



Later Impacts of Early Childhood Interventions: A Systematic Review

IEG Working Paper 2015/3

Jeffery C. Tanner, Tara Candland, and Whitney S. Odden

© 2015
Independent Evaluation Group, The World Bank Group
1818 H St., NW
Washington, DC 20433

IEG: Improving World Bank Group Development Results Through Excellence in Evaluation

The Independent Evaluation Group is an independent unit within the World Bank Group; it reports directly to the Bank's Board of Executive Directors. IEG assesses what works, and what does not; how a borrower plans to run and maintain a project; and the lasting contribution of the Bank to a country's overall development. The goals of evaluation are to learn from experience, to provide an objective basis for assessing the results of the Bank's work, and to provide accountability in the achievement of its objectives. It also improves Bank work by identifying and disseminating the lessons learned from experience and by framing recommendations drawn from evaluation findings.

IEG Working Papers are an informal series to disseminate the findings of work in progress to encourage the exchange of ideas about development effectiveness through evaluation.

The findings, interpretations, and conclusions expressed here are those of the author(s) and do not necessarily reflect the views of the Board of Executive Directors of the World Bank or the governments they represent, or IEG management.

The World Bank cannot guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply on the part of the World Bank any judgment of the legal status of any territory or the endorsement or acceptance of such boundaries.

ISBN-10: 1-60244-261-4
ISBN-13: 978-1-60244-261-0

Contact: IEG Communication, Learning and Strategies (IEGCS)
e-mail: ieg@worldbank.org
Telephone: 202-458-4497
Facsimile: 202-522-3125
<http://www.worldbank.org/ieg>

Contents

| | |
|---|-----------|
| ABBREVIATIONS | VI |
| GLOSSARY | VII |
| ACKNOWLEDGMENTS | IX |
| OVERVIEW | XI |
| Highlights | xi |
| Methodology and Scope | xii |
| Findings..... | xiii |
| Physical Development | xiii |
| Cognitive Development | xiv |
| Language Development | xvi |
| Socioemotional Development..... | xvii |
| Schooling Outcomes | xviii |
| Employment and Labor Market Outcomes..... | xix |
| Heterogeneous Effects..... | xix |
| Time Effects | xx |
| Knowledge Gaps..... | xxi |
| Challenges | xxii |
| Implications | xxiii |
| INTRODUCTION: REVIEW QUESTIONS AND STRATEGY | 1 |
| Search Strategy for Identification of Relevant Studies..... | 4 |
| Criteria for Inclusion and Exclusion of Studies..... | 6 |
| PART I: EFFECTS BY OUTCOME | 13 |
| 1. PHYSICAL DEVELOPMENT | 15 |
| Weight | 22 |
| Height..... | 25 |
| Mid-Upper-Arm Circumference and Head Circumference | 29 |
| Fine Motor Skills..... | 29 |
| Physical Development Summary | 29 |
| 2. COGNITIVE DEVELOPMENT..... | 32 |
| Explanation of Cognitive Tests..... | 32 |
| Evidence from Developed Countries..... | 33 |
| Evidence from Developing Countries..... | 34 |
| Importance of Measuring Cognition | 35 |
| General Cognition | 41 |
| Nonverbal Cognition..... | 45 |
| Executive Function..... | 46 |
| Cognition Summary..... | 47 |
| 3. LANGUAGE DEVELOPMENT | 49 |
| Verbal Ability | 53 |
| Reading and Literacy | 57 |
| Vocabulary | 60 |
| Language Summary..... | 62 |
| 4. SOCIOEMOTIONAL DEVELOPMENT | 63 |
| Externalizing Behavior | 71 |

CONTENTS

| | |
|---|------------|
| Internalizing Behavior..... | 72 |
| Strengths and Difficulties Questionnaire..... | 72 |
| Socioemotional Summary..... | 73 |
| 5. SCHOOLING OUTCOMES..... | 75 |
| On-Time Primary School Enrollment..... | 79 |
| Years of Schooling Completed..... | 81 |
| Post-Primary Attendance..... | 86 |
| School Performance..... | 86 |
| Schooling Summary..... | 88 |
| 6. EMPLOYMENT AND LABOR MARKET OUTCOMES..... | 89 |
| Employment and Labor Market Summary..... | 91 |
| PART II: FACTORS INFLUENCING OUTCOMES..... | 93 |
| 7. HETEROGENEOUS EFFECTS—GIRLS, BOYS, AND THE POOR..... | 94 |
| Effects by Gender..... | 95 |
| Effects by Socioeconomic Status..... | 97 |
| Heterogeneous Effects Summary..... | 98 |
| 8. TIME EFFECTS—LONGER, LATER, AND LASTING?..... | 100 |
| Temporal Trajectories..... | 100 |
| From Early Childhood to Later Outcomes..... | 101 |
| Beyond Early Childhood..... | 102 |
| Temporal Trajectories Summary..... | 105 |
| Age at Intervention: The First 1,000 Days..... | 106 |
| Length of Exposure..... | 108 |
| Schooling Outcomes..... | 109 |
| Cognitive and Language Development..... | 110 |
| Socioemotional Development..... | 111 |
| Physical Development..... | 111 |
| Summarizing Time Effects..... | 112 |
| PART III: IMPROVING STUDY QUALITY AND COVERAGE..... | 115 |
| 9. IMPROVING STUDY QUALITY..... | 116 |
| Confounding Characteristics..... | 116 |
| Attrition..... | 118 |
| Challenges in Design Allowing for Follow-Up..... | 121 |
| External Validity..... | 122 |
| 10. KNOWLEDGE GAPS TO FILL..... | 125 |
| Gaps in Outcomes by Intervention Type..... | 125 |
| Gaps by Evaluation Components..... | 129 |
| Gaps by Region..... | 130 |
| Addressing the Gaps..... | 133 |
| REFERENCES..... | 135 |
| List of Numbered Studies..... | 135 |
| Other Citations..... | 140 |
| APPENDIX A. DESCRIPTIONS OF INTERVENTIONS FROM INCLUDED STUDIES..... | 151 |
| APPENDIX B. COMPLETE LIST OF OUTCOMES APPEARING IN INCLUDED STUDIES..... | 166 |

| | |
|--|-----|
| APPENDIX C. SEARCH STRATEGY DETAILS—IDENTIFYING RELEVANT STUDIES | 182 |
| APPENDIX D. CODING PROTOCOL AND INSTRUMENTS..... | 190 |
| APPENDIX E. ANTHOLOGY OF IDENTIFIED ECD SYSTEMATIC REVIEWS | 194 |
| APPENDIX F. LIST OF STUDIES GIVEN AN A RATING | 197 |
| APPENDIX G. INTERVENTION-OUTCOME GAP MAP | 203 |

INCLUDED BOXES

| | |
|---|-----|
| Introduction Box 1. Analyzed Outcomes: The Rule of Three | 6 |
| Introduction Box 2. Challenges and Cautions for Systematic Reviews..... | 10 |
| Box 1.1 Vignette on Physical Development: Breastfeeding Promotion in Belarus has few Lasting Physical Benefits...20 | |
| Box 2.1 Vignette on Cognitive Development: The Psychosocial Component of the Jamaica Supplementation and Stimulation ECD Program Produces Lasting Cognitive Benefits..... | 39 |
| Box 3.1 Vignette on Language Development: Fostering and Stimulation in Romania Improved Language and Other Outcomes..... | 53 |
| Box 4.1 Vignette on Socioemotional Development: Early Education and Health Enrichment in Mauritius Results in Improved Socioemotional Outcomes through early Adulthood..... | 68 |
| Box 5.1 Vignette on Schooling Outcomes: Mexico's Progresa Conditional Cash Transfer Program Improves Post-early Childhood Schooling Outcomes..... | 79 |
| Box 5.2 What Helps Numeracy and Maths?..... | 87 |
| Box 9.1 Identifying Challenges of Follow-Up Evaluations of ECD Interventions: Confounding Characteristics in the Instituto de Nutrición de Centroamérica y Panamá (INCAP) Program in Guatemala..... | 117 |
| Box 9.2 Identifying Challenges of Long-Term Evaluations: Attrition in the Turkish Early Enrichment Project (TEEP).. | 120 |

Abbreviations

| | |
|--------|---|
| BMI | body mass index |
| BMIZ | BMI for age z-score |
| CCT | conditional cash transfer |
| CPC | Child Parent Center |
| CSG | Child Support Grant |
| DQ | development quotient |
| ECCE | Early Childhood Care and Education |
| ECD | early childhood development |
| HAZ | height-for-age z-score |
| HC | Hogares Comunitarios |
| IBRD | International Bank for Reconstruction and Development |
| IDA | International Development Assistance |
| IEG | Independent Evaluation Group |
| INCAP | Instituto de Nutrición de Centroamérica y Panamá |
| IQ | intelligence quotient |
| MUAC | mid-upper-arm circumference |
| SIEF | Strategic Impact Evaluation Fund |
| PROBIT | Promotion of Breastfeeding Intervention Trial |
| RCT | randomized controlled trial |
| RPS | Red de Protección Social |
| SDQ | Strengths and Difficulties Questionnaire |
| SMP | Safe Motherhood Program |
| TEEP | Turkish Early Enrichment Project |
| TVIP | Test de Vocabulario en Imagenes Peabody |
| UCT | unconditional cash transfer |
| UNICEF | United Nations Children's Fund |
| UNIT | Universal Nonverbal Intelligence Test |
| WAZ | weight-for-age z-score |
| WAIS | Wechsler Adult Intelligence Scale |
| WASI | Wechsler Abbreviated Scale of Intelligence |
| WISC | Wechsler Intelligence Scale for Children |
| WPPSI | Wechsler Preschool and Primary Scale of Intelligence |
| WHZ | weight-for-height z-score |
| WHO | World Health Organization |

Glossary

| | |
|----------------------------|---|
| Estimate | The empirical result of the application of an impact evaluation (IE) methodology to a set or subset of data for an intervention. Multiple estimates are possible for the same intervention and study if multiple IE methods are used. |
| Impact evaluation | A methodology of empirical analysis allowing causal inference through the use of a counterfactual. For the purposes of counting impact evaluations in this review, the number of impact evaluations within a study is the number of unique interventions within that study that reports an estimate for one of the included outcomes. Multiple arms are counted separately. |
| Intervention | The most disaggregated combination of policy or project components for which there is a unique impact evaluation estimate; an arm of a randomized control trial or pieces of an intervention whose effect can be separately estimated in a quasi-experimental IE. |
| Intervention family | The broadest categorization of interventions used for this review (see figure 1.1). |
| Intervention type | The more specific categorization of interventions used for this review (see figure 1.2). |
| Outcome | The construct measured by an estimate. |
| Project | The full bundle of interventions carried out for a population over a period of time. |
| Study | An article, working paper, or other publication that has at least one unique estimate for use in this systematic review. For counting purposes, a publication with several effectiveness for multiple arms of a project are still counted as a single study. |

Acknowledgments

This study was prepared by a team led by Jeffery C. Tanner. The core writing team consisted of Tara L. Candland, Whitney S. Odden, and Jeffery C. Tanner. The team benefited from excellent research assistance provided by Ryotaro Hayashi. In addition to these individuals, the search and coding team also included Tomoko Kato, Elizabeth Koshe, Yunsun Li, Karol Acon Monge, and Maya Sherpa. Valuable topical expertise was provided by Patricia Kariger. Editorial support was given by Barbara Rice and Bruce Ross-Larson. Yezena Yimer offered administrative support.

The team thanks Harold Alderman, Lucy Basset, Susan Caceres, Leslie Elder, Julieta Trias, and Quentin Wodon for insightful discussions and comments at various stages of the review. Members of the Independent Evaluation Group management team also provided useful comments, as did participants of presentations of the work from the Japan International Cooperation Agency, the University of Tokyo, and the World Bank. The team is particularly grateful to the government of Japan for partially funding this work through a Policy and Human Resources Development Grant.

Finally, the team wishes to thank and acknowledge the authors and implementers of the evaluations contained herein, together with their funders, for their vision and dedication in breaking new ground to aid in the understanding of the later-life effects of early childhood interventions.

The work was conducted under the general guidance of Mark Sundberg (manager), Emmanuel Jimenez (former director over the course of the majority of the work), and Nicholas York (acting director). The team is grateful for comments proffered by its peer reviewers, Lia Fernald and Sophie Naudeau.

Overview

Highlights

The economic rationale for investing in young children goes beyond improving quality of life during early childhood; it hinges on the belief that the benefits of these investments persist into school age and beyond. This report is the first systematic review devoted exclusively to investigating this theory.

By identifying and analyzing all 55 studies that provide reliable causal estimates, the report provides the most complete, credible evidence to date on the post-early childhood effects of early childhood interventions. It serves three important functions. First, it provides analysis on early childhood interventions whose sustained effects have been evaluated across six areas of human development. Second, it examines how effects change – both within a population (yielding shared prosperity implications) and over time. Finally, the review aims to improve the quality and coverage of ECD knowledge by enumerating commonly observed evaluation challenges and identifying research gaps on the question of benefits beyond early childhood.

Four important findings emerge from this effort:

- ❖ Early childhood interventions can, but do not always, lead to benefits later in life in the areas of cognition, language, socioemotional health, education, and the labor market. Evaluated interventions have not demonstrated consistent lasting advantages for physical development, although these outcomes are less salient to adult welfare.
- ❖ Gender-neutrality dominates outcomes generally, but schooling does tend to improve for girls, the poor, and those who are in quality preschool and supplemental feeding programs for longer.
- ❖ Nutrition interventions may need to be in place throughout and beyond the first 1,000 days in order to leverage the window of opportunity from conception to age 2 and achieve sustained effects beyond early childhood.
- ❖ Sizeable knowledge gaps persist but can be closed with careful planning and design.

Early childhood development holds considerable promise for making progress on the World Bank’s dual objectives of reducing poverty and increasing shared prosperity while encouraging economic growth. More investment and evaluation are required to enable interventions, and the children they serve, to realize their potential.

In developed countries, well-documented evidence shows that interventions aimed at improving early childhood development (ECD) can play

a major role in shaping the arc of young children’s lives and livelihoods. This evidence prompts many in the international development

community – including the World Bank – to focus attention on the years before children enroll in primary school or reach the age of six as the first step in a sequenced strategy to build the skills needed for productivity and economic growth.

In developing countries, the ability of interventions to improve outcomes beyond early childhood is less well studied. Impact evaluations in middle- and low-income countries are scarce; it may be problematic to assume findings from high-income countries apply to middle- and low-income economies.

In an effort to bridge the evidence gap, the Independent Evaluation Group (IEG) prepared a systematic review that gathers and analyzes the available impact evaluation evidence in developing countries from 1990 to 2015 on whether early childhood interventions shape future outcomes. Its purpose is not to supplant existing evidence but rather to help practitioners understand how evidence from impact evaluations supports or challenges beliefs about interventions and can be used to inform development policy.

This review aims to answer two questions:

- What is the evidence of attributable effects on outcomes in primary school and beyond from early childhood interventions in low- and middle-income countries?

- How do the post-early childhood effects of early childhood interventions vary by socio-economic status, gender, age at intervention (particularly the first 1,000 days from conception to age two), and age at evaluation?

Methodology and Scope

Most systematic reviews are intervention based, meaning they track the outcomes from a narrow set of interventions. This report, on the other hand, reviews all interventions from developing countries that occur during early childhood for which impact evaluation estimates exist for effects observed at primary school age and older. From an initial search of thousands of studies, the search process – which included database searches, hand searches, and snowballing – identified more than 500 ECD-related impact evaluations written from 1990 to 2015. After excluding clinical trials as well as outcome estimates for which at least part of the sample was still in the early childhood period, 116 studies remained. The team then analyzed the risk of bias for these 116 studies, including conducting a rigorous check of the credibility of the causal estimates, identifying assumptions, and assessing construct validity and representativeness. Studies were given a quality rating of A, AA, or AAA. The 55 studies with ratings of AA or AAA were included in the review.

Although this review contains all of the causally robust evaluations within its inclusion criteria, the evidence base for any given combination of interventions and outcomes may still be thin. Furthermore, there are many reasons why a study may yield a null result besides the intervention having no impact, including challenges of statistical power, contamination, attrition, uptake, and implementation. Therefore, IEG's aim is to elucidate what is known about the long-term effects of early childhood interventions and identify the remaining research gaps.

Findings

This review covers six areas—or domains—of human development: physical development, cognitive development, language development, socioemotional development, schooling outcomes, and employment and labor market outcomes. These domains are commonly included in evaluations of early interventions targeting poor children because they are negatively affected by early poverty, can benefit from early intervention, and are important for overall well-being or adult productivity. Some outcomes can be measured repeatedly starting from early childhood (for example, height and weight) while others are measurable only later in life (for example, cognition, schooling, and employment and labor market status).

PHYSICAL DEVELOPMENT

Some improvement is seen in height but no effect is found on weight.

Taken as a whole, physical development benefits appear difficult to sustain past early childhood. Cohort studies find strong correlations between early anthropometrics and later cognition, schooling and labor market outcomes, and the broader literature contains many examples of impact evaluations that demonstrate the ability of early health and nutrition interventions to improve physical development among children under age 6 in the near term. However, those effects appear to dissipate quickly beyond early childhood.

Despite the review of seven different intervention types in five geographical Regions of the World Bank over a range of ages from six to 17 years old, none had a lasting effect on weight.

The ability of early childhood interventions to change height may also be limited. Only 2 of 10 nutritional supplementation interventions improved height, and both were only marginally significant: Providing iron and zinc supplements to infants in Thailand improved height at six to eight years old; giving pregnant Nepalese mothers folic acid, iron, and zinc slightly improved children's height at age nine. Yet the eight nutrition interventions affected height either negatively or not at all.

OVERVIEW

The cash transfer from Mexico's Progresa, and the family planning and maternal and child health program in Matlab, Bangladesh, improved beneficiaries' height-for-age, and female beneficiaries of the South African Child Support Grant were significantly taller. It is difficult to draw firm conclusions from these results because other evaluations of cash transfers and early stimulation programs found no effect.

The third physical outcome – fine motor skills – has a rather thin evidence base but was included for its intrinsic value as a measure of school readiness. Fine motor skills improved in a Mozambican preschool program, but the Bucharest Early Intervention Program had no lasting effect on fine motor skills.

Even so, physical outcomes such as stunting and wasting are important not for their intrinsic value but as functional correlates of impeded cognitive development, school achievement, and future economic activity. Therefore, the finding that physical development benefits are rarely sustained past early childhood is less alarming in light of the many sustained benefits in the functional outcomes more closely tied to lifelong well-being.

COGNITIVE DEVELOPMENT

Stimulation interventions most consistently impacted outcomes over the spectrum of cognitive outcome measures; "general cognition" was improved by a wide range of interventions.

Cognitive ability reflects an individual's problem-solving and analytical skills, memory functions, general knowledge

and ability to apply logic, and reactions to new situations. Cognitive outcomes are of particular interest in ECD research because of their influence on an individual's future productivity in areas such as schooling and the labor market.

The Nutrition Enigma

Small-scale and near-term studies as well as others from high-income countries have shown that nutrition interventions can lead to significant improvements in child development, including better morbidity, mortality, and cognitive outcomes.

Longitudinal studies, including those in developing countries, indicate a strong correlation between early nutritional status and later-life outcomes.

Yet these later-life effects in developing countries are less well-established. Across human development domains, nutrition interventions had little impact on post-early childhood outcomes.

The way forward for nutrition interventions may lie in both timing and duration. The only nutrition project that demonstrated sustained effects was also the only one that had a sustained intervention throughout (and beyond) the critical period of the first 1,000 days of a child's life from conception through age two.

Early sustained nutrition promises later sustained results.

Quantifying outcomes in cognition can be difficult because of the lack of a universal, standardized measurement of cognitive ability. Therefore, the reviewed studies report outcomes from a range of tests, from brief screening assessments to comprehensive full-scale intelligence tests. Given that it is inaccurate to compare results across

different types of cognitive tests, the review groups results from similar assessment tools to compare outcomes across intervention types and contexts.

Improvements in both full-scale and abbreviated measures of general cognition were caused by four different intervention types and six separate programs: breastfeeding promotion in Belarus, stimulation-related interventions in Jamaica and Romania, CCTs in Mexico and Nicaragua, and a deworming program in Kenya.

The breastfeeding promotion program in Belarus improved IQ outcomes at age 6.5 years old, though the effects were only marginally significant. The stimulation component of the Jamaica program improved cognition scores for low birthweight children at age 6 and stunted children at the ages of 11–12, 17–18, and 22 years. In Romania, children who received greater stimulation by being randomly assigned to foster care had marginally higher full-scale IQ scores at age 8 years than children who had remained in institutional care. The cash component of the Progresá CCT in Mexico produced a marginally significant effect on general cognition between the ages of 8 and 10, and Nicaragua's conditional cash transfer improved cognition in 10-year-old boys. Siblings of children who had participated in a deworming program in Kenya also had measurable improvements between the ages of 8 and 15 years old.

While several interventions proved effective in improving nonverbal cognition, the evidence base for each intervention type is too thin to be more than suggestive. Only a single nutritional intervention, Guatemala's Instituto de Nutrición de Centroamérica y Panamá (INCAP) program, improved nonverbal cognition in the post-early childhood period. Stunted children who participated in the stimulation component of the Jamaica program scored higher in nonverbal cognition at ages 11–12 and 17–18. Only one of two social protection interventions, the CCT in Nicaragua, improved nonverbal outcomes. The deworming program in Kenya also improved nonverbal cognition among siblings of children who participated in the intervention.

There is similarly little evidence around executive function, a measure of cognition that reflects how an individual responds to new or challenging situations. While stimulation for low birthweight babies in Jamaica improved short-term memory at age six years, related measures did not show improvement at age 11–12 (processing speed) or 17–18 years old (working memory). The Romanian foster care program did not result in noticeable improvements in executive function across a range of measurements assessed at eight years of age. The Nicaragua CCT program improved executive function in 10-year-old boys but not girls.

OVERVIEW

Most nutrition interventions did not yield lasting cognitive benefits. The one (of six) that demonstrated strong impacts was also the only one to have supplied nutrition in a sustained fashion through (and in this case beyond) the first thousand days from conception to age two. On the other hand, both stimulation programs, Jamaica and Bucharest, improved four different cognitive outcomes, as did health access in Indonesia and deworming in Kenya.

LANGUAGE DEVELOPMENT

Early childhood interventions can improve language outcomes, although the evidence is mixed across intervention types.

During the second year of life (12–24 months), children experience a vocabulary explosion. As they enter the preschool years, vocabulary, spoken grammar, and sentence structure become more sophisticated, and children develop the ability to identify letters and, later, words. These skills are important for enabling them to read and do well in school.

Language outcomes were measured in three subdomains: verbal ability, reading and literacy, and vocabulary. The existing evidence suggests that stimulation, nutrition, and social protection programs can improve language. In addition, three nontraditional ECD interventions – deworming, sanitation, and governance – all improved language outcomes, although only one study was found for each type.

Two of the three stimulation programs included in this review improved long-term language outcomes. Verbal abilities seemed to be particularly sensