and summing these points to create an individual score for each test. Exceptions to this method are mentioned below. When a child

[9]Details on the cultural adaptations, pretesting procedures, bivariate correlations, and empirical frequency distributions of raw test scores are presented in Appendix 7.C.2.

was unable to complete the practice trial of a test, a score of zero was assigned for this test, as long as the child participated in the other tests. Standardization of each test score was done by subtracting its sample mean and dividing it by its sample standard deviation. All standardized test scores of one domain (e.g., executive function) were then summed into a domain score and standardized again by subtracting its sample mean and dividing by the sample standard deviation of the domain score for better interpretability. After these steps, we obtained the following composite scores.

Executive function ($\alpha = .74, r_{sp} = .60$).[10] The final executive function score includes four tests, of which the first two listed here are from the MELQO toolkit. The construct inhibitory control is assessed with the head-knee task (McClelland et al., 2014). The test has two stages. In the first stage, the child stands in front of the enumerator and is asked five times to touch his/her head or knees. In the second stage, the child is asked to do the opposite of what the enumerator says. Working memory (short-term auditory memory) is assessed with a forward digit span test in which children have to repeat sequences of digits, which increase in length. The Dimensional Change Card Sort test is used as a measure of cognitive flexibility. We followed the procedures outlined in Zelazo (2006) using cards with two colors (blue and red) and two pictures (boat and rabbit). To reduce the burden on tested children, we followed the protocol with the exception that children needed to pass the pre-switch phase (at least 5 out of 6 correct) to participate in the post-switch phase. The border version of the test was not administered. The demonstration phase of the test included one practice trial. As per protocol, this practice trial was not used to determine whether a child was eligible for the test, as the child could have performed well by chance.

We use a self-developed cancellation task to measure sustained attention. In this test, children see a printed matrix with different symbols and are asked to cross out all the symbols that match the given one (e.g., cross out all flowers). When completed, a larger matrix is given, and a new symbol has to be crossed-out. The test continues until the child has completed four matrices, crossed out more wrong than correct images in a matrix, loses attention, or states that he or she is done. The test was scored by using the difference between correctly and incorrectly crossed-out images.

Language development ($\alpha = .71, r_{sp} = .65$). Receptive vocabulary skills are assessed with a test derived from the TVIP. In this test, the child is asked to match a word to one of four pictures. The version used in the Cambodian context was culturally

[10]Cronbach's α for the composite test scores is based on the covariance between the standardized test scores of the individual tests. The rank correlation coefficient for the composite test scores gives the correlation between baseline and follow-up composite scores with 11 months between on average.

adapted during piloting and validation exercises prior to baseline data collection and with the support of key informants. The final instrument includes 82 pictures, with the rule that the test stops after six of the last eight pictures were wrong.[11]

All other language development tests were taken from the MELQO. Expressive language skill is assessed by asking the child to name up to 10 things that can be eaten and up to 10 animals he or she knows. The final score is the number of recalled items. Receptive language is assessed with a listening comprehension test in which a short story (116 words) is read to the child. After reading the story, the child is asked five questions about the content of the story. Knowledge of reading concepts is assessed by showing a children's storybook and asking how the book should be opened and where and in which direction one should start reading the story. Reading skills are assessed with a letter name knowledge test in which the child has to identify common letters of Khmer script. A drawing test, where the child copies shapes, like circles or squares, is used to assess fine motor skills. This test was used to replace a letter writing test, which turned out to be too difficult during the pretests. Since the ability to copy simple shapes is a prerequisite for literacy skills, such as writing letters, we decided to keep the test in this category, although it is a fine motor skill test.

Early numeracy ($\alpha = .73, r_{sp} = .56$). The tests for early numeracy include a self-developed test for measurement concepts, for example, if the child understands concepts such as tallest/shortest, in which the child had to point to different printed objects. In a test for verbal counting, the child was asked to count as high as he or she could and was stopped at reaching 30. Numbers and operations were also administered with a self-developed quantitative comparison test in which the child had to compare the number of printed objects on two sides of a page. A number identification test analogous to the letter name knowledge test was used. A self-developed shape recognition test was used to test whether the child could identify basic geometric shapes.

Socioemotional problems ($\alpha = .69, r_{sp} = .26$). To test for behavioral problems, we use the total difficulties score of the Strengths and Difficulties Questionnaire. The score is composed of five questions for each of four categories: emotional symptoms, conduct problems, hyperactivity/inattention, and peer relationship problems. This test for socioemotional problems is the only child development measure that is based on the caregiver survey instead of the direct child assessment. The Strengths and Dif-

[11] As for all the tests, the raw scores of the receptive vocabulary test are standardized with the sample mean and standard deviation. We do not attempt to use standardized scores using reference populations from different settings, as this would imply benchmarking against noncomparable reference populations who took, essentially, a different test in a different language.

ficulties Questionnaire results are not reversed: a higher score means more behavioral difficulties.

7.2.3 Procedure

The survey firm recruited survey interviewers based on their familiarity with data collection and experience with young children. To ensure that the children and parents correctly understood the questionnaire and the instruments were reliable, the research team pretested every instrument at least three times before collecting data in the sample villages. The survey firm translated the questionnaires into Khmer, and an independent third party back-translated them into English, which led to further refinements in the instruments. The research team participated in the interviewer training conducted by the local firm's fieldwork manager. Field staff was organized into six groups, each comprising of four interviewers, one supervisor, and one field editor. All the supervisors had several years of data collection experience in Cambodia and were responsible for household sampling and quality control procedures. The editors supported the supervisors doing spot-checks and interviewer observations and conducted independent re-interviews of at least 20% of the interviews in each village.[12]

Upon arrival at a household's home, the enumerator first identified the main caregiver and the eligible children,[13] and then administered the household and caregiver surveys as well as completed the anthropometric measures and development tests. The child's testing session lasted an average of 31 minutes ($p50 = 30$).

7.2.4 Data Analyses Plan

The data analyses plan was designed to test the following research questions, which reflect our initial hypotheses: (1) are there gradients in child development based on family wealth for children ages 3 to 5 years, and do parenting and nutritional status account for a portion of the variance in child development due to wealth gradients; (2) does parenting show unique components and, if so; (3) how do parenting components differentially relate to undernutrition, family wealth, and child development; and (4) how do parenting and nutritional status interact in association with child development? All the analyses were conducted with follow-up variables as dependent variables. Independent variables were measured at follow-up as well, with the exception of the wealth

[12]Initial re-interviews were limited to a set of 20 questions to reduce the burden on the respondents. In the case of inconsistencies between the initial interview and the re-interview, households were called or visited again to obtain clarity.

[13]The caregiver is defined as a direct relative who takes care of the child most of the time. In our sample, the primary caregiver is most often the mother (52% of the children), followed by the grandmother (35%), father (6%), grandfather (2%), aunt or uncle (3%), or an older sibling (1%).

index, living standard measure, and baseline test scores.

Research question 1: Wealth gradients in child development. To assess the presence of wealth gradients within the four measures of child development, we used a regression model with wealth quintile dummies 2 to 5 ($Quintile_j$), with quintile 1 being the left out category; age category dummies (Age_k) for children ages 4 and 5, with age 3 being the left out category; and interaction terms:

$$\begin{aligned}\text{Age-adjusted test score} =& \alpha + \beta_4 \text{Age}_4 + \beta_5 \text{Age}_5 \\ &+ \sum_{j=2}^{5} \left(\gamma_j \text{Quintile}_j + \sum_{k=4}^{5} \phi_k \text{Quintile}_j \text{Age}_k \right) + \eta' \text{W} + \epsilon.\end{aligned}$$

Age-adjusted test scores are residuals from a regression of test scores on the first three polynomials of age as defined in the method section, standardized with their sample mean and standard deviation. This approach ensures that wealth gradients refer to test scores with a standard deviation of 1, independently of whether it is additionally controlled for age. Before estimating the full model, we first ran a regression just on quintile dummies to describe the wealth gap for all the children ages 3 to 5. Second, we estimate a specification that adds the age category dummies and interaction terms. Although age category dummies alone do not explain any variation when the outcome is mean-independent from age, the dummies were used to test for differences between wealth gaps among children at different ages. For example, testing $\gamma_5 = 0$ is equivalent to testing whether there exists a test score gap between the highest and lowest quintiles among children at age 3. Testing $\phi_4 = 0$ and $\phi_5 = 0$ allowed us to test whether the quintile wealth gap at ages 4 and 5 is larger or smaller than at age 3. Testing $\phi_4 = \phi_5$ allowed us to test whether the gap at age 4 is different than at age 5. In a third regression (the full model), we used additional control variables W (parenting measures or HAZ scores) to see how much of the wealth gradients can be explained by them.

Research question 2: Components of parenting. We conducted exploratory factor analyses to determine whether parenting data could be grouped by distinct dimensions of parenting. Sampling adequacy was supported by a Kaiser-Meyer-Olkin statistic of 0.868 (Kaiser, 1974), confirming a compact pattern of correlations between the 25 parenting items. Bartlett's test for sphericity rejected the null hypothesis of zero covariance, $p < .001$. Parallel analysis suggested an 11-factor solution, equaling the number of factors with eigenvalue greater than zero. We followed a similar procedure as Jones et al. (2017) to reduce the number of factors retained. We rejected a large number of factors, as only three of them had more than one item with a factor loading greater than 0.3, and all but three factors had poor reliability ($\alpha < .6$) . We gradually reduced the number of factors retained until all factors had reliability of at least $\alpha \geq .6$ and at least two items with loadings greater than 0.3.

Research question 3: Associations of parenting components with nutritional status, family wealth, and child development. To assess the extent to which the three parenting scores capture meaningful differences in parenting behavior, we first correlated the parenting composite scores with each other and with family characteristics and child outcomes using Sidak corrections for multiple comparisons. Since parenting is significantly associated with several other household characteristics, isolating the independent role of parenting for child development is important. Therefore, we also explored the roles of parenting, nutrition, and their joint roles in explaining variation in child development conditional on other variables, using a hierarchical regression model. The full model is given by:

$$Test\ score = \alpha + \beta_1' X_1 + \beta_2' X_2 + \gamma' Parenting + \delta Stunting + \phi'\ Test\ score_{baseline} + \epsilon$$

In the first step, test scores (not adjusted for age) were regressed on age and gender (X_1) , to show that these variables account for a large fraction of the variance in our outcomes. In the second step, measures of household size, low living standard, caregiver age, primary education, and nonverbal reasoning skills were added to the equation (X_3) . In the third step, we separately added our three parenting measures and stunting indicator to the regression to determine their unique contributions to child development. Differences in the β_2 parameters were examined to gauge how much of their effect can be explained by parenting and stunting. In the next step, baseline test scores were added to the regression to address concerns about correlation of the error term with the parenting scores. If a correlated component of the error term is already reflected in the baseline test scores, estimates of a causal effect can be obtained by controlling for baseline test scores. Although the interpretation and point estimates of our coefficients change, since we also control for effects of variables captured by the baseline test scores, the baseline test scores reduce the set of potential confounders and the model with baseline test scores is our preferred specification.

Research question 4: Joint associations of parenting and stunting with child development. In the last step, we examined the joint associations of parenting and stunting by first adding a cognitive parenting-stunting interaction term to the model. In a second step, we added a caregiver education-stunting interaction term to the model. The full model is given by:

$$
\begin{aligned}
Test\ score = & \alpha + \beta_1^{'}X_1 + \beta_2^{'}X_2 + \delta_1^{'}Parenting + \delta_2 Stunting + \phi^{'}Test\ score_{baseline} \\
& + \lambda_1 Cognitive\ parenting \times Stunting \\
& + \lambda_2 Caregiver\ education \times Stunting + \epsilon
\end{aligned}
$$

We hypothesized that λ_1 would be negative, since we expected the underlying causal effect of cognitive parenting on child development to be weaker for stunted children. We also conducted a series of robustness checks, most of which are discussed in Appendix 7.C.3. We tested an alternative explanation for a negative λ_1, i.e., that unobserved quality of cognitive parent interactions, as opposed to the frequency of parent-child interactions measured by our questionnaire, varies by stunting status. If unobserved quality of parent-child interactions is negatively correlated with stunting and positively correlated with frequency of interactions, this would cause the interaction term in the regression to be negative, even if the effect of our cognitive parenting measure is constant. As a proxy for parenting quality, we therefore chose caregiver education, which is also negatively correlated with stunting, and added a second interaction term to see if this affects the estimate of λ_1. Although we recognize that a causal interpretation goes beyond the scope of our observational study design, we see this as a useful exercise to rule out the confounding effect of unobserved quality differences.

7.3 Results

Research question 1: Are there gradients in child development based on family wealth for children ages 3 to 5 years? Table 7.1 ("No control" columns) shows large differences between children from households with low and high levels of wealth. The top wealth quintile performs between 0.46 and 0.70 SD higher than the lowest quintile in cognitive competences (language, early numeracy, and executive function). The gap for socioemotional problems (-0.23 SD between the top and bottom quintiles) is smaller than for cognitive competences. Table 7.1 (No control) shows that the wealth gaps widen with time. Between ages 3 and 4 years, and again between ages 4 and 5 years, the gap increases significantly for children's language and early numeracy scores. From a small 0.12 SD at age 3 in early numeracy (0.32 SD in language), the gap reaches a large magnitude of 0.88 SD in early numeracy (1.11 SD in language) by age 5. The gap also increases for executive function (from 0.40 to 0.55 SD) and socioemotional problems (from -0.21 to -0.27 SD), although these differences are not statistically significant. To visualize the differences in child development by socioeconomic status over age, we also used separate nonparametric regressions for the lowest and highest quintiles of the socioeconomic status index (Figure 7.1). The figure shows that the gap in language development at age 5 is of substantial magnitude and equivalent to a delay of 14-18 months (a child progressing 0.75 SD per year in our sample). The size of the raw wealth gap is substantially reduced after controlling for parenting (the three separate parenting composite scores) or anthropometrics (HAZs). Both sets of control variables reduce the gap in executive function by the same amount, about 8% (Table 7.1). The gap in language skills is reduced by about 10% when controlling for parenting and 6% when controlling for height-for-age. Similarly, the gap in early numeracy skills is reduced by 14% when controlling for parenting, and by 7% when controlling for nutrition. For socioemotional problems, parenting measures account for 43% of the gap and height-for-age does not account for any of the gap (0%).[14]

Research question 2: Does parenting show unique components? The exploratory factor analysis resulted in a three-factor solution, which is summarized in Table 7.2. The three retained parenting factors capture three dimensions of parenting. First, "cognitive parenting" measures how parents actively interact with their child in ways that are likely to develop cognitive competencies: playing games, reading books, and playing with toys or objects. The second dimension, "socioemotional parenting," captures the emotional support and responsiveness of the child and the parent. Third, "negative parenting," in which higher values indicate more negative parenting, refers to the degree to which parents use harsh or punitive statements to discipline their child. The constructs of emotional support and responsiveness and the use of harsh or punitive statements appear to be unique constructs in our sample.

Research question 3: How do the three parenting components relate to undernutrition, family wealth, and child development? The results, reported in Table 7.3, show that wealthier and more educated families tend to adopt parenting behaviors that are more cognitively stimulating and positive toward children. Cognitive parenting is positively correlated with caregiver education ($0.26,\ p < .001$) and caregiver nonverbal reasoning ($0.18,\ p < .001$) , but negatively correlated with a low household living standard ($-0.14,\ p < .001$) , the caregiver's age ($-0.15,\ p < .001$) . A similar pattern is observed for socioemotional parenting, although the correlations are somewhat smaller. There is no association between parenting measures and chil-

[14]As with all these results, the estimates should not necessarily be interpreted as causal impacts. In this case, in particular, for example, it is possible that socioemotional problems could be the cause rather than the result of negative parenting practices (i.e., reverse causation).

dren's stunted status.[15] Interestingly, although cognitive parenting and socioemotional parenting are positively correlated (0.38, $p < .001$) , negative parenting is negatively correlated with cognitive parenting (-0.05, $p = .002$) but positively correlated with socioemotional parenting (0.14, $p < .001$) . That is, punitive parenting behaviors are not inconsistent with other behaviors linked to positive socioemotional interactions. Cognitive parenting is significantly positively correlated with scores on child assessments, while negative parenting is significantly negatively correlated with these scores. No clear pattern emerges for socioemotional parenting. Socioemotional parenting is somewhat negatively correlated with socioemotional problems (-0.04, $p = .051$) while correlations with other composite scores are negative but significant. We use multivariate regressions in the next step to show that, also after controlling for household and caregiver characteristics and even baseline test scores, the association of parenting with child outcomes remains significant, and that socioemotional parenting is not negatively associated with executive function, language or early numeracy skills. To investigate the role of parenting and nutrition in wealth gradients in child development, we next use four specifications in a hierarchical regression model (see Table 7.4). In the first step (column 1), we regress the dependent variable on child age and child gender to determine how much of the total variation is explained by these exogenous characteristics. The share of explained variation in these regressions (R-squared) varies from 2.1% for socioemotional problems, to 38% for early numeracy, and approximately 43% for executive function and language skills. The three child-administered test scores increased by approximately 0.72 to 0.78 SD by year of age; the cognitive gap that separates a child in the bottom quintile from a child in the top quintile (between 0.46 and 0.70 SD, see Table 7.1) corresponds to between 65% and 100% of one year of average development. Girls perform significantly better than boys across all domains (9-12% of a SD). The gender difference is particularly pronounced for executive function. The caregiver and household control variables added in column 2 explain additional variation for the domains of executive function ($R^2\Delta = .021$) , language ($R^2\Delta = .039$) , early numeracy ($R^2\Delta = .023$) , and socioemotional problems ($R^2\Delta = .016$) . Joint significance of the added set of variables is confirmed with an F -test ($p < .001$ for all four outcomes). Low living standard, household size, caregiver education, and caregiver nonverbal reasoning affect children cognition in the expected directions, with all coefficients highly significant and of relatively large magnitude. To test if the three parenting variables explain additional variation in child outcomes, they were added to the regression model in Table 7.4, column 3. Parenting explains additional variance across all child out-

[15]We show in Table 7.12 that the association between parenting measures and stunting remains insignificant for all three parenting measures when we control for household and caregiver characteristics.

comes. The change in the R-squared is by far the highest for socioemotional problems ($\Delta R^2 = .122$) and of relatively smaller size for executive function ($\Delta R^2 = .003$) , language ($\Delta R^2 = .010$) , and early numeracy ($\Delta R^2 = .009$) . Joint significance of the three parenting variables is confirmed with an F -test ($p < .001$ for all four outcomes). For the cognitive domains, the parameters of all the household and caregiver controls are reduced once parenting is added to the model. The largest change is for caregiver education and household living standard for the socioemotional problems outcome. Of our three parenting measures, cognitive parenting shows the strongest association with the three cognitive scores. A one-SD increase in our measure of cognitive parenting is associated with a 0.092-0.098 SD increase in language and early numeracy scores, an increase that is comparable between one-half to two-thirds of the negative effect of stunting, low living standard, or low caregiver education. Socioemotional parenting is positively related to executive function skills but not to the other cognitive competences. Higher scores on socioemotional parenting are also significantly associated with fewer socioemotional problems. Although smaller in size than for cognitive parenting, the association of negative parenting with child outcomes is significantly negative for all the cognitive competences. The size of the relationship is especially large for the socioemotional problem score (0.346 SD). To test whether stunting explains additional variance in child development outcomes, we performed a similar exercise by adding it to the model in column 4 instead of the parenting variables. Stunting does not explain variance in socioemotional problems ($\Delta R^2 = .000$) , similar to the results in Table 7.1. Stunting explains a larger share of variance in executive function than our measures of parenting do ($\Delta R^2 = .008$) , a similar share for language ($\Delta R^2 = .008$), and a slightly smaller share for early numeracy ($\Delta R^2 = .006$). The coefficients of the household and caregiver variables change only marginally when stunting is added to the model. Thus, although the relationships with outcomes are significant—equivalent to about three months of cognitive development, for example—stunting seems to be a weaker mediator of caregiver education than parenting behaviors are.[16]

We explore additional associations and mediating factors in Table 7.5. First, we include the baseline test scores collected 11 months prior to the main outcomes of this study (which would also have the effect of controlling for unobserved factors already captured in those scores). Although the magnitudes fall, the relationships with parent-

[16]That the wealth gradients remains large even after controlling for anthropometrics and parenting should be interpreted with care. As our measure of parenting is only self-reported and based on recall, it is possible that it does not capture the full dynamic of parenting throughout the children's life time. Our measure of parenting could for instance be imperfectly correlated with past parenting (during pregnancy or breastfeeding). Our measure of parenting may also imperfectly measure the quality of parenting. Longitudinal and external measures of parenting would be necessary to enhance our measures.

ing typically remain significant, suggesting that the effect of parenting is cumulative (and not "one-shot").[17]

Research question 4: How do parenting and nutrition interact in explaining variation in child development? We also hypothesized that parenting may have differential effects on child development, depending on children's nutritional status. To test the hypothesis that the association of cognitive parenting and child outcomes is weaker for stunted children, we analyze the role of parenting, separately for stunted and non-stunted children and with the baseline controls (Table 7.5, column 3). We focus on cognitive parenting because of the notable role of nutritional status in cognitive development; we thus hypothesize that cognitive parenting and subsequent child development would be especially sensitive to nutritional status. Although no significant interactions between parenting and child nutritional status are observed for executive function or socioemotional problems, the association of cognitive parenting and language and early numeracy outcomes differ by stunting. It is about 35% smaller for language and 54% smaller for early numeracy, showing that the positive association of cognitive parenting and cognitive development is weaker for stunted children. Although the cognitive parenting-stunting interaction is statistically significant for the language and early numeracy outcomes, this finding could also be explained by the unobserved quality of cognitive parenting interactions. We add a cognitive parenting-caregiver education interaction to the models in the last columns of Table 7.5, to see if the coefficient of the cognitive parenting–stunting interaction changes. The inclusion of caregiver education and its interaction with nutritional status does not change the significance of the previous interaction. We thus conclude that cognitive parenting shows stronger associations with some domains of child development for non-stunted versus stunted children, and that this cannot be explained be differences in the quality of interactions. We conducted a series of sensitivity analyses, shown in Appendix 7.C.3. The conclusions from Table 7.5 do not change when a model without baseline test scores is used. Models controlling for village fixed effects and preschool enrollment also yield similar findings. Since three different cognitive measures were used throughout this study and we did not account for multiple hypothesis testing, we test whether the same associations can be observed when a summary measure of cognitive skills is used. As expected, the associations with parenting measures are still significant. The interaction between parenting and child nutritional status is muted, compared with the regressions on language and early numeracy, due to the influence of the executive function measures, and it is no longer significant. Finally, we test stunting cutoffs dif-

[17]In Table 7.5, we control for parenting and stunting. The effects of each are relatively unchanged.

ferent from HAZ < -2 SD, to show that the relatively weaker association of cognitive parenting and test scores is predominantly driven by children in the lower tail of the HAZ distribution.

7.4 Discussion

Consistent with previous findings, our results confirm that wealth gaps exist in rural Cambodia, where the overall level of wealth is not high, and parenting and undernutrition jointly contribute to the differences between higher- and lower-income children's development. However, our results also suggest that children who have experienced undernutrition are less sensitive to parenting and, moreover, that parenting interventions may have a large positive influence on child development but are unlikely to compensate fully for the negative impacts of early undernutrition. Consistent with previous work, the wealth gradients documented in this sample amount to about 7-9 months of child development for preschool-age children. The gap appears to be larger for cognitive competences (executive function, language, and early numeracy) than for behavioral problems. The gap in child development due to family wealth widens with age (consistent with, e.g., Paxson and Schady, 2007; Fernald et al., 2012). The gap at ages 4 and 5 is much larger than at age 3, particularly for language development (from 0.32 SD between high- and low-income families at age 3 to 1.11 SD at age 5) and early numeracy, but less for executive function (from 0.40 to 0.55 SD). The gap is also relatively constant across children's ages for socioemotional problems. That the executive function gap is already present at age 3 and only .46 SD larger at age 5 suggests that before age 3, the impacts of family wealth on executive function are already evident. Consistent with previous work, we confirm that cognitively stimulating and emotionally responsive parenting are important explanatory factors of wealth gradients in child development. Our measures of parenting behaviors are strongly correlated with children's performance, accounting for about 8% -14% of the wealth gap in cognitive competences observed among children ages 3 to 5-years. Although the explanatory power of parenting is robust, the share of explained variation in our test measures is relatively small in comparison with other explanatory factors: we show that a large 1 SD improvement in the main parental competences—an improvement beyond the scope of most parenting programs—would only compensate about half of the effect of stunting. Finally, our study also provides evidence that the positive effects of stimulating and supportive parenting may be stronger for children who are not stunted. This is an important finding for two reasons. First, it demonstrates the importance of designing and testing integrated program models of child development that anticipate the complex interactions that will arise between child characteristics and environmental stimuli.

It is unlikely that "one size fits all" (Shonkoff et al., 2017), but rather that the effects of early childhood interventions may vary based on child and family characteristics. Second, this interaction—albeit a small one, and which is not consistent across all domains of parenting or child development—adds to the growing body of research documenting the variable effects of environments on child development, by adding undernutrition as a condition that may moderate the effects of later environmental stimuli on child development. Finally, children's socioemotional development is less strongly associated with stunting, and it is also less strongly associated with wealth gradients. Although it is important to note that this study only included a measure of behavioral problems, which does not comprise the totality of socioemotional development, our findings suggest that supporting socioemotional development may be an important pathway for building resilience among children living in adversity. There are several limitations to this study that should be noted. First, parents reported on their parenting behaviors, and a more objective measure of parenting behavior may have yielded different results. A parents' perception of his or her parenting behaviors could also be influenced by the child's characteristics, given the bidirectional nature of parent/child dynamics, and our study was not able to capture that dynamic. Second, in the regression models, the additional variance of child development accounted for by parenting and the interaction of parenting and nutrition was small. Although the results were significant, our findings should be replicated in other studies to confirm that these effects are reliable and meaningful. Finally, a longer-term longitudinal study design is needed to describe the interactions between environmental stimulation as defined by parenting, nutrition, and child development over time. For example, a longitudinal design would produce more insight into how child nutritional status and other factors may elicit parenting responses over time, leading to a cascade of effects on child development. (Prado and Dewey, 2014) call for more comprehensive analyses of how environmental stimulation (including but not limited to parenting) could interact with child nutritional status, using additive and interactive models. This study should be replicated and expanded upon with other types of environmental stimulation, such as schooling. The results reported here suggest that children may also show differential responsiveness to preschool and other school settings based on early nutritional status. Findings in this area have been reported for Latin America (e.g. Cueto et al., 2016) by demonstrating that more nourished children benefit more from formal preschools than less nourished children do. If it is the case that children who have experienced early stunting respond less to environmental stimulation, the implications for addressing inequity are profound. Parenting interventions, informal and formal schooling environments, and other forms of environmental stimulation may contribute to the gap between stunted children and their peers without more careful attention to the types of environmental stimulation that are effective for stunted children. Education systems and programs should continue to acknowledge the importance of attending to early nutrition, consistent with the messages from many advocates on the importance of building integrated early childhood systems (e.g. Zonji, 2018). This study adds to an important body of literature documenting the emergence and growth of wealth gradients in child development, even within populations that are not wealthy overall. We also establish that parenting behaviors, as self-reported by parents, meaningfully relate to child development outcomes in the predicted ways, suggesting that although cultural influences are of course critical to parenting, some degree of cross-cultural similarity in parenting young children may also be evident. The documentation of the interaction between parenting and undernutrition highlights the complex interactions between environmental stimulation and nutrition. More work is needed to examine the interactions between the environmental conditions that contribute to wealth gradients. Such analyses would help guide the design of effective interventions, which would exploit complementarities between parenting, nutrition, and other factors for the promotion of child development in challenging contexts. Encouraging stimulating and responsive parenting and improving children's educational environments are key to reducing cognitive inequality and improving children's performance during that critical time when the cognitive gap widens. As with all elements of early development, prevention and early intervention are more effective than amelioration—and, in this case, that may pertain to addressing undernutrition before conception and during pregnancy. At a minimum, the findings from this study confirm that a combination of interventions may be needed to provide effective support for young children's development in low-income countries.

7.A Tables

Table 7.1: Results of linear regressions predicting different domains of child development from quintile dummies, quintile dummy–child age interactions, and parenting variables or child anthropometrics

	Executive function			Language			Early numeracy			Socio-emotional problems		
	No control	**Parenting**	**Height for age**	**No control**	**Parenting**	**Height for age**	**No control**	**Parenting**	**Height for age**	**No control**	**Parenting**	**Height for age**
Quintile gap: Full sample	0.487***	0.445***	0.446***	0.701***	0.632***	0.661***	0.460***	0.395***	0.426***	-0.234***	-0.134***	-0.238***
... age 3	0.397***	0.359***	0.352***	0.318***	0.256***	0.273***	0.117**	0.060	0.081*	-0.205***	-0.118*	-0.209***
... age 4	0.523***	0.482***	0.488***	0.725***	0.654***	0.690***	0.439***	0.374***	0.408***	-0.234***	-0.154**	-0.242***
... age 5	0.545***	0.496***	0.498***	1.111***	1.036***	1.068***	0.881***	0.811***	0.843***	-0.267***	-0.129*	-0.266***
... age 4 - age 3	0.126	0.123	0.136	0.408***	0.397***	0.418***	0.322***	0.314***	0.327***	-0.029	-0.036	-0.033
... age 5 - age 3	0.148	0.137	0.146	0.793***	0.780***	0.796***	0.764***	0.750***	0.762***	-0.062	-0.011	-0.056
... age 5 - age 4	0.022	0.014	0.010	0.386***	0.383***	0.378***	0.442***	0.437***	0.435***	-0.033	0.024	-0.024

Notes: The table shows gaps in child development between the highest and lowest wealth quintiles. The “No control” columns show results using only quintile dummies as explanatory variables. The “Parenting” columns add the three parenting composite scores to the explanatory variables. The “Height for age” columns add height-for-age z-scores to the control variables. The dependent variables are standardized residuals of a linear regression of the original test scores on the first three polynomials of age. See the main text for more details. Two-tailed p-values for t-tests using robust standard errors clustered at the village level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7.2: Pattern matrix for exploratory factor analysis (iterated principal factor) with promax oblique rotation of the parenting measures

Items (abbreviated)	Mean	SD	Cognitive parenting	Socio-emotional parenting	Negative parenting
Read book or magazine and look at pictures	1.78	1.25	**0.57**	-0.02	-0.03
Play games with words	1.70	1.23	**0.50**	0.00	0.01
Tell stories	1.57	1.09	**0.50**	-0.01	0.01
Sing songs	2.16	1.51	**0.46**	0.12	-0.01
Play games with numbers	2.64	1.57	**0.51**	0.10	-0.03
Play active games, e.g. with running or ball	1.42	1.00	**0.42**	0.01	-0.02
Draw or paint	1.49	1.01	**0.56**	-0.07	-0.01
Spend time with child so he/she can feel love and care	4.18	1.20	-0.04	**0.47**	-0.02
Show love by hugging, kissing, caring	4.44	1.07	0.02	**0.48**	-0.08
Teach to become self-suficient in daily routines	4.50	0.98	-0.01	**0.46**	-0.01
Talk to child when he/she did something wrong	4.02	1.16	-0.01	**0.45**	0.17
Comfort child when he/she feels sad	3.89	1.22	-0.03	**0.49**	0.07
Encourage, compliment when child does something	3.77	1.20	0.11	**0.51**	-0.08
Tell child that he/she makes you happy	2.66	1.50	0.21	**0.30**	-0.02
Talk about reasons when forbidding something	4.18	1.11	-0.01	**0.40**	0.14
Feel that it is hard taking care of child	3.06	1.65	0.03	0.00	**0.44**
Physical punishment (kicking, slapping, beating)	2.04	1.20	0.04	-0.04	**0.56**
When angry, throw anger at child by shouting	2.84	1.38	-0.05	0.04	**0.63**
Feel annoyed with what child has done	2.68	1.37	0.01	-0.03	**0.72**
Feel annoyed or angry when child cries	2.59	1.40	-0.01	0.05	**0.59**
Shout or speak loudly to child	3.00	1.41	-0.07	0.04	**0.65**
Call child dumb, lazy or similar	1.59	1.11	0.05	-0.12	**0.37**
Talk, e..g name objects, describe daily routines	3.29	1.53	0.23	0.24	0.03
Ignore when child cries	1.47	0.97	0.08	-0.12	0.24
Go for a walk with child	3.36	1.51	0.01	0.29	0.06
Correlations between factor scales:					
Cognitive parenting				0.61	-0.06
Socio-emotional parenting					0.21
Cronbach's alpha of composite score			0.72	0.69	0.77

Notes: Factor loadings > .3 in bold are the items of the three parenting composite scores. Possible responses ranged from 1 (never or almost never) to 5 (every day or almost every day).

Table 7.3: Pairwise correlations between follow-up child test scores, child, household, and caregiver control variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Executive function	1.00													
2 Language development	0.75***	1.00												
3 Early numeracy	0.66***	0.77***	1.00											
4 Socio-emotional problems	-0.15***	-0.16***	-0.16***	1.00										
5 Age (years)	0.65***	0.66***	0.61***	-0.14***	1.00									
6 Female	0.06***	0.04*	0.04	-0.05***	-0.01	1.00								
7 Stunted	-0.12***	-0.12***	-0.11***	0.01	-0.02	-0.03	1.00							
8 Household size	-0.01	-0.05***	-0.02	-0.02	0.01	-0.01	0.03	1.00						
9 Low living standard	-0.08***	-0.13***	-0.10***	0.08***	0.00	0.01	0.05***	-0.11***	1.00					
10 Caregiver age	0.01	0.03	0.03	0.04	0.01	-0.01	0.01	0.05***	-0.13***	1.00				
11 Caregiver primary school	0.07***	0.09***	0.07***	-0.11***	-0.03	0.00	-0.06***	-0.05***	-0.13***	-0.32***	1.00			
12 Caregiver non-verbal reasoning	0.08***	0.10***	0.07***	-0.06***	-0.02	-0.00	-0.06***	-0.04*	-0.07***	-0.37***	0.38***	1.00		
13 Cognitive parenting	0.05***	0.11***	0.10***	-0.12***	-0.03	0.01	-0.03	-0.07***	-0.14***	-0.15***	0.26***	0.18***	1.00	
14 Socio-emotional parenting	-0.03	-0.02	-0.01	-0.04*	-0.11***	-0.01	-0.03	-0.06***	-0.09***	-0.03	0.11***	0.08***	0.38***	1.00
15 Negative parenting	-0.07***	-0.07***	-0.08***	0.35***	-0.04**	-0.10***	-0.00	0.02	0.07***	-0.04	-0.05***	-0.01	-0.05***	0.14***

Notes: The table shows pairwise correlations (Sidac-corrected for multiple testing). Two-tailed p-values: $^{*}p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Table 7.4: Results of hierarchical regression analysis predicting different domains of child development from parental stimulation, nutrition, and other household, caregiver, and child variables

	Executive function				Language				Early numeracy				Socio-emotional problems			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Age (years)	0.770***	0.772***	0.774***	0.768***	0.777***	0.780***	0.782***	0.776***	0.722***	0.724***	0.726***	0.721***	-0.158***	-0.161***	-0.153***	-0.161***
Female	0.124***	0.126***	0.121***	0.121***	0.092***	0.095***	0.089***	0.090***	0.088***	0.091***	0.083***	0.087***	-0.106***	-0.108***	-0.042*	-0.108***
Household size		-0.013**	-0.011*	-0.012**		-0.038***	-0.034***	-0.036***		-0.018***	-0.014**	-0.017***		-0.012*	-0.021***	-0.012*
Low living standard		-0.170***	-0.161***	-0.157***		-0.238***	-0.220***	-0.225***		-0.187***	-0.169***	-0.176***		0.092***	0.053**	0.092***
Caregiver age		0.003***	0.003***	0.003***		0.005***	0.005***	0.005***		0.004***	0.004***	0.004***		0.001	0.002**	0.001
Caregiver primary school		0.131***	0.110***	0.125***		0.172***	0.126***	0.166***		0.148***	0.103***	0.143***		-0.220***	-0.139***	-0.220***
Caregiver non-verbal reasoning		0.076***	0.072***	0.072***		0.089***	0.082***	0.085***		0.063***	0.056***	0.060***		-0.017	-0.007	-0.017
Cognitive parenting			0.032***				0.098***				0.092***				-0.051***	
Socio-emotional parenting			0.023**				0.003				0.007				-0.077***	
Negative parenting			-0.026**				-0.022**				-0.033***				0.346***	
Stunted				-0.187***				-0.183***				-0.154***				0.001
Constant	-3.447***	-3.473***	-3.490***	-3.392***	-3.464***	-3.446***	-3.481***	-3.368***	-3.222***	-3.268***	-3.299***	-3.202***	0.750***	0.802***	0.723***	0.802***
Observations	6699	6699	6699	6699	6699	6699	6699	6699	6699	6699	6699	6699	6704	6704	6704	6704
R^2	0.432	0.453	0.455	0.461	0.437	0.477	0.486	0.485	0.378	0.401	0.410	0.407	0.021	0.037	0.159	0.037
F	2846.5**	891.5***	629.3***	816.9***	1755.3**	557.8***	400.3***	502.6***	1018.2**	322.0***	233.4***	288.5***	72.3***	33.2***	127.7***	29.1***
$R^2\Delta$		0.021	0.003	0.008		0.039	0.010	0.008		0.023	0.009	0.006		0.016	0.122	0.000
$F\Delta$		41.10***	9.57***	86.05***		68.47***	34.81***	101.02***		35.79***	28.35***	73.21***		16.97***	306.33**	0.011

Notes: Changes in the R^2 ($R^2\Delta$) and F-tests for joint significance of additional variables ($F\Delta$) refer to a comparison of (2) vs. (1) in column (2), (3) vs. (2) in column (3), and (4) vs. (2) in column (4). Two-tailed p-values for t-tests and F-tests using robust standard errors clustered at the village level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

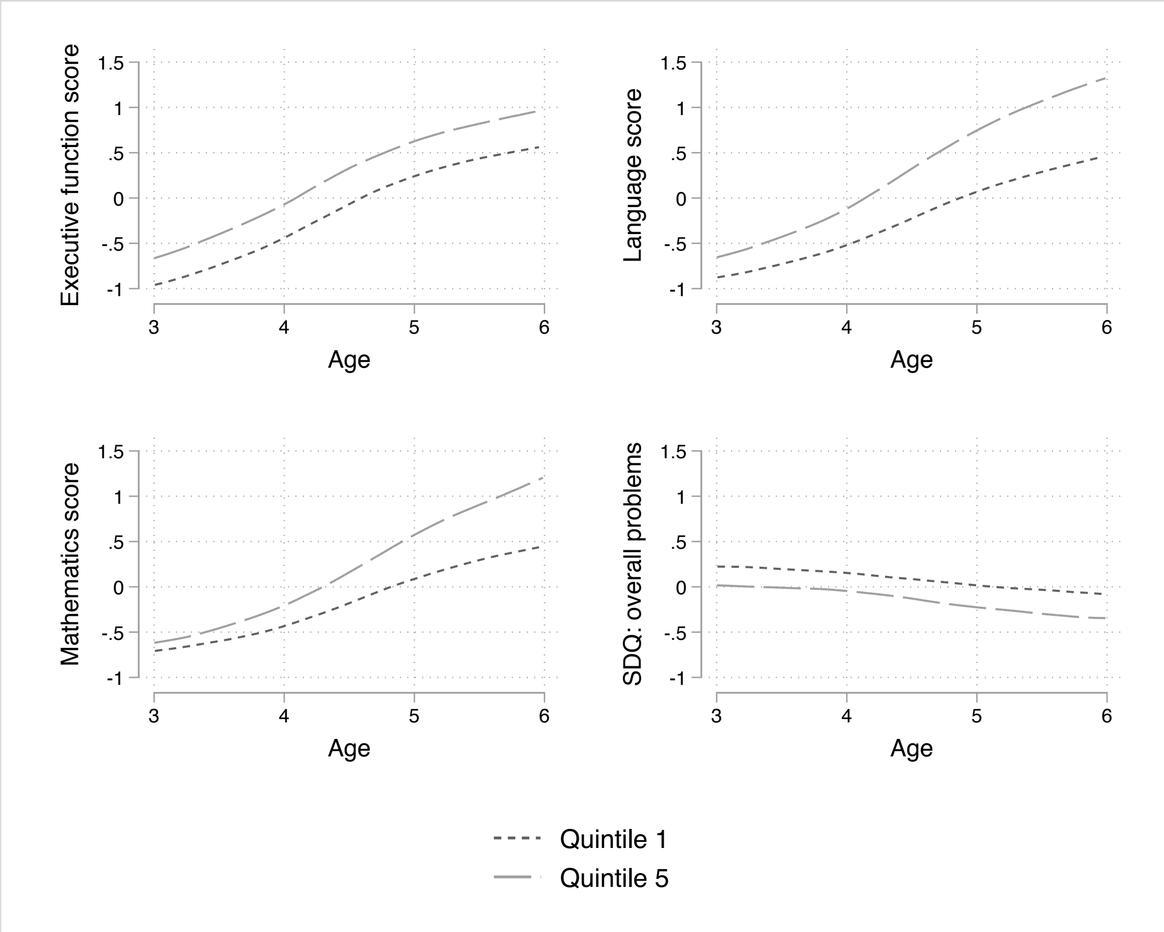
Table 7.5: Results of hierarchical regression analysis predicting different domains of child development using baseline controls, a cognitive parenting x stunted interaction term, and a cognitive parenting x caregiver education interaction term

	Executive function				Language				Early numeracy				Socio-emotional problems			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Age (years)	0.770***	0.502***	0.502***	0.502***	0.778***	0.424***	0.424***	0.424***	0.723***	0.379***	0.379***	0.379***	-0.153***	-0.101***	-0.101***	-0.101***
Female	0.116***	0.095***	0.095***	0.095***	0.084***	0.059***	0.059***	0.059***	0.079***	0.053***	0.053***	0.053***	-0.041*	-0.032	-0.032	-0.032
Household size	-0.010*	-0.003	-0.003	-0.003	-0.033***	-0.023***	-0.023***	-0.023***	-0.013**	-0.004	-0.004	-0.004	-0.021***	-0.017**	-0.017**	-0.017**
Low living standard	-0.148***	-0.084***	-0.084***	-0.084***	-0.208***	-0.121***	-0.121***	-0.121***	-0.159***	-0.080***	-0.080***	-0.080***	0.052**	0.045*	0.045*	0.045*
Caregiver age	0.003***	0.001	0.001	0.001	0.005***	0.003***	0.003***	0.003***	0.004***	0.002***	0.002***	0.002***	0.002**	0.002*	0.002*	0.002*
Caregiver primary school	0.103***	0.076***	0.076***	0.071***	0.120***	0.075***	0.075***	0.075***	0.097***	0.054**	0.054**	0.052**	-0.139***	-0.119***	-0.119***	-0.115***
Caregiver non-verbal reasoning	0.069***	0.051***	0.051***	0.051***	0.078***	0.056***	0.056***	0.056***	0.053***	0.031***	0.030***	0.030***	-0.006	-0.000	-0.000	-0.000
Cognitive parenting	0.032***	0.011	0.012	0.004	0.098***	0.065***	0.075***	0.075***	0.091***	0.057***	0.071***	0.069***	-0.051***	-0.041***	-0.046***	-0.041**
Socio-emotional parenting	0.021*	0.017	0.017	0.017*	0.001	-0.006	-0.005	-0.005	0.006	0.000	0.001	0.001	-0.077***	-0.067***	-0.067***	-0.067***
Negative parenting	-0.026**	-0.019**	-0.019**	-0.019**	-0.023**	-0.011	-0.011	-0.011	-0.033***	-0.022**	-0.022**	-0.022**	0.346***	0.311***	0.312***	0.311***
Stunted	-0.186***	-0.130***	-0.131***	-0.130***	-0.182***	-0.108***	-0.108***	-0.108***	-0.153***	-0.080***	-0.080***	-0.080***	0.004	-0.004	-0.004	-0.004
Child executive function (baseline)		0.164***	0.165***	0.165***		0.061***	0.061***	0.061***		0.029*	0.029*	0.029*		0.018	0.018	0.018
Child language (baseline)		0.114***	0.114***	0.113***		0.327***	0.327***	0.327***		0.246***	0.245***	0.245***		-0.042**	-0.042**	-0.042**
Child early numeracy (baseline)		0.115***	0.115***	0.115***		0.115***	0.115***	0.115***		0.235***	0.235***	0.235***		-0.012	-0.012	-0.012
Socio-emotional problems (baseline)		0.004	0.004	0.004		0.004	0.004	0.004		0.003	0.003	0.003		0.197***	0.197***	0.197***
Cognitive parenting **x** Stunted			-0.003	-0.001			-0.026*	-0.026*			-0.038**	-0.037**			0.014	0.013
Cognitive parenting **x** Caregiver primary school				0.021				-0.000				0.005				-0.014
Constant	-3.409***	-2.217***	-2.217***	-2.218***	-3.402***	-1.837***	-1.837***	-1.837***	-3.232***	-1.707***	-1.707***	-1.707***	0.721***	0.483***	0.483***	0.484***
Observations	6699	6699	6699	6699	6699	6699	6699	6699	6699	6699	6699	6699	6704	6704	6704	6704
R^2	0.464	0.531	0.531	0.531	0.494	0.611	0.612	0.612	0.416	0.538	0.539	0.539	0.159	0.197	0.197	0.197
F	599.5***	515.2***	483.2***	455.1***	372.1***	432.1***	407.1***	391.0***	216.5***	265.7***	251.0***	236.4***	116.6***	110.2***	103.4***	97.3***
$R^2\Delta$		0.067	0.000	0.000		0.117	0.000	0.000		0.122	0.000	0.000		0.038	0.000	0.000
$F\Delta$		163.8***	0.027	1.032		296.2***	3.322*	0.001		272.7***	4.941**	0.056		89.10***	0,315	0.317

Notes: Changes in the R^2 ($R^2\Delta$) and F-tests for joint significance of additional variables ($F\Delta$) refer to a comparison of (2) vs. (1) in column (2), (3) vs. (2) in column (3), and (4) vs. (3) in column (4). Two-tailed p-values for t-tests and F-tests using robust standard errors clustered at the village level. $*p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$

7.B Figures

Figure 7.1: Nonparametric regressions of composite scores on age using a bandwidth of 6 months



7.C Appendix

7.C.1 Parenting questionnaire

This section includes the full list of questions of the follow-up parenting questionnaire. An abbreviated version of the questions with descriptive statistics is shown in Table 7.2 . Questions in the parenting questionnaire used at follow-up:

1. How often do you feel that it is hard taking care of (. . .)?
2. How often do you read a book/magazine or look at the pictures in a book or magazine with (. . .)?
3. How often do you physically punish (. . .) (kicking, slapping, beating)?
4. How often do you play games with words with (. . .)?
5. When you feel angry, how often do you throw your anger to (. . .) by shouting at him/her?
6. How often do you tell stories to (. . .)?
7. How often do you spend time with (. . .) so that he/she can feel love and care from you?
8. How often do you feel annoyed with what (. . .) has done?
9. How often do you sing songs with/to (. . .)?
10. How often do you talk with (. . .), for example naming objects, describing daily routines, etc.?
11. How often do you talk to (. . .) about reasons, when you forbid him/her to do something?
12. How often does (. . .) tell you about what he/she has done?
13. How often do you play games with numbers with (. . .)?
14. How often do you feel annoyed or angry when (. . .) cries?
15. How often do you show your love to (. . .) such as hugging, kissing or caring him/her?
16. How often do you teach (. . .) how to become self-sufficient in daily routines, for example getting dressed, toileting, etc.?
17. How often do you shout or speak loudly to (...)?

18. How often do you play active games with (...), for example playing with a ball, running, etc.?
19. How often do you tell (...) that he/she makes you happy?
20. How often do you go for a walk with (...)?
21. How often do you talk with (...), when he/she does something that is not good/suitable for him/her?
22. How often do you draw or paint with (...)?
23. How often do you call (...) dumb, lazy, or another name like that?
24. How often do you comfort (...) when he/she feels sad?
25. How often do you encourage/compliment (...) when he/she does something?
26. How often do you ignore (...) when he/she cries?

Answer scale:

1. Never or almost never
2. A few times in the month
3. About once a week
4. Several times a week
5. Every day or almost every day

7.C.2 Child test adaptations, correlations, and distributions

To ensure that the child assessments were suitable for the studied context, baseline and follow-up data collections were preceded by pretests in the field. Three child assessment pretests were conducted at baseline ($n = 25$, $n = 31$, $n = 156$) and two at follow-up ($n = 22$, $n = 62$). The first pretests of each wave were conducted by field supervisors, and the last pretest of each wave was conducted by enumerators during the end of their training. The third pretest at baseline lasted two days to ensure that all the enumerators gained enough field experience with the child assessments before the start of data collection. For each wave, a selected set of tests for each of the domains was first tested with a small sample of children. This allowed the study to assess whether the tests were generally well understood by the field staff and children and whether they are too easy or too difficult for the children at their current age. Some of the tests from MELQO were removed during this adaptation process. For example, we removed a backward-digit span test and a mental transformation task, because very few of the children were able to understand them. Likewise, as the children in our sample were unable to write Khmer letters, we removed tests that included writing. Sound (phoneme) discrimination tests were also too difficult for our sample of children. Based on the pretest findings, other tests included in our battery were adapted to the local context or modified to fit the competences of the sampled children. For example, as some of the children rarely visit markets or stores, a question asking for things that can be bought and eaten at a market was changed to things that can be eaten. Likewise, the letters and numbers used in some of the tests were selected to fit the local preschool curriculum. A number identification test was replaced with a self-developed quantity comparison test where children had to compare quantities of printed objects instead of printed numbers. A simple addition test was merged with the same quantity comparison test, by asking the children questions about the total number of objects obtained after adding or subtracting a given number of objects. Since the overall purpose was to generate a well-functioning test for Cambodia (as opposed to maintaining consistency for other purposes, such as international comparisons), the adaptations were prioritized to ensure adequate fit to the local context. In Tables 7.6 and 7.7, we show pairwise correlations of raw test scores and covariates. Figure 7.2 shows empirical distributions of raw test scores and composite test scores. Additional data and details on the executive functioning assessments and receptive vocabulary test:

- Head-knee task (1.1). Before each stage of this test, the enumerator explains what to do and uses two practice trials to ensure that the child knows what to do. Only children who understood the concept of the first stage ($M = 0.866$) participated in it and transitioned to the second stage. Of the children who participated in the first stage, 85.1% scored 3 or more out of 5 correct, i.e., better than random. Only after they understood the concept of the second stage ($M = 0.210$), the complete test was administered. Children were able to score 2 points per movement in the second stage when they were able to perform it without self-correction and 1 point when they self-corrected their spontaneous response. Among the children participating in the second stage, an average score of 4.90 out of 10 was reached, with 65.46% of the children being able to score at least once without self-correction and 59.93% of the children scoring in at least 3 out of 5 movements.

- Forward digit span (1.2). Two practice trials with two digits were used to determine whether the child understood the concept of the test, which was the case for 84.8% of the children.

- Dimensional Change Card Sort (1.3). On average, the children scored 4.71 out of 6 cards correct in the pre-switch phase and 25.02% of the children scored at chance (3 out of 6) or worse. The post-switch phase was reached by 36.76% of the children age 3, 69-68% of the children age 4, and 87.59% of the children age 5. The children who reached the post-switch phase scored 4.9 points on average (children ages 3, 4, and 5 scored 4.3, 4.9, and 5.2 on average, respectively). In the post-switch phase, 18.61% of the children scored at chance (3 out of 6) or worse.
- Cancellation task (1.4). Because the children had to cross out symbols by drawing at least a line or a scribble, this test was only administered when a child was able to score at least 1 point in the copying test (age 3: $M = 0.418$, age 4: $M = 0.805$, age 5: $M = 0.943$). The test was otherwise scored zero. Children who were not able to participate in the cancellation task were younger, slightly less likely to be female ($r = -0.024$,$p = 0.049$), less likely to be raised by a caregiver with primary school education ($r = -0.060$,$p < .001$), and more likely to be stunted ($r = 0.077$,$p < .001$) or from a household with a low living standard ($r = 0.093$,$p < .001$). Of the children participating in the test, 80.35% were able to reach the second matrix, and 60.03% reached the fourth matrix.
- Receptive vocabulary test (2.1). Up to six practice trials were conducted to test whether the child was able to understand the concept and willing to participate: 99.0% of the children passed the practice trials; the others were assigned a score of zero. Of those passing, 79.76% reached the 10th picture, 51.58% reached the 20th picture, 32.96% reached the 30th picture, and 16.3% reached the 40th picture. All the other tests did not use practice trials. Figure 7.2 provides details on the share of the children obtaining a score of zero.

Table 7.6: Pairwise correlations between raw test scores

	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	2.5	2.6	3.1	3.2	3.3	3.4	3.5	4.1	4.2	4.3	4.4
1.1: Head-knee task	1.00																		
1.2: Forward digit span	0.45***	1.00																	
1.3: DCCS	0.35***	0.37***	1.00																
1.4: Cancellation task	0.42***	0.43***	0.49***	1.00															
2.1: Receptive vocabulary	0.43***	0.41***	0.44***	0.60***	1.00														
2.2: Expressive language	0.42***	0.50***	0.38***	0.49***	0.48***	1.00													
2.3: Receptive language (short sto	0.45***	0.46***	0.43***	0.53***	0.57***	0.54***	1.00												
2.4: Knowledge of reading concep	0.26***	0.27***	0.27***	0.32***	0.33***	0.33***	0.31***	1.00											
2.5: Letter recognition test	0.20***	0.14***	0.13***	0.20***	0.27***	0.20***	0.20***	0.17***	1.00										
2.6: Shape drawing test	0.46***	0.49***	0.51***	0.69***	0.57***	0.50***	0.55***	0.37***	0.21***	1.00									
3.1: Measurement vocabulary	0.30***	0.36***	0.37***	0.40***	0.42***	0.37***	0.39***	0.25***	0.12***	0.43***	1.00								
3.2: Verbal counting	0.41***	0.37***	0.34***	0.47***	0.52***	0.48***	0.48***	0.32***	0.38***	0.50***	0.32***	1.00							
3.3: Quantitative comparison	0.44***	0.47***	0.44***	0.52***	0.56***	0.51***	0.53***	0.33***	0.28***	0.58***	0.44***	0.63***	1.00						
3.4: Number identification	0.33***	0.25***	0.24***	0.35***	0.44***	0.33***	0.36***	0.25***	0.63***	0.37***	0.21***	0.61***	0.47***	1.00					
3.5: Shape identification	0.18***	0.20***	0.20***	0.22***	0.23***	0.21***	0.23***	0.14***	0.09***	0.25***	0.19***	0.20***	0.24***	0.15***	1.00				
4.1: SDQ: emotional symptoms	-0.01	-0.05***	-0.06***	-0.05**	-0.06***	-0.03	-0.03	-0.04	-0.01	-0.04	-0.06***	-0.04	-0.05***	-0.02	-0.02	1.00			
4.2: SDQ: conduct problems	-0.03	-0.04	-0.07***	-0.07***	-0.07***	-0.05***	-0.07***	-0.04	-0.03	-0.07***	-0.05***	-0.08***	-0.08***	-0.06***	-0.03	0.21***	1.00		
4.3: SDQ: hyperactivity/inattentio	-0.08***	-0.09***	-0.10***	-0.10***	-0.12***	-0.07***	-0.11***	-0.07***	-0.07***	-0.15***	-0.08***	-0.12***	-0.12***	-0.11***	-0.07***	0.14***	0.32***	1.00	
4.4: SDQ: peer relationship proble	-0.07***	-0.08***	-0.10***	-0.12***	-0.10***	-0.14***	-0.09***	-0.07***	-0.05***	-0.09***	-0.07***	-0.12***	-0.11***	-0.09***	-0.02	0.06***	0.19***	-0.01	1.00

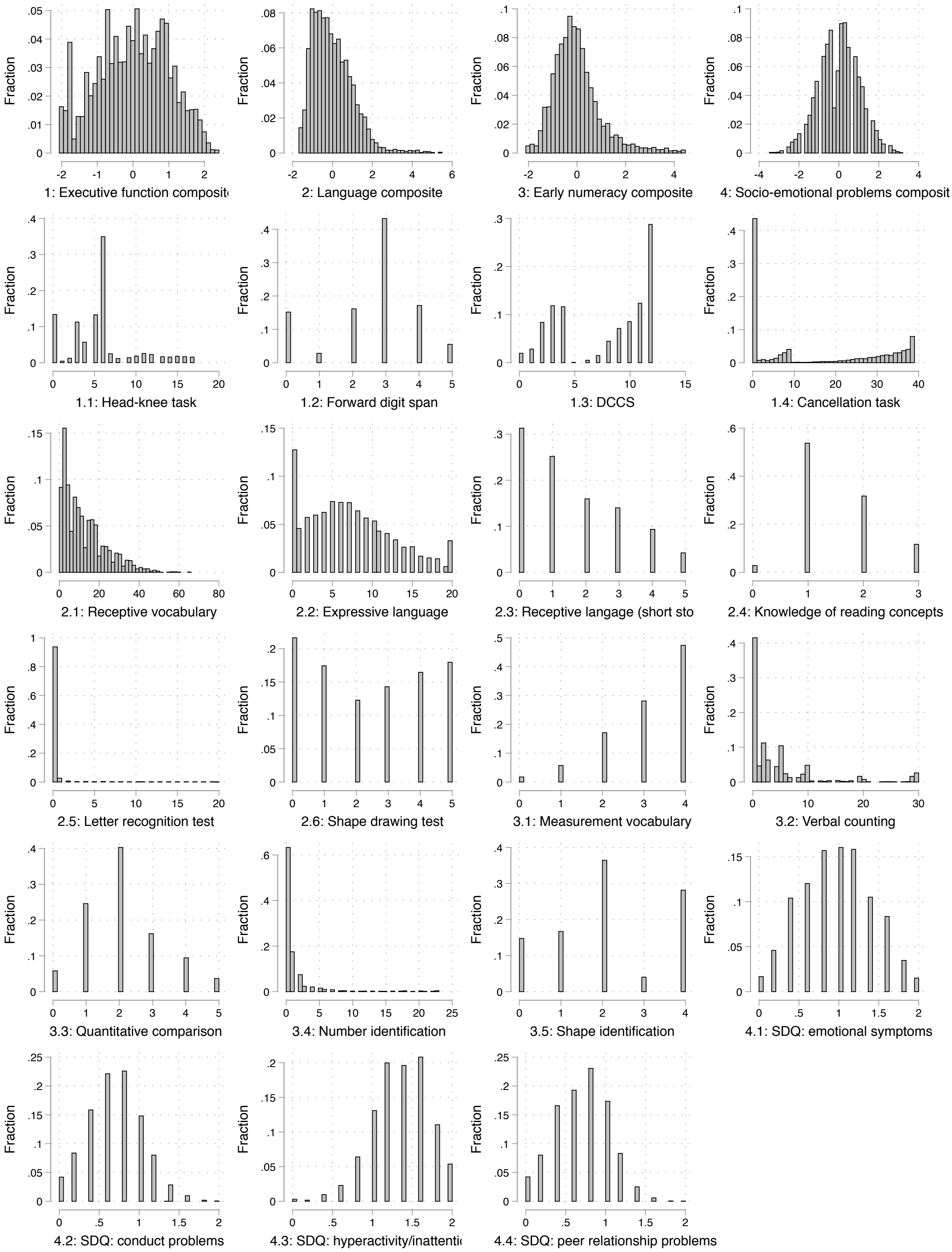
Notes: The table shows pairwise correlations (Sidac-corrected for multiple testing). Two-tailed p-values: *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7.7: Pairwise correlations between covariates and raw test scores

	1	2	3	4	5	6	7	8	9	10	11
1 Age (years)	1.00										
2 Female	-0.01	1.00									
3 Household size	0.01	-0.00	1.00								
4 Low living standard	0.00	0.01	-0.11***	1.00							
5 Caregiver age	0.01	-0.01	0.05*	-0.13***	1.00						
6 Caregiver primary school	-0.03	0.00	-0.05***	-0.13***	-0.32***	1.00					
7 Caregiver non-verbal reasoning	-0.02	0.00	-0.04	-0.07***	-0.37***	0.38***	1.00				
8 Cognitive parenting	-0.03	0.01	-0.06***	-0.14***	-0.15***	0.26***	0.18***	1.00			
9 Socio-emotional parenting	-0.11***	-0.01	-0.06***	-0.09***	-0.02	0.11***	0.08***	0.38***	1.00		
10 Negative parenting	-0.05*	-0.10***	0.02	0.07***	-0.03	-0.05**	-0.02	-0.05***	0.13***	1.00	
11 Stunted	-0.02	-0.03	0.03	0.05***	0.01	-0.06***	-0.06***	-0.03	-0.03	-0.00	1.00
1.1: Head-knee task	0.43***	0.05**	-0.01	-0.05***	0.01	0.05**	0.03	0.03	-0.02	-0.03	-0.09***
1.2: Forward digit span	0.47***	0.04	-0.02	-0.05***	0.04	0.03	0.04	0.04	-0.00	-0.05***	-0.09***
1.3: DCCS	0.49***	0.05**	0.00	-0.04	-0.01	0.05***	0.07***	0.02	-0.03	-0.06***	-0.10***
1.4: Cancellation task	0.57***	0.04	-0.00	-0.09***	-0.01	0.08***	0.11***	0.06***	-0.02	-0.06***	-0.09***
2.1: Receptive vocabulary	0.57***	-0.00	-0.01	-0.12***	0.03	0.09***	0.09***	0.09***	-0.01	-0.07***	-0.10***
2.2: Expressive language	0.47***	0.02	-0.07***	-0.10***	0.05***	0.06***	0.08***	0.11***	0.03	-0.05*	-0.07***
2.3: Receptive language (short story)	0.53***	0.04	-0.03	-0.10***	0.04	0.06***	0.07***	0.07***	-0.02	-0.07***	-0.11***
2.4: Knowledge of reading concepts	0.33***	0.05**	-0.04	-0.07***	0.01	0.05**	0.05*	0.05**	-0.02	-0.05***	-0.08***
2.5: Letter recognition test	0.20***	0.02	-0.04	-0.06***	-0.01	0.06***	0.06***	0.08***	-0.00	-0.01	-0.06***
2.6: Shape drawing test	0.68***	0.05**	-0.01	-0.09***	0.01	0.05***	0.05***	0.06***	-0.05*	-0.07***	-0.09***
3.1: Measurement vocabulary	0.45***	-0.01	-0.00	-0.08***	0.04	0.04	0.05**	0.05**	0.01	-0.05**	-0.06***
3.2: Verbal counting	0.48***	0.06***	-0.02	-0.08***	0.03	0.06***	0.06***	0.10***	-0.02	-0.07***	-0.10***
3.3: Quantitative comparison	0.58***	0.04	-0.01	-0.09***	0.03	0.07***	0.06***	0.08***	-0.02	-0.07***	-0.09***
3.4: Number identification	0.36***	0.04	-0.04	-0.08***	-0.00	0.06***	0.05***	0.08***	-0.00	-0.04	-0.10***
3.5: Shape identification	0.25***	0.00	0.02	-0.02	0.01	0.02	0.01	0.03	-0.01	-0.04	-0.02
4.1: SDQ: emotional symptoms	-0.04	0.04	-0.01	0.07***	0.05***	-0.11***	-0.07***	-0.02	0.03	0.16***	0.02
4.2: SDQ: conduct problems	-0.07***	-0.07***	-0.00	0.02	-0.02	-0.03	-0.01	-0.07***	-0.01	0.35***	-0.00
4.3: SDQ: hyperactivity/inattention	-0.14***	-0.08***	-0.01	0.06***	0.04	-0.07***	0.02	-0.17***	-0.06***	0.27***	-0.00
4.4: SDQ: peer relationship problems	-0.09***	-0.04	-0.02	0.02	0.02	-0.05**	-0.08***	-0.03	-0.07***	0.08***	0.01

Notes: The table shows pairwise correlations (Sidac-corrected for multiple testing). Two-tailed p-values: $*p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$.

Figure 7.2: Empirical distribution of composite and raw test scores.



7.C.3 Sensitivity analyses

In Table 7.8, we repeated the same exercise as in Table 7.5 but without baseline controls and observed similar interaction effect sizes. Although the interaction term for the language outcome (-0.25 ,$p = .139$) is no longer statistically significant at the 10% level, it is relatively unchanged in magnitude. For all the regression analyses so far, we used standard (Huber-Eicker-White) village-level clustered standard errors, i.e., we assumed that our standard errors might be correlated for individuals within the same village but not between villages. This assumption might not be sufficient for consistent estimates if the error term is correlated with the covariates, despite using a broad set of control variables. Although we cannot control for this entirely with an observational study design, we run a robustness test with village fixed effects in Table 7.9, column (1). Although the standard errors slightly decrease, the point estimates remain robust. We therefore conclude that the village-specific error term components (e.g., due to the presence of ethnic minorities or effects of the preschool construction program) are not systematically correlated with our covariates. So far in our analysis, we have ignored that some of the children in our sample are enrolled in preschool (enrollment rates for children ages 3, 4, and 5 are 29.6%, 55.2%, and 76.6%, respectively). Controlling for the effect of preschool enrollment in our analysis is not trivial, since the relationship between enrollment and child outcomes is bidirectional, particularly in a country like Cambodia. The preschools in our study sample accommodate up to 25 children ages 3 to 5 in a small classroom with one teacher. In this environment, enrollment in preschool strongly depends on the self-sufficiency and emotional maturity of the child. For example, the most important reasons for not enrolling their child, reported by the caregivers in our sample, are that they are afraid the child might get hurt on the way to school (17% of the caregivers), there is no one to bring and pick up the child (17%), the child is too young (17%), enrollment was turned down (13%), and the child refuses to go/cries/is afraid (11%). Teachers also reported that preschool enrollment is usually turned down when a child is not mature enough to stay without close supervision. It is very likely that adding preschool enrollment into our regression model would control for observed and unobserved components of cognitive and socioemotional development and hence also control for the effect of parenting to some extent. Nevertheless, we conducted a robustness test with preschool enrollment in the model. The estimates are shown in Table 7.9, column (2). As expected, the effect sizes become slightly smaller, but the conclusions do not change. Since our tests were also conducted among non-Khmer speaking ethnic minorities, we added a robustness test in Table 7.9, column (3), to show that our findings do not change when non-Khmer speaking households are excluded from the sample. Further, since three different cognitive measures were used throughout this study and it was not accounted for in multiple hypothesis testing so far, we test whether the same associations can be observed when a summary measure of cognitive skills is used. We constructed a cognition index by using the first principal component of the individual test scores from the domains of executive function, language, and early numeracy and repeat our analyses. We used this index to repeat the analyses from Tables 4 and 5. The results are shown in Table 7.10. As expected, the associations with the parenting measures are still significant. Yet, the cognitive parenting-stunting interaction effect is muted compared with the regressions on language and early numeracy and no longer significant (-0.0158, $p = .237$). Hence, a significant parenting-stunting interaction effect can only be confirmed for two of the three domains of cognitive development but not for the summary measure of cognitive development. Finally, we also tested whether our results are robust to different cutoff points for the stunting indicator. As Perumal et al. (2018) point out, there is no biological basis for the commonly used cutoff at -2 SD in the height-for-age z-score (HAZ) and children slightly below and above the cutoff might be equally affected by undesirable health outcomes due to growth impairments. The results are reported in Table 7.11. Using a binary indicator equal to one if $HAZ < -1.5$ ($M = 0.592$) yields smaller and insignificant estimates for the cognitive parenting-stunted interaction term in Table 7.5, column 3, for language (-0.0233, $p = .183$) and early numeracy development (-0.0165, $p = .396$). Yet, using a lower cutoff score of $HAZ < -2.5$ ($M = 0.201$) yields slightly larger and significant interaction effects on language (-0.0311, $p = .057$) and early numeracy development (-0.0433, $p = .014$) than the results in Table 7.5. The same pattern is observed without baseline control variables. This demonstrates that the weaker association of cognitive parenting and test scores is predominantly driven by children in the lower tail of the HAZ distribution.

Table 7.8: Results of hierarchical regression analysis predicting different domains of child development using a cognitive parenting x stunted and cognitive parenting x caregiver education interaction term

	Executive function			Language			Early numeracy			Socio-emotional problems		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Age (years)	0.770***	0.770***	0.771***	0.778***	0.778***	0.778***	0.723***	0.723***	0.723***	-0.153***	-0.153***	-0.153***
Female	0.116***	0.116***	0.116***	0.084***	0.084***	0.084***	0.079***	0.079***	0.079***	-0.041*	-0.041*	-0.041*
Household size	-0.010*	-0.010*	-0.010*	-0.033***	-0.033***	-0.033***	-0.013**	-0.013**	-0.013**	-0.021***	-0.021***	-0.021***
Low living standard	-0.148***	-0.148***	-0.148***	-0.208***	-0.208***	-0.208***	-0.159***	-0.159***	-0.159***	0.052**	0.053**	0.052**
Caregiver age	0.003***	0.003***	0.003***	0.005***	0.005***	0.005***	0.004***	0.004***	0.004***	0.002**	0.002**	0.002**
Caregiver primary school	0.103***	0.103***	0.096***	0.120***	0.119***	0.115***	0.097***	0.097***	0.092***	-0.139***	-0.138***	-0.132***
Caregiver non-verbal reasoning	0.069***	0.069***	0.069***	0.078***	0.078***	0.078***	0.053***	0.053***	0.053***	-0.006	-0.006	-0.007
Cognitive parenting	0.032***	0.033**	0.022	0.098***	0.107***	0.101***	0.091***	0.104***	0.097***	-0.051***	-0.056***	-0.046**
Socio-emotional parenting	0.021*	0.021*	0.022*	0.001	0.002	0.002	0.006	0.007	0.007	-0.077***	-0.077***	-0.077***
Negative parenting	-0.026**	-0.026**	-0.026**	-0.023**	-0.023**	-0.023**	-0.033***	-0.033***	-0.033***	0.346***	0.346***	0.346***
Stunted	-0.186***	-0.186***	-0.186***	-0.182***	-0.182***	-0.182***	-0.153***	-0.154***	-0.154***	0.004	0.004	0.004
Cognitive parenting **x** Stunted		-0.001	0.001		-0.025	-0.024		-0.035*	-0.034*		0.015	0.013
Cognitive parenting **x** Caregiver primary school			0.030			0.016			0.022			-0.027
Constant	-3.409***	-3.409***	-3.409***	-3.402***	-3.402***	-3.402***	-3.232***	-3.232***	-3.232***	0.721***	0.721***	0.721***
Observations	6699	6699	6699	6699	6699	6699	6699	6699	6699	6704	6704	6704
R^2	0.464	0.464	0.464	0.494	0.494	0.494	0.416	0.416	0.416	0.159	0.159	0.159
F	599.5***	554.2***	513.2***	372.1***	340.8***	320.8***	216.5***	199.1***	184.2***	116.6***	106.9***	99.00**
$R^2\Delta$		0.000	0.000		0.000	0.000		0.000	0.000		0.000	0.000
$F\Delta$		0.005	2.039		2.274	0.606		3.467*	0.894		0.337	1.138

Notes: Changes in the R^2 ($R^2\Delta$) and F-tests for joint significance of additional variables ($F\Delta$) refer to a comparison of (2) vs. (1) in column (2) and (3) vs. (2) in column (3). Two-tailed p-values for t-tests and F-tests using robust standard errors clustered at the village level. $*p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$.

Table 7.9: Robustness checks for the results from Table 5 using village fixed effects (1), controlling for preschool enrollment (2), and using Khmer-speaking households only (3)

	Executive function			Language			Early numeracy			Socio-emotional problems		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Age (years)	0.770***	0.770***	0.771***	0.778***	0.778***	0.778***	0.723***	0.723***	0.723***	-0.153***	-0.153***	-0.153***
Female	0.116***	0.116***	0.116***	0.084***	0.084***	0.084***	0.079***	0.079***	0.079***	-0.041*	-0.041*	-0.041*
Household size	-0.010*	-0.010*	-0.010*	-0.033***	-0.033***	-0.033***	-0.013**	-0.013**	-0.013**	-0.021***	-0.021***	-0.021***
Low living standard	-0.148***	-0.148***	-0.148***	-0.208***	-0.208***	-0.208***	-0.159***	-0.159***	-0.159***	0.052**	0.053**	0.052**
Caregiver age	0.003***	0.003***	0.003***	0.005***	0.005***	0.005***	0.004***	0.004***	0.004***	0.002**	0.002**	0.002**
Caregiver primary school	0.103***	0.103***	0.096***	0.120***	0.119***	0.115***	0.097***	0.097***	0.092***	-0.139***	-0.138***	-0.132***
Caregiver non-verbal reasoning	0.069***	0.069***	0.069***	0.078***	0.078***	0.078***	0.053***	0.053***	0.053***	-0.006	-0.006	-0.007
Cognitive parenting	0.032***	0.033**	0.022	0.098***	0.107***	0.101***	0.091***	0.104***	0.097***	-0.051***	-0.056***	-0.046**
Socio-emotional parenting	0.021*	0.021*	0.022*	0.001	0.002	0.002	0.006	0.007	0.007	-0.077***	-0.077***	-0.077***
Negative parenting	-0.026**	-0.026**	-0.026**	-0.023**	-0.023**	-0.023**	-0.033***	-0.033***	-0.033***	0.346***	0.346***	0.346***
Stunted	-0.186***	-0.186***	-0.186***	-0.182***	-0.182***	-0.182***	-0.153***	-0.154***	-0.154***	0.004	0.004	0.004
Cognitive parenting **x** Stunted		-0.001	0.001		-0.025	-0.024		-0.035*	-0.034*		0.015	0.013
Cognitive parenting **x** Caregiver primary school			0.030			0.016			0.022			-0.027
Constant	-3.409***	-3.409***	-3.409***	-3.402***	-3.402***	-3.402***	-3.232***	-3.232***	-3.232***	0.721***	0.721***	0.721***
Observations	6699	6699	6699	6699	6699	6699	6699	6699	6699	6704	6704	6704
R^2	0.464	0.464	0.464	0.494	0.494	0.494	0.416	0.416	0.416	0.159	0.159	0.159
F	599.5***	554.2***	513.2***	372.1***	340.8***	320.8***	216.5***	199.1***	184.2***	116.6***	106.9***	99.00**
$R^2\Delta$		0.000	0.000		0.000	0.000		0.000	0.000		0.000	0.000
$F\Delta$		0.005	2.039		2.274	0.606		3.467*	0.894		0.337	1.138

Notes: Two-tailed p-values for t-tests using robust standard errors clustered at the village level. $*p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$.

Table 7.10: Robustness checks for results from Tables 7.4 and 7.5 for a cognition index as the outcome variable

	Cognition index					
	(1)	(2)	(3)	(4)	(5)	(6)
Age (years)	0.855***	0.850***	0.853***	0.495***	0.495***	0.495***
Female	0.104***	0.106***	0.099***	0.073***	0.073***	0.073***
Household size	-0.022***	-0.024***	-0.021***	-0.012***	-0.012***	-0.012***
Low living standard	-0.208***	-0.211***	-0.195***	-0.110***	-0.110***	-0.110***
Caregiver age	0.005***	0.004***	0.005***	0.002***	0.002***	0.002***
Caregiver primary school	0.128***	0.162***	0.120***	0.076***	0.076***	0.075***
Caregiver non-verbal reasoning	0.078***	0.080***	0.074***	0.052***	0.052***	0.053***
Cognitive parenting	0.083***		0.083***	0.051***	0.057***	0.055***
Socio-emotional parenting	0.013		0.012	0.005	0.006	0.006
Negative parenting	-0.031***		-0.032***	-0.021***	-0.021***	-0.021***
Stunted		-0.197***	-0.196***	-0.119***	-0.120***	-0.120***
Child executive function (baseline)				0.097***	0.097***	0.097***
Child language (baseline)				0.257***	0.257***	0.257***
Child early numeracy (baseline)				0.163***	0.164***	0.164***
Socio-emotional problems (baseline)				0.006	0.006	0.006
Cognitive parenting **x** Stunted					-0.016	-0.016
Cognitive parenting **x** Caregiver primary school						0.004
Constant	-3.837***	-3.728***	-3.758***	-2.174***	-2.174***	-2.174***
Observations	6665	6665	6665	6665	6665	6665
R^2	0.567	0.568	0.576	0.695	0.695	0.695

Notes: Two-tailed p-values for t-tests using robust standard errors clustered at the village level. $*p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$.

Table 7.11: Robustness checks for results from Table 7.5 with different cutoff points for the stunting indicator

	Executive function				Language				Early numeracy				Socio-emotional problems			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Age (years)	0.501***	0.502***	0.501***	0.501***	0.423***	0.423***	0.423***	0.423***	0.378***	0.378***	0.378***	0.378***	-0.101***	-0.101***	-0.100***	-0.101***
Female	0.096***	0.096***	0.094***	0.094***	0.060***	0.060***	0.058***	0.058***	0.053***	0.053***	0.052***	0.052***	-0.033	-0.032	-0.033	-0.033
Household size	-0.004	-0.004	-0.003	-0.004	-0.023***	-0.023***	-0.023***	-0.023***	-0.004	-0.004	-0.004	-0.004	-0.017**	-0.017**	-0.017**	-0.017**
Low living standard	-0.085***	-0.085***	-0.085***	-0.085***	-0.123***	-0.123***	-0.122***	-0.122***	-0.080***	-0.080***	-0.081***	-0.081***	0.046*	0.046*	0.046*	0.046*
Caregiver age	0.001	0.001	0.001	0.001	0.003***	0.003***	0.003***	0.003***	0.002***	0.002***	0.002***	0.002***	0.002*	0.002*	0.002*	0.002*
Caregiver primary school	0.075***	0.070***	0.076***	0.071***	0.074***	0.074***	0.075***	0.075***	0.053**	0.052**	0.054**	0.053**	-0.120***	-0.116***	-0.120***	-0.116***
Caregiver non-verbal reasoning	0.052***	0.053***	0.051***	0.051***	0.057***	0.057***	0.056***	0.056***	0.031***	0.031***	0.030***	0.030***	-0.001	-0.001	-0.001	-0.001
Cognitive parenting	0.016	0.009	0.015	0.008	0.080***	0.080***	0.072***	0.072***	0.068***	0.065***	0.066***	0.064***	-0.036*	-0.030	-0.035**	-0.030*
Socio-emotional parenting	0.017	0.017	0.018*	0.018*	-0.005	-0.005	-0.005	-0.005	0.000	0.000	0.001	0.001	-0.067***	-0.067***	-0.067***	-0.067***
Negative parenting	-0.019*	-0.019*	-0.020**	-0.020**	-0.011	-0.011	-0.012	-0.012	-0.022**	-0.022**	-0.023**	-0.023**	0.311***	0.311***	0.311***	0.311***
Stunted (haz<=-1.5)	-0.113***	-0.113***			-0.087***	-0.087***			-0.073***	-0.073***			-0.020	-0.020		
Stunted (haz<=-2.5)			-0.164***	-0.164***			-0.134***	-0.134***			-0.092***	-0.092***			-0.027	-0.027
Child executive function (baseline)	0.166***	0.166***	0.163***	0.164***	0.063***	0.063***	0.060***	0.060***	0.030*	0.030*	0.029*	0.029*	0.018	0.017	0.017	0.017
Child language (baseline)	0.115***	0.115***	0.116***	0.115***	0.329***	0.329***	0.329***	0.329***	0.246***	0.246***	0.247***	0.247***	-0.043**	-0.042**	-0.043**	-0.043**
Child early numeracy (baseline)	0.114***	0.114***	0.114***	0.114***	0.115***	0.115***	0.114***	0.114***	0.234***	0.234***	0.235***	0.235***	-0.013	-0.013	-0.013	-0.013
Socio-emotional problems (baseline)	0.004	0.004	0.003	0.003	0.003	0.003	0.003	0.003	0.002	0.002	0.002	0.002	0.197***	0.196***	0.197***	0.197***
Cognitive parenting **x** Stunted (haz<=-1.5)	-0.007	-0.006			-0.023	-0.023			-0.016	-0.016			-0.008	-0.009		
Cognitive parenting **x** Stunted (haz<=-2.5)			-0.015	-0.014			-0.031*	-0.031*			-0.043**	-0.043**			-0.028	-0.030
Cognitive parenting x Caregiver primary school		0.020		0.020		0.000		-0.000		0.007		0.005		-0.016		-0.017
Constant	-2.198***	-2.199***	-2.228***	-2.229***	-1.821***	-1.821***	-1.846***	-1.846***	-1.693***	-1.693***	-1.716***	-1.716***	0.491***	0.492***	0.484***	0.485***
Observations	6699	6699	6699	6699	6699	6699	6699	6699	6699	6699	6699	6699	6704	6704	6704	6704
R^2	0.530	0.530	0.531	0.531	0.611	0.611	0.612	0.612	0.538	0.538	0.538	0.538	0.197	0.198	0.198	0.198

Notes: Two-tailed p-values for t-tests using robust standard errors clustered at the village level. $*p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$.

Table 7.12: Results of multivariate regression of parenting scores on explanatory variables

	Cognitive parenting	ocio-emotional parentir	Negative parenting
Child age (years)	-0.033**	-0.126***	-0.057***
Child female	0.017	-0.031	-0.196***
Household size	-0.032***	-0.031***	0.014*
Low living standard	-0.160***	-0.088***	0.071**
Caregiver age	-0.005***	0.001	-0.003***
Caregiver primary school	0.436***	0.188***	-0.130***
Caregiver non-verbal reasoning	0.067***	0.049***	-0.009
Stunted	-0.017	-0.038	-0.020
Constant	0.445***	0.666***	0.446***
Observations	6706	6706	6706
R^2	0.085	0.031	0.019

Notes: Two-tailed p-values for t-tests using robust standard errors clustered at the village level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

CHAPTER 8

Conclusion

The six main chapters in this dissertation cover different questions of household decision-making and educational attainment. In particular, it focuses on questions of early childhood development, the role of information of education decisions, the role of social norms for household decisions, and the use and validation of detailed micro-level primary data. In this last chapter, conclusions are drawn, research and policy implications, and areas for further research are discussed.

Chapter 2 extends the literature on effects of information provision for education choice with experimental evidence on the effects of information provision on beliefs about postgraduate returns, enrollment intentions, and realized enrollment into postgraduate education. A significant updating of beliefs about postgraduate returns is observed half a year later. The downward-adjustment of expectations regarding the postgraduate premium is driven by males. For these, corresponding changes in enrollment intentions six months after treatment are observed as well. The results also suggest that the information provision materialised into differences in realised postgraduate enrollment two years after initial treatment. Overall, the results show that even for undergraduate students, information frictions regarding returns to postgraduate education exist.

Policywise, the findings of the study suggest that information about labour market consequences of tertiary education decisions could be provided systematically to effectively reduce information constraints. The online-treatment could be scaled up at low costs which makes it a particularly attractive policy instrument.

However, no impacts were measured for girls and no significant heterogeneity was found by socio-economic background. One possible explanation for this is that other factors than information constraints cause gaps by gender or socio-economic background in the postgraduate education decisions of German undergraduate students.

One limitation of the employed study design is that effects of information can only be estimated for groups where significant belief updating takes place. Hence, this study can only draw conclusions about the effect of information updating on education decisions for men. It is possible that with a larger sample to identify effects for subgroups more precisely, more refined policy conclusions would be possible. Further, with even more detailed information, more specific information constraints might be identified. The significant effects of this study highlight the importance of a larger replication study. In addition, a longer term follow-up might be worthwhile to conduct since some study participants had not yet graduated at the two-year follow-up and full impacts on their enrollment might show up later.

Chapter 3 studies if the amount of individual housework contribution depends on the spouses labour supply. To study the relationship between time spent for market work and non-market work the analysis builds on time-use data from the GDR as well as reunified Germany. The results show that women in the GDR worked more in the labour market and less in the household than in Western Germany. The key finding from the analysis is that once individual labour supply (for market work) is accounted for, the gender housework gap becomes identical for East and West Germany. This complements recent findings that males do not adjust their own labour supply when their wives start working. As a result, households in the GDR/East and West Germany only differ in their household-level time allocation because of the different amounts of time women spend on the labour market, and not because of different, i.e. more egalitarian, behaviour in the household.

Labour market policies that strengthen women's economic position by increasing their hours of market work run the risk of putting an unintended burden on women if their spouse does not provide additional support with housework. At least the extreme labour market policies and propaganda imposed by the Soviet rule and GDR over East Germans, for a duration of four decades, did not make partners' time allocation to housework more responsive to their spouses labour supply, and thus lead to gender inequality in the domestic domain.

The findings suggest that time allocation to market work and non-market work are governed by a separate set of norms. Theories of household allocation should take this into account, in particular to explain the missing partner reactions. If preferences and social norms about housework evolve isolated from other domains of gender equality such as market work, policies can have unintended, detrimental consequences for women.

The findings of **Chapters 4 to 7** highlight how strongly questions about parenting, nutrition and early childhood education are connected, at least in the studied low-income country context. The results of **Chapter 4** show that constructing the community preschools increased enrollment and moved children who had been attending informal schools into structurally better schools, as intended. However, the information campaigns didn't further increase demand, and no large average impacts on short- (one-year) or medium-term (two-year) impacts on child cognitive or socio-emotional outcomes were observed. The findings highlight that, notwithstanding the robust evidence that high-quality preschool experiences boost child development, it is difficult to engender those experiences in a program at scale in a low-income, capacity-constrained environment.

While the results are context-specific, they nevertheless highlight the general importance of childcare arrangements counterfactual to preschool attendance. These arrangements are common in many countries and include family care along with informal preschool arrangements. **Chapter 5** highlights that in order to substantively impact child development outcomes, formal preschool provision needs to deliver substantial value added over the counterfactual. Ideally, targeted programs take this into account and try to reach children for which the shift to the new program is a substantial improvement in the environmental stimuli.

While the documented positive effects of the evaluated preschool construction program on cognitive development on children from the highest wealth quartile are encouraging, they also imply that the program further increased wealth gradients in early development. **Chapter 7** demonstrates that children who have experienced undernutrition are less sensitive to parenting and, moreover, that parenting interventions may have a large positive influence on child development but are unlikely to compensate fully for the negative impacts of early undernutrition. The same mechanism can explain why the preschool program did not improve children's cognitive and socio-emotional development. This highlights the importance of designing and testing integrated program models of child development that anticipate the complex interactions that will arise between children's characteristics and their environment.

Even ECEC services, starting as early as the age of three, can be considered a compensatory investment if socio-economic gradients in early development are prevalent at such early age. Quality is key to improving child development through ECEC services and particularly key to foster development of the most disadvantaged children – which are often not reached by high quality ECEC services (Stahl et al., 2018). The results of **Chapters 4 and 5** suggest that in Cambodia, and likely in other similar contexts, a key focus should be on improving the process quality aspects of preschool, namely improvements in teacher pedagogical skills and ultimately teacher-child interactions. A key place to start in Cambodia would be the training of community preschool teachers — which is much shorter and less intensive than the training the formal preschool teachers received.

Chapter 6 highlights the association between preschool teacher training and process quality. Similarly, the large pay-differential between the two types of teachers, which might have affected the selection of trained teachers, their motivation and their ability to stay in the job. While structural quality has been shown to play a smaller role than process quality for child development in other contexts (Andrew et al., 2019), the analysis of associations between detailed measures of preschool quality and child development in Cambodia has shown that severe, regularly occurring, deficiencies in structural quality are detrimental to child development as well. To maintain and further improve the quality of the current preschool infrastructure, active monitoring and ongoing quality assurance is required to ensure optimal effects on child development. While the findings have shown that associations between process quality and child development can be low, the results do not imply that process quality has a low relevance for child development in general. It is material for further research if an improvement in process quality levels that goes beyond the ones observed in this study would make a significant difference to child development.

Bibliography

Alkire, S. and M.E. Santos (2010): Acute Multidimensional Poverty: A New Index for Developing Countries, *United Nations Development Programme Human Development Report Office Background Paper* (2010/11).

Almond, D., J. Currie and V. Duque (2018): Childhood Circumstances and Adult Outcomes: Act II, *Journal of Economic Literature*, 56(4), 1360–1446.

Altonji, J., P. Arcidiacono and A. Maurel (2016): The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects, in: E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education*: Elsevier, Vol. 5, Chapter 7, pp. 305 – 396.

Altonji, J.G. and R.M. Blank (1999): Race and gender in the labor market, *Handbook of Labor Economics*, 3, 3143–3259.

Anderson, M.L. (2008): Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects, *Journal of the American Statistical Association*, 103(484), 1481–1495.

Andrew, A., O. Attanasio, R. Bernal, L.C. Sosa, S. Krutikova and M. Rubio-Codina (2019): Preschool Quality and Child Development, Working Paper No. 26191, National Bureau of Economic Research.

Andrew, A., O. Attanasio, E. Fitzsimons, S. Grantham-McGregor, C. Meghir and M. Rubio-Codina (2018): Impacts 2 Years After a Scalable Early Childhood Development Intervention to Increase Psychosocial Stimulation in the Home: A Follow-Up of a Cluster Randomised Controlled Trial in Colombia, *PLoS medicine*, 15(4), e1002556.

Angrist, J.D., G.W. Imbens and D.B. Rubin (1996): Identification of causal effects using instrumental variables, *Journal of the American Statistical Association*, 91(434), 444–455.

Araujo, M.C., P. Carneiro, Y. Cruz-Aguayo and N. Schady (2016): Teacher Quality and Learning Outcomes in Kindergarten, *The Quarterly Journal of Economics*, 131(3), 1415–1453.

Araujo, M.C., M. Dormal and N. Schady (2019): Childcare Quality and Child Development, *Journal of Human Resources*, 54(3), 656–682.

Autorengruppe Bildungsberichterstattung (2018): *Bildung in Deutschland 2018. Ein indikatorengestützter Bericht mit einer Analyse zu Wirkungen und Erträgen von Bildung*: Bielefeld: Bertelsmann.

Autorengruppe Bildungsberichterstattung (2020): *Bildung in Deutschland 2020. Ein indikatorengestützter Bericht mit einer Analyse zur Bildung in einer digitalisierten Welt*: Bielefeld: WBV.

Baker, M., J. Gruber and K. Milligan (2008): Universal Child Care, Maternal Labor Supply, and Family Well-Being, *Journal of Political Economy*, 116(4), 709–745.

Baker, M., J. Gruber and K. Milligan (2019): The long-run impacts of a universal child care program, *American Economic Journal: Economic Policy*, 11(3), 1–26.

Barth, D., J. Jessen, C.K. Spieß and K. Wrohlich (2020): Mothers in Eastern and Western Germany: Employment rates and attitudes are converging, full-time employment is not, *DIW Weekly Report*, 10(38), 403–410.

Bassok, D., A.J. Markowitz, D. Player and M. Zagardo (2017): Do parents know "high quality" preschool when they see it, in: *EdPolicyWorks Working Paper Series No. 54. University of Virginia Curry School.*

Bauernschuster, S. and H. Rainer (2012): Political regimes and the family: how sex-role attitudes continue to differ in reunified Germany, *Journal of Population Economics*, 25(1), 5–27.

Beblo, M. (2001): *Bargaining over time allocation: Economic modeling and econometric investigation of time use within families*: Springer Science & Business Media.

Beblo, M. and L. Görges (2018): On the nature of nurture. The malleability of gender differences in work preferences, *Journal of Economic Behavior & Organization*, 151, 19–41.

Becker, G.S. (1962): Investment in human capital: A theoretical analysis, *Journal of Political Economy*, 70(5, Part 2), 9–49.

Becker, G.S. (1981): *A Treatise on the Family*: Harvard university press.

Belfield, C., T. Boneva, C. Rauh and J. Shaw (2019): What Drives Enrolment Gaps in Further Education? The Role of Beliefs in Sequential Schooling Decisions, *Economica* (online first 11. April 2019).

Belfield, C.R., M. Nores, S. Barnett and L. Schweinhart (2006): The high/scope perry preschool program cost–benefit analysis using data from the age-40 followup, *Journal of Human Resources*, 41(1), 162–190.

Belloni, A., V. Chernozhukov and C. Hansen (2013): Inference on Treatment Effects after Selection among High-Dimensional Controls†, *The Review of Economic Studies*, 81(2), 608–650.

Belsky, J., M.J. Bakermans-Kranenburg and M.H. Van IJzendoorn (2007): For better and for worse: Differential susceptibility to environmental influences, *Current Directions in Psychological Science*, 16(6), 300–304.

Bentley, M.E., H.M. Wasser and H.M. Creed-Kanashiro (2011): Responsive feeding and child undernutrition in low-and middle-income countries, *The Journal of Nutrition*, 141(3), 502–507.

Berkes, J. and A. Bouguen (2019): Heterogeneous Preschool Impact and Close Substitutes: Evidence from a Preschool Construction Program in Cambodia, *UC Berkeley: Center for Effective Global Action.*.

Berkes, J., A. Bouguen and D. Filmer (2017). *Increasing Early Childhood Care and Development Through Community Preschools in Cambodia: Evaluating the Impacts.* `https://doi.org/10.1257/rct.1045-3.0`: AEA RCT Registry.

Berkes, J., A. Bouguen, D. Filmer and T. Fukao (2019a): Combining Supply and Demand-side Interventions : Evidence from a Large Preschool Program in Cambodia : Impact Evaluation Final Report, *http://documents.worldbank.org/curated/en/110771561664132545/Impact-Evaluation-Final-Report, last accessed: 20-04-2021.*

Berkes, J., A. Bouguen, D. Filmer and T. Fukao (2019b): Improving Preschool Provision and Encouraging Demand : Heterogeneous Impacts of a Large-Scale Program, Impact Evaluation series, Washington, D.C. : World Bank Group.

Berkes, J., A. Raikes, A. Bouguen and D. Filmer (2019): Joint Roles of Parenting and Nutritional Status for Child Development: Evidence from Rural Cambodia, *Developmental Science*, 22(5), e12874.

Berlinski, S., S. Galiani and P. Gertler (2009): The Effect of Pre-Primary Education on Primary School Performance, *Journal of Public Economics*, 93(1), 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), 1416–1440.

Bernal, R., O. Attanasio, X. Peña and M. Vera-Hernández (2019): The Effects of the Transition from Home-Based Childcare to Childcare Centers on Children's Health and Development in Colombia, *Early Childhood Research Quarterly*, 47, 418–431.

Bertrand, M. (2018): Coase lecture–the glass ceiling, *Economica*, 85(338), 205–231.

Bertrand, M., E. Kamenica and J. Pan (2015): Gender identity and relative income within households, *The Quarterly Journal of Economics*, 130(2), 571–614.

Bettinger, E.P. and R.B. Baker (2014): The Effects of Student Coaching – An Evaluation of a Randomized Experiment in Student Advising, *Educational Evaluation and Policy Analysis*, 36(1), 3–19.

Bettinger, E.P., B.T. Long, P. Oreopoulos and L. Sanbonmatsu (2012): The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment, *The Quarterly Journal of Economics*, 127(3), 1205–1242.

Bicchieri, C., R. Muldoon and A. Sontuoso (2018): Social Norms, , Philosophy, Politics and Economics, University of Pennsylvania.

Black, M.M., R. Pérez-Escamilla and S. Fernandez Rao (2015): Integrating nutrition and child development interventions: scientific basis, evidence of impact, and implementation considerations, *Advances in Nutrition*, 6(6), 852–859.

Black, M.M., S.P. Walker, L.C. Fernald, C.T. Andersen, A.M. DiGirolamo, C. Lu, D.C. McCoy, G. Fink, Y.R. Shawar, J. Shiffman et al. (2017): Early Childhood Development Coming of Age: Science through the Life Course, *The Lancet*, 389(10064), 77–90.

Black, M.M., D.P. Yimgang, K.M. Hurley, K.B. Harding, S. Fernandez-Rao, N. Balakrishna, K.V. Radhakrishna, G.A. Reinhart and K.M. Nair (2019): Mechanisms linking height to early child development among infants and preschoolers in rural India, *Developmental Science*, 22(5), e12806.

Blair, C. and C.C. Raver (2012): Child development in the context of adversity: experiential canalization of brain and behavior., *American Psychologist*, 67(4), 309.

Blau, D. and J. Currie (2006): Pre-School, Day Care, and After-School Care: Who's Minding the Kids?, *Handbook of the Economics of Education*, 2, 1163–1278.

Blimpo, M.P., P.M. Amaro Da Costa Luz Carneiro, P. Jervis and T. Pugatch (2019). Improving Access and Quality in Early Childhood Development Programs: Experimental Evidence from The Gambia.

Blossfeld, H.P., J. von Maurice and T. Schneider (2011): The National Education Panel Study: need, main features and research potential, *Zeitrschrift für Erziehungswissenschaft*, 14, 5–17.

Blundell, R., L. Dearden, C. Meghir and B. Sianesi (1999): Human capital investment: the returns from education and training to the individual, the firm and the economy, *Fiscal Studies*, 20(1), 1–23.

Boelmann, B., A. Raute and U. Schönberg (2020): Wind of Change? Cultural Determinants of Maternal Labor Supply, *CReAM Discussion Paper 20/20*.

Boneva, T., M. Golin, C. Rauh et al. (2019): Can Perceived Returns Explain Enrollment Gaps in Postgraduate Education?, HCEO Working Paper No. 2019-045, University of Chicago.

Boneva, T. and C. Rauh (2017): Socio-Economic Gaps in University Enrollment: The Role of Perceived Pecuniary and Non-Pecuniary Returns, HCEO Working Paper No. 2017-080, University of Chicago.

Bornstein, M.H. and R.H. Bradley (2014): Socioeconomic Status, *Parenting, and Child Development. London: Routledge*.

Bornstein, M.H. and D.L. Putnick (2012): Cognitive and socioemotional caregiving in developing countries, *Child Development*, 83(1), 46–61.

Bouguen, A., D. Filmer, K. Macours and S. Naudeau (2013): Impact Evaluation of Three Types of Early Childhood Development Interventions in Cambodia (English), *Policy Research Working Paper*, IE 97(WPS 6540).

Bouguen, A., D. Filmer, K. Macours and S. Naudeau (2018a): Preschool and Parental Response in a Second Best World: Evidence from a School Construction Experiment, *Journal of Human Resources*, 53(2), 474–512.

Bouguen, A., D. Filmer, K. Macours and S. Naudeau (2018b): Preschool and parental response in a second best world evidence from a school construction experiment, *Journal of Human Resources*, 53(2), 474–512.

Bouguen, A., J. Grenet and M. Gurgand (2017): Does class size influence student achievement?, *Notes IPP* (28).

Bradbury, B., M. Corak, J. Waldfogel and E. Washbrook (2015): *Too many children left behind: The US achievement gap in comparative perspective*: Russell Sage Foundation.

Bradley, R.H. and R.F. Corwyn (2002): Socioeconomic status and child development, *Annual Review of Psychology*, 53(1), 371–399.

Brinkman, S.A., S. Alanna and D. Plamen Nikolov (2016): *The Status of Early Childhood Health and Development in Northern Lao PDR (English)*: Washington, D.C.: World Bank Group.

Brinkman, S.A., A. Hasan, H. Jung, A. Kinnell, N. Nakajima and M. Pradhan (2017): The role of preschool quality in promoting child development: evidence from rural Indonesia, *European Early Childhood Education Research Journal*, 25(4), 1–23.

Brinkman, S.A., A. Hasan, H. Jung, A. Kinnell and M. Pradhan (2017): The Impact of Expanding Access to Early Childhood Education Services in Rural Indonesia, *Journal of Labor Economics*, 35(S1), S305–S335.

Britto, P.R., S.J. Lye, K. Proulx, A.K. Yousafzai, S.G. Matthews, T. Vaivada, R. Perez-Escamilla, N. Rao, P. Ip, L.C. Fernald et al. (2017): Nurturing Care: Promoting Early Childhood Development, *The Lancet*, 389(10064), 91–102.

Britto, P.R., H. Yoshikawa and K. Boller (2011): Quality of Early Childhood Development Programs in Global Contexts: Rationale for Investment, Conceptual Framework and Implications for Equity. Social Policy Report. Volume 25, Number 2., *Society for Research in Child Development*.

Bruhn, M. and D. McKenzie (2009): In Pursuit of Balance: Randomization in Practice in Development Field Experiments, *American Economic Journal: Applied Economics*, 1(4), 200–232.

Busch-Heizmann, A. (2015): Supply-side explanations for occupational gender segregation: Adolescents' work values and gender-(a) typical occupational aspirations, *European Sociological Review*, 31(1), 48–64.

Camehl, G.F., P.S. Schober and C.K. Spiess (2018): Information asymmetries between parents and educators in German childcare institutions, *Education Economics*, 26(6), 624–646.

Campa, P. and M. Serafinelli (2019): Politico-Economic Regimes and Attitudes: Female Workers under State Socialism, *Review of Economics and Statistics*, 101(2), 233–248.

Campbell, F.A., C.T. Ramey, E. Pungello, J. Sparling and S. Miller-Johnson (2002): Early Childhood Education: Young Adult Outcomes From the Abecedarian Project, *Applied Developmental Science*, 6(1), 42–57.

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

Carneiro, P., J.J. Heckman and E.J. Vytlacil (2011): Estimating Marginal Returns to Education, *The American Economic Review*, 101(6), 2754–2781.

Carneiro, P., M. Lokshin and N. Umapathi (2017): Average and Marginal Returns to Upper Secondary Schooling in Indonesia, *Journal of Applied Econometrics*, 32(1), 16–36.

Carpenter, P.A., M.A. Just and P. Shell (1990): What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test., *Psychological Review*, 97(3), 404.

Carrell, S. and B. Sacerdote (2017): Why Do College-Going Interventions Work?, *American Economic Journal: Applied Economics*, 9(3), 124–151.

Castleman, B.L. and B.T. Long (2016): Looking beyond Enrollment: The Causal Effect of Need-Based Grants on College Access, Persistence, and Graduation, *Journal of Labor Economics*, 34(4), 1023–1073.

Castleman, B.L., L.C. Page and K. Schooley (2014): The Forgotten Summer: Does the Offer of College Counseling After High School Mitigate Summer Melt Among College-Intending, Low-Income High School Graduates?, *Journal of Policy Analysis and Management*, 33(2), 320–344.

Chang, L., D. Schwartz, K.A. Dodge and C. McBride-Chang (2003): Harsh parenting in relation to child emotion regulation and aggression., *Journal of Family Psychology*, 17(4), 598.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey and J. Robins (2018): Double/debiased machine learning for treatment and structural parameters, *The Econometrics Journal*, 21(1), C1–C68.

Chetty, R., J.N. Friedman and J.E. Rockoff (2014): Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates, *American Economic Review*, 104(9), 2593–2632.

Coltrane, S. (2000): Research on household labor: Modeling and measuring the social embeddedness of routine family work, *Journal of Marriage and family*, 62(4), 1208–1233.

Conti, G., J. Heckman and S. Urzua (2010): The education-health gradient, *American Economic Review*, 100(2), 234–38.

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.

Cueto, S., J. León, A. Miranda, K. Dearden, B.T. Crookston and J.R. Behrman (2016): Does pre-school improve cognitive abilities among children with early-life stunting? A longitudinal study for Peru, *International journal of Educational Research*, 75, 102–114.

Cuevas, K., K. Deater-Deckard, J. Kim-Spoon, A.J. Watson, K.C. Morasch and M.A. Bell (2014): What's mom got to do with it? Contributions of maternal executive function and caregiving to the development of executive function across early childhood, *Developmental Science*, 17(2), 224–238.

Cunha, F., I.T. Elo and J. Culhane (2013): Eliciting Maternal Expectations About the Technology of Cognitive Skill Formation, *NBER Working Paper* (w19144).

Cunha, F. and J. Heckman (2007): The Technology of Skill Formation, *The American Economic Review*, 97(2), 31.

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, *The Journal of Economic Perspectives*, 15(2), 213–238.

Currie, J. and D. Almond (2011): Human capital development before age five, in: *Handbook of Labor Economics*: Elsevier, Vol. 4, pp. 1315–1486.

Currie, J. and D. Thomas (1995): Does Head Start Make a Difference?, *The American Economic Review*, 341–364.

Datta Gupta, N. and M. Simonsen (2016): Academic Performance and Type of Early Childhood Care, *Economics of Education Review*, 53, 217–229.

Dean, J.T. and S. Jayachandran (2020): Attending kindergarten improves cognitive but not socioemotional development in India.

Delavande, A. and B. Zafar (2019): University choice: The role of expected earnings, nonpecuniary outcomes, and financial constraints, *Journal of Political Economy*, 127(5), 2343–2393.

Diekman, A.B., E.R. Brown, A.M. Johnston and E.K. Clark (2010): Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers, *Psychological Science*, 21(8), 1051–1057.

Duflo, E., R. Glennerster and M. Kremer (2008). *Using Randomization in Development Economics Research: A Toolkit*, Vol. 4 of Handbook of Development Economics, Chapter 61, pp. 3895–3962: Elsevier.

Duncan, G.J. and K. Magnuson (2013): Investing in Preschool Programs, *Journal of Economic Perspectives*, 27(2), 109–132.

Dunn, L.M. and D.M. Dunn (2007): *PPVT-4: Peabody Picture Vocabulary Test*: Pearson Assessments.

Dunn, L.M., E.R. Padilla, D.E. Lugo and L.M. Dunn (1986): *TVIP: Test de Vocabulario en Imágenes Peabody: Adaptación Hispanoamericana*: American Guidance Service.

Dynarski, S. (2002): The behavioral and distributional implications of aid for college, *American Economic Review*, 92(2), 279–285.

Dynarski, S., C.J. Libassi, K. Michelmore and S. Owen (2018): Closing the Gap: The Effect of a Targeted, Tuition-Free Promise on College Choices of High-Achieving, Low-Income Students, NBER Working Paper No. 25349, NBER.

Dynarski, S.M. and J.E. Scott-Clayton (2006): The Cost Of Complexity In Federal Student Aid: Lessons From Optimal Tax Theory And Behavioral Economics, *National Tax Journal*, 59(2), 319–356.

Ehlert, M., F. Peter, C. Finger, A. Rusconi, H. Solga, C.K. Spiess and V. Zambre (2017): The Berliner-Studienberechtigen-Panel (Best Up): Methodological and Data Report, DIW Data Documentation No. 90, DIW Berlin.

Elango, S., J.L. García, J.J. Heckman and A. Hojman (2015). *Early Childhood Education*, pp. 235–297: University of Chicago Press.

Ellis, B.J., W.T. Boyce, J. Belsky, M.J. Bakermans-Kranenburg and M.H. Van IJzendoorn (2011): Differential susceptibility to the environment: An evolutionary–neurodevelopmental theory, *Development and Psychopathology*, 23(1), 7–28.

Engle, P.L., L.C. Fernald, H. Alderman, J. Behrman, C. O'Gara, A. Yousafzai, M.C. de Mello, M. Hidrobo, N. Ulkuer, I. Ertem et al. (2011): Strategies for Reducing Inequalities and Improving Developmental Outcomes for Young Children in Low-Income and Middle-Income Countries, *The Lancet*, 378(9799), 1339–1353.

Fabian, G., J. Hillmann, F. Trennt and K. Briedis (2016): *Hochschulabschlüsse nach Bologna: Werdegänge der Bachelor-und Masterabsolvent(inn)en des Prüfungsjahrgangs 2013*: DZHW Deutsches Zentrum für Hochschul-und Wissenschaftsforschung.

Feller, A., T. Grindal, L. Miratrix and L.C. Page (2016): Compared to what? Variation in the impacts of early childhood education by alternative care type, *Ann. Appl. Stat.*, 10(3), 1245–1285.

Fernald, L.C., P. Kariger, M. Hidrobo and P.J. Gertler (2012): Socioeconomic gradients in child development in very young children: Evidence from India, Indonesia, Peru, and Senegal, *Proceedings of the National Academy of Sciences*, 109(Supplement 2), 17273–17280.

Fernald, L.C., E. Prado, P. Kariger, A. Raikes et al. (2017): A Toolkit for Measuring Early Childhood Development in Low and Middle-Income Countries, *World Bank Publications*.

Fernández, R. (2013): Cultural change as learning: The evolution of female labor force participation over a century, *American Economic Review*, 103(1), 472–500.

Fiebiger, H. (1991): Die Zeitbudgeterhebung 1990 in der ehemaligen DDR, *M. Ehling/Rv Schweitzer, Zeitbudgeterhebung der amtlichen Statistik* (17), 12–38.

Fogli, A. and L. Veldkamp (2011): Nature or nurture? Learning and the geography of female labor force participation, *Econometrica*, 79(4), 1103–1138.

Fort, M., A. Ichino and G. Zanella (2020): Cognitive and noncognitive costs of day care at age 0–2 for children in advantaged families, *Journal of Political Economy*, 128(1), 158–205.

Galasso, E., A. Wagstaff, S. Naudeau and M. Shekar (2016): The economic costs of stunting and how to reduce them, *Policy Research Note World Bank, Washington, DC*.

Gelbach, J.B. (2016): When do covariates matter? And which ones, and how much?, *Journal of Labor Economics*, 34(2), 509–543.

Gershuny, J. and K. Fisher (2013). Multinational Time Use Study.

Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S.M. Chang and S. Grantham-McGregor (2014): Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica, *Science*, 344(6187), 998–1001.

Goldin, C. (2014): A grand gender convergence: Its last chapter, *American Economic Review*, 104(4), 1091–1119.

Goodman, R. (2001): Psychometric properties of the strengths and difficulties questionnaire, *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(11), 1337–1345.

Gordon, R.A., K. Fujimoto, R. Kaestner, S. Korenman and K. Abner (2013): An assessment of the validity of the ECERS-R with implications for measures of child care quality and relations to child development., *Developmental Psychology*, 49(1), 146.

Grantham-McGregor, S.M., C.A. Powell, S.P. Walker and J.H. Himes (1991): Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican Study, *The Lancet*, 338(8758), 1–5.

Grunow, D., K. Begall and S. Buchler (2018): Gender ideologies in Europe: A multidimensional framework, *Journal of Marriage and Family*, 80(1), 42–60.

Gunderson, M. and P. Oreopolous (2020): Returns to education in developed countries, in: *The Economics of Education*: Elsevier, pp. 39–51.

Hakim, C. (2000): *Work-lifestyle choices in the 21st century: Preference theory*: OUp Oxford.

Hamadani, J.D., F. Tofail, S.N. Huda, D.S. Alam, D.A. Ridout, O. Attanasio and S.M. Grantham-McGregor (2014): Cognitive deficit and poverty in the first 5 years of childhood in Bangladesh, *Pediatrics*, 134(4), e1001–e1008.

Hamre, B.K., R.C. Pianta, A.J. Mashburn and J.T. Downer (2007): Building a science of classrooms: Application of the CLASS framework in over 4,000 US early childhood and elementary classrooms, *Foundation for Childhood Development*, 30(2008).

Hanushek, E.A. and S.G. Rivkin (2012): The distribution of teacher quality and implications for policy, *Annu. Rev. Econ.*, 4(1), 131–157.

Havnes, T. and M. Mogstad (2015): Is Universal Child Care Leveling the Playing Field?, *Journal of Public Economics*, 127, 100–114.

Heckman, J. (2010): Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy, *Journal of Economic Literature*, 48(2), 356–98.

Heckman, J., N. Hohmann, J. Smith and M. Khoo (2000): Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment*, *The Quarterly Journal of Economics*, 115(2), 651–694.

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), 2052–86.

Heckman, J.J. (2006): Skill Formation and the Economics of Investing in Disadvantaged Children, *Science*, 312(5782), 1900–1902.

Heckman, J.J. and D.V. Masterov (2007): The productivity argument for investing in young children, *Applied Economic Perspectives and Policy*, 29(3), 446–493.

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(1-2), 114–128.

Heckman, J.J., S. Urzua and E. Vytlacil (2006): Understanding instrumental variables in models with essential heterogeneity, *The Review of Economics and Statistics*, 88(3), 389–432.

Heidemeyer, H. (1994): Flucht und Zuwanderung aus der SBZ, *DDR: Flüchtlingspolitik der Bundesrepublik Deutschland bis zum Bau der Mauer (Dusseldorf, 1993)*.

Henderson, R.H. and T. Sundaresan (1982): Cluster Sampling to Assess Immunization Coverage: A Review of Experience with a Simplified Sampling Method., *Bulletin of the World Health Organization*, 60(2), 253.

Herbst, C.M. (2018): The impact of quality rating and improvement systems on families' child care choices and the supply of child care labor, *Labour Economics*, 54, 172–190.

Hochschild, A. and A. Machung (1989): The second shift, *New York, NY: Viking*.

Hochschild, A. and A. Machung (2012): *The second shift: Working families and the revolution at home*: Penguin.

Hughes, C. and R. Ensor (2006): Behavioural problems in 2-year-olds: links with individual differences in theory of mind, executive function and harsh parenting, *Journal of Child Psychology and Psychiatry*, 47(5), 488–497.

Hull, P. (2018): IsoLATEing: Identifying Counterfactual-Specific Treatment Effects with Cross-Stratum Comparisons, *Working Paper*.

Humphreys, J. (2015): Education Premiums in Cambodia: Dummy Variables Revisited and Recent Data, *Econ Journal Watch*, 12(3), 339–345.

Hwang, J. (2016): The second shift: assimilation in housework time among immigrants, *Review of Economics of the Household*, 14(4), 941–959.

Ichino, A., M. Fort and G. Zanella (2019): Cognitive and Non-Cognitive Costs of Daycare 0-2 for Children in Advantaged Families, *Journal of Political Economy*.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, P. Lagos, M. Norris, E. Ponarin, B. Puranen and et al. (2014): World Values Survey: All Rounds - Country-Pooled Datafile Version.

Jackson, C.K., J.E. Rockoff and D.O. Staiger (2014): Teacher effects and teacher-related policies, *Annu. Rev. Econ.*, 6(1), 801–825.

Jensen, R. (2010): The (Perceived) Returns to Education and the Demand for Schooling, *The Quarterly Journal of Economics*, 125(2), 515–548.

Jeong, J., D.C. McCoy and G. Fink (2017): Pathways between paternal and maternal education, caregivers' support for learning, and early child development in 44 low-and middle-income countries, *Early Childhood Research Quarterly*, 41, 136–148.

Jessen, J. (2021): Culture, Children and Couple Gender Inequality, *mimeo*.

Johnson, R.C. and C.K. Jackson (2018): Reducing Inequality Through Dynamic Complementarity: Evidence from Head Start and Public School Spending, NBER Working Paper Series No. 23489, National Bureau of Economic Research.

Jones, P.C., L.L. Pendergast, B.A. Schaefer, M. Rasheed, E. Svensen, R. Scharf, R. Shrestha, A. Maphula, R. Roshan, Z. Rasmussen et al. (2017): Measuring home environments across cultures: Invariance of the HOME scale across eight international sites from the MAL-ED study, *Journal of School Psychology*, 64, 109–127.

Jung, H. and A. Hasan (2016): The impact of early childhood education on early achievement gaps in Indonesia, *Journal of Development Effectiveness*, 8(2), 216–233.

Kaiser, H.F. (1974): An index of factorial simplicity, *Psychometrika*, 39(1), 31–36.

Katz, L.F. and K.M. Murphy (1992): Changes in Relative Wages, 1963-1987: Supply and Demand Factors, *The Quarterly Journal of Economics*, 107(1), 35–78.

Kautz, T., J.J. Heckman, R. Diris, B. ter Weel, L. Borghans et al. (2014): Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success, *IZA Discussion Papers* (8696).

Kerr, S.P., T. Pekkarinen, M. Sarvimäki and R. Uusitalo (2020): Post-secondary education and information on labor market prospects: A randomized field experiment, *Labour Economics*, 66, 101888.

Kirkeboen, L.J., E. Leuven and M. Mogstad (2016): Field of Study, Earnings, and Self-Selection*, *The Quarterly Journal of Economics*, 131(3), 1057–1111.

Kline, P. and C.R. Walters (2016): Evaluating Public Programs with Close Substitutes: The Case of Head Start, *The Quarterly Journal of Economics*, 131(4), 1795–1848.

Knowles, J.A. (2013): Why are married men working so much? An aggregate analysis of intra-household bargaining and labour supply, *Review of Economic Studies*, 80(3), 1055–1085.

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

Krueger, A.B. and J.S. Pischke (1995): A comparative analysis of East and West German labor markets: before and after unification, in: *Differences and Changes in Wage Structures*: University of Chicago Press, pp. 405–446.

Kuklina, E.V., U. Ramakrishnan, A.D. Stein, H.H. Barnhart and R. Martorell (2006): Early childhood growth and development in rural Guatemala, *Early Human Development*, 82(7), 425–433.

Lavecchia, A.M., H. Liu and P. Oreopoulos (2016): Behavioral economics of education: Progress and possibilities, in: *Handbook of the Economics of Education*: Elsevier, Vol. 5, pp. 1–74.

Leroy, J.L. and E.A. Frongillo (2019): Perspective: what does stunting really mean? A critical review of the evidence, *Advances in Nutrition*, 10(2), 196–204.

Lindley, J. and S. Machin (2016): The rising postgraduate wage premium, *Economica*, 83(330), 281–306.

Lippa, R.A. (2005): Subdomains of gender-related occupational interests: Do they form a cohesive bipolar M-F dimension?, *Journal of Personality*, 73(3), 693–730.

Lippmann, Q., A. Georgieff and C. Senik (2020): Undoing gender with institutions: Lessons from the German division and reunification, *The Economic Journal*, 130(629), 1445–1470.

Lugo-Gil, J. and C.S. Tamis-LeMonda (2008): Family resources and parenting quality: Links to children's cognitive development across the first 3 years, *Child Development*, 79(4), 1065–1085.

Lundberg, S. and R.A. Pollak (1993): Separate spheres bargaining and the marriage market, *Journal of Political Economy*, 101(6), 988–1010.

Lundberg, S. and R.A. Pollak (1996): Bargaining and distribution in marriage, *Journal of Economic Perspectives*, 10(4), 139–158.

Maier, L. (2014): Methodik und Durchführung der Zeitverwendungserhebung 2012/2013, *Wirtschaft und Statistik*, 11(2014), 672–679.

Martinez, S., S. Naudeau and V. Pereira (2017a): The promise of preschool in Africa: A randomized impact evaluation of early childhood development in rural Mozambique, *Washington, DC: The World Bank*.

Martinez, S., S. Naudeau and V.A. Pereira (2017b): Preschool and Child Development Under Extreme Poverty: Evidence from a Randomized Experiment in Rural Mozambique, *World Bank Policy Research Working Paper* (8290).

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.

McLoyd, V.C. (1998): Socioeconomic disadvantage and child development., *American Psychologist*, 53(2), 185.

Meager, R. (2018): Understanding the Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of 7 Randomised Experiments, *American Economic Journal: Applied Economics (forthcoming)* (1506.06669).

Mejia, A., R. Calam and M.R. Sanders (2012): A review of parenting programs in developing countries: opportunities and challenges for preventing emotional and behavioral difficulties in children, *Clinical Child and Family Psychology Review*, 15(2), 163–175.

Miller, A.C., M.B. Murray, D.R. Thomson and M.C. Arbour (2016): How consistent are associations between stunting and child development? Evidence from a meta-analysis of associations between stunting and multidimensional child development in fifteen low-and middle-income countries, *Public Health Nutrition*, 19(8), 1339–1347.

Mincer, J. (1958): Investment in human capital and personal income distribution, *Journal of Political Economy*, 66(4), 281–302.

Mocan, N. (2007): Can consumers detect lemons? An empirical analysis of information asymmetry in the market for child care, *Journal of Population Economics*, 20(4), 743–780.

MoEYS (2014): Education Strategic Plan 2014-2018, `http://www.moeys.gov.kh/en/policies-and-strategies/559.html`, last accessed: 20-04-2021.

MoEYS (2017): The Education, Youth and Sport Performance in the Academic Year 2015-2016 and Goals for the Academic Year 2016-2017, `https://moeys.gov.kh/index.php/en/education-congress-2018/reports.html#.YH7cqKkza0d`, last accessed: 20-04-2021.

MOEYS (2019a): Education Strategic Plan 2019-2023, `https://www.moeys.gov.kh/index.php/en/policies-and-strategies/3206.html`, last accessed: 20-04-2021.

MOEYS (2019b): Public Education Statistics & Indicators, `https://www.moeys.gov.kh/index.php/en/emis/3069.html`, last accessed: 20-04-2021.

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

Naudeau, S., S. Martinez, P. Premand and D. Filmer (2011): Cognitive development among young children in low-income countries, *No small matter: The impact of poverty, shocks, and human capital investments in early childhood development*, 9–50.

Nguyen, P.H., D. Headey, E.A. Frongillo, L.M. Tran, R. Rawat, M.T. Ruel and P. Menon (2017): Changes in underlying determinants explain rapid increases in child linear growth in Alive & Thrive study areas between 2010 and 2014 in Bangladesh and Vietnam, *The Journal of Nutrition*, 147(3), 462–469.

Nguyen, P.H., K.K. Saha, D. Ali, P. Menon, S. Manohar, L.T. Mai, R. Rawat and M.T. Ruel (2014): Maternal mental health is associated with child undernutrition and illness in Bangladesh, Vietnam and Ethiopia, *Public Health Nutrition*, 17(6), 1318–1327.

Nguyen, T. (2008): Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar, *Unpublished manuscript*, 6.

Nybom, M. (2017): The Distribution of Lifetime Earnings Returns to College, *Journal of Labor Economics*, 35(4), 000–000.

OECD (2020): Education at a Glance 2020: OECD Indicators,

Olivetti, C. and B. Petrongolo (2016): The evolution of gender gaps in industrialized countries, *Annual Review of Economics*, 8, 405–434.

Olivetti, C. and B. Petrongolo (2017): The economic consequences of family policies: lessons from a century of legislation in high-income countries, *Journal of Economic Perspectives*, 31(1), 205–30.

Oreopoulos, P. and R. Dunn (2013): Information and college access: Evidence from a randomized field experiment, *The Scandinavian Journal of Economics*, 115(1), 3–26.

Oreopoulos, P. and R. Ford (2016): Keeping College Options Open: A Field Experiment to Help All High School Seniors Through the College Application Process, NBER Working Paper No. 22320, NBER, Cambridge MA.

Ozler, B., L.C. Fernald, P. Kariger, C. McConnell, M. Neuman and E. Fraga (2018): Combining pre-school teacher training with parenting education: A cluster-randomized controlled trial, *Journal of Development Economics*, 133, 448 – 467.

Padilla, C.M.L. (2019). *Moving Beyond the Average Effect: Quantifying and Exploring Variation in Head Start Treatment Effects on Parenting Behavior*. Ph. D. thesis, Georgetown University.

Patrinos, H.A. and G. Psacharopoulos (2020): Returns to education in developing countries, in: *The Economics of Education*: Elsevier, pp. 53–64.

Paxson, C. and N. Schady (2007): Cognitive development among young children in Ecuador the roles of wealth, health, and parenting, *Journal of Human Resources*, 42(1), 49–84.

Perlman, M., G.L. Zellman and V.N. Le (2004): Examining the psychometric properties of the early childhood environment rating scale-revised (ECERS-R), *Early Childhood Research Quarterly*, 19(3), 398–412.

Perumal, N., D.G. Bassani and D.E. Roth (2018): Use and misuse of stunting as a measure of child health, *The Journal of Nutrition*, 148(3), 311–315.

Peter, F., C.K. Spiess and V. Zambre (2018): Informing Students about College: An Efficient Way to Decrease the Socio-Economic Gap in Enrollment – Evidence from a Randomized Field Experiment, DIW Discussion Paper No.1770, DIW Berlin.

Peter, F.H. and V. Zambre (2017): Intended college enrollment and educational inequality: Do students lack information?, *Economics of Education Review*, 60, 125–141.

Pianta, R.C., B.K. Hamre and S. Mintz (2012): *Classroom assessment scoring system: Secondary manual*: Teachstone.

Ponguta, L.A., C. Maldonado-Carreño, S.L. Kagan, H. Yoshikawa, A.M. Nieto, C.A. Aragón, A. Mateus, A.M. Rodríguez, A. Motta, L.R. Varela et al. (2019): Adaptation and Application of the Measuring Early Learning Quality and Outcomes (MELQO) Framework to Early Childhood Education Settings in Colombia, *Zeitschrift für Psychologie*.

Prado, E.L. and K.G. Dewey (2014): Nutrition and brain development in early life, *Nutrition Reviews*, 72(4), 267–284.

Priller, E. (1993): Zeitverwendung im Wandel: Analysen mit Zeitbudgetdaten der DDR-Statistik, in: *Amtliche Daten der DDR und der neuen Bundesländer: Informationsquelle für die Sozialwissenschaften*: Mannheim: ZUMA, pp. 46–64.

Psacharopoulos, G. (1985): Returns to education: a further international update and implications, *Journal of Human Resources*, 583–604.

Puma, M., S. Bell, R. Cook, C. Heid, P. Broene, F. Jenkins, A. Mashburn and J. Downer (2012): Third Grade Follow-Up to the Head Start Impact Study, , Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Raikes, A., N. Koziol, D. Davis and A. Burton (2020): Measuring quality of preprimary education in sub-Saharan Africa: Evaluation of the Measuring Early Learning Environments scale, *Early Childhood Research Quarterly*, 53, 571–585.

Raikes, A., N. Koziol, M. Janus, L. Platas, T. Weatherholt, A. Smeby and R. Sayre (2019): Examination of School Readiness Constructs in Tanzania: Psychometric Evaluation of the MELQO Scales, *Journal of Applied Developmental Psychology*, 62, 122–134.

Raikes, A., R. Sayre, D. Davis, K. Anderson, M. Hyson, E. Seminario and A. Burton (2019): The Measuring Early Learning Quality & Outcomes initiative: purpose, process and results, *Early Years*, 39(4), 360–375.

Raikes, A., H. Yoshikawa, P.R. Britto and I. Iruka (2017): Children, youth and developmental science in the 2015-2030 global sustainable development goals, *Social Policy Report*, 30(3), 1–23.

Rao, N., J. Sun, J.M. Wong, B. Weekes, P. Ip, S. Shaeffer, M. Young, M. Bray, E. Chen and D. Lee (2014): *Early Childhood Development and Cognitive Development in Developing Countries*: DFID, UK Government.

Raven, J. (2000): The Raven's progressive matrices: change and stability over culture and time, *Cognitive Psychology*, 41(1), 1–48.

Reuben, E., M. Wiswall and B. Zafar (2017): Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender, *The Economic Journal*, 127(604), 2153–2186.

Rivkin, S.G., E.A. Hanushek and J.F. Kain (2005): Teachers, schools, and academic achievement, *Econometrica*, 73(2), 417–458.

Rosenfeld, R.A., H. Trappe and J.C. Gornick (2004): Gender and work in Germany: Before and after reunification, *Annual Review of Sociology*, 30, 103–124.

Rubio-Codina, M., O. Attanasio, C. Meghir, N. Varela and S. Grantham-McGregor (2015): The socioeconomic gradient of child development: Cross-sectional evidence from children 6–42 months in Bogota, *Journal of Human Resources*, 50(2), 464–483.

Ruhm, C. and J. Waldfogel (2012): Long-term effects of early childhood care and education, *Nordic Economic Policy Review*, 1(1), 23–51.

Sanderson, E. and F. Windmeijer (2016): A weak instrument F-test in linear IV models with multiple endogenous variables, *Journal of Econometrics*, 190(2), 212–221.

Sayer, L.C. (2005): Gender, time and inequality: Trends in women's and men's paid work, unpaid work and free time, *Social Forces*, 84(1), 285–303.

Schady, N., J. Behrman, M.C. Araujo, R. Azuero, R. Bernal, D. Bravo, F. Lopez-Boo, K. Macours, D. Marshall, C. Paxson et al. (2015): Wealth gradients in early childhood cognitive development in five Latin American countries, *The Journal of Human Resources*, 50(2), 446.

Schmitz, S. and F. Weinhardt (2019): Immigration and the Evolution of Local Cultural Norms, *IZA DP No. 12509*.

Schultz, T.W. (1961): Investment in human capital, *The American Economic Review*, 51(1), 1–17.

Selin, H. (2013): *Parenting across cultures: Childrearing, motherhood and fatherhood in Non-Western Cultures*, Vol. 7: Springer.

Shonkoff, J., J. Richmond, P. Levitt, S. Bunge, J. Cameron, G. Duncan and C. Nelson III (2016): From Best Practices to Breakthrough Impacts a Science-Based Approach to Building a More Promising Future for Young Children and Families, *Cambirdge, MA: Harvard University, Center on the Developing Child*.

Shonkoff, J.P., D.A. Phillips, N.R. Council et al. (2000): Communicating and learning, in: *From neurons to neighborhoods: The science of early childhood development*: National Academies Press (US).

Shonkoff, J.P., J.M. Radner and N. Foote (2017): Expanding the evidence base to drive more productive early childhood investment, *The Lancet*, 389(10064), 14–16.

Sørensen, A. and H. Trappe (1995): The persistence of gender inequality in earnings in the German Democratic Republic, *American Sociological Review*, 398–406.

Spangenberg, H. and H. Quast (2016): *Bildungsentscheidungen und Umorientierungen im nachschulischen Verlauf: Dritte Befragung der Studienberechtigten 2010 viereinhalb Jahre nach Schulabschluss*: Deutsches Zentrum für Hochschul-und Wissenschaftsforschung (DZHW).

Spiess, C.K. (2017). *Early Childhood Education and Care Services and Child Development: Economic Perspectives for Universal Approaches*, pp. 1–14: American Cancer Society.

Sprengholz, M., A. Wieber and E. Holst (2020): Gender identity and wives' labor market outcomes in West and East Germany between 1983 and 2016, *Socio-Economic Review*.

Stahl, J.F., P.S. Schober and C.K. Spiess (2018): Parental socio-economic status and childcare quality: Early inequalities in educational opportunity?, *Early childhood research quarterly*, 44, 304–317.

Stallings, J. (1977): *Learning to look: A handbook on classroom observation and teaching models*: Wadsworth Publishing Company.

Stallings, J. and G. Mohlman (1988): Classroom observation techniques, *Educational research, methodology, and measurement: An international handbook*, 469–474.

Stallings, J.A., S.L. Knight and D. Markham (2014): Using the stallings observation system to investigate time on task in four countries.

Sudfeld, C.R., D.C. McCoy, G. Danaei, G. Fink, M. Ezzati, K.G. Andrews and W.W. Fawzi (2015): Linear growth and child development in low-and middle-income countries: a meta-analysis, *Pediatrics*, 135(5), e1266–e1275.

Surkan, P.J., C.E. Kennedy, K.M. Hurley and M.M. Black (2011): Maternal depression and early childhood growth in developing countries: systematic review and meta-analysis, *Bulletin of the World Health Organization*, 89, 607–615.

Trappe, H. (1996): Work and family in women's lives in the German Democratic Republic, *Work and Occupations*, 23(4), 354–377.

UNESCO, UNICEF, B. (2017): Overview MELQO: Measuring Early Learning Quality Outcomes.

UNICEF (2019): *A World Ready to Learn: Prioritizing Quality Early Childhood Education*: UNICEF, New York, April 2019.

Vargas, A.J. (2016): Assimilation effects beyond the labor market: time allocations of Mexican immigrants to the US, *Review of Economics of the Household*, 14(3), 625–668.

Vermeulen, F. (2002): Collective household models: principles and main results, *Journal of Economic Surveys*, 16(4), 533–564.

Wagner, G.G., J.R. Frick and J. Schupp (2007): The German Socio-Economic Panel Study (SOEP) - Scope, evolution, and enhancements, *Schmollers Jahrbuch*, 127(1), 139–169.

Waldfogel, J. (2015): The role of preschool in reducing inequality, *IZA World of Labor*.

Wales, P. (2013): Access all areas? The impact of fees and background on student demand for postgraduate higher education in the UK, SERC Discussion Paper No. 128, Spatial Economics Research Centre (SERC), London School of Economics.

Walker, S.P., S.M. Chang, M. Vera-Hernández and S. Grantham-McGregor (2011): Early Childhood Stimulation Benefits Adult Competence and Reduces Violent Behavior, *Pediatrics*, 127(5), 849–857.

Weisgram, E.S. and R.S. Bigler (2006): Girls and science careers: The role of altruistic values and attitudes about scientific tasks, *Journal of Applied Developmental Psychology*, 27(4), 326–348.

Weisgram, E.S., L.M. Dinella and M. Fulcher (2011): The role of masculinity/femininity, values, and occupational value affordances in shaping young men's and women's occupational choices, *Sex Roles*, 65(3-4), 243–258.

Wiswall, M. and B. Zafar (2015): How Do College Students Respond to Public Information about Earnings?, *Journal of Human Capital*, 9(2), 117–169.

Woerner, W., B. Fleitlich-Bilyk, R. Martinussen, J. Fletcher, G. Cucchiaro, P. Dalgalarrondo, M. Lui and R. Tannock (2004): The Strengths and Difficulties Questionnaire overseas: evaluations and applications of the SDQ beyond Europe, *European Child & Adolescent Psychiatry*, 13(2), ii47–ii54.

Wolf, S., J.L. Aber, J.R. Behrman and E. Tsinigo (2019): Experimental Impacts of the "Quality Preschool for Ghana" Interventions on Teacher Professional Well-Being, Classroom Quality, and Children's School Readiness, *Journal of Research on Educational Effectiveness*, 12(1), 10–37.

World Bank (2018): *World Development Report 2018: Learning to Realize Education's Promise*: The World Bank.

World Health Organization (2006): *WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development*: World Health Organization.

Yoshikawa, H., D. Leyva, C.E. Snow, E. Treviño, M. Barata, C. Weiland, C.J. Gomez, L. Moreno, A. Rolla, N. D'Sa et al. (2015): Experimental Impacts of a Teacher Professional Development Program in Chile on Preschool Classroom Quality and Child Outcomes., *Developmental Psychology*, 51(3), 309.

Young, A. (2019): Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results, *The Quarterly Journal of Economics*, 134(2), 557–598.

Zafar, B. (2013): College major choice and the gender gap, *Journal of Human Resources*, 48(3), 545–595.

Zambre, V. (2018): The Gender Gap in Wage Expectations: Do Young Women Trade off Higher Wages for Lower Wage Risk?, DIW Discussion Paper No.1742 No. 1742, DIW Berlin.

Zelazo, P.D. (2006): The Dimensional Change Card Sort (DCCS): A Method of Assessing Executive Function in Children, *Nature Protocols*, 1(1), 297.

Zoch, G. (2021): Thirty Years after the Fall of the Berlin Wall—Do East and West Germans Still Differ in Their Attitudes to Female Employment and the Division of Housework?, *European Sociological Review*, 1–20.

Zonji, S. (2018): A network for early childhood development, *Annals of the New York Academy of Sciences*, 1419(1), 20–22.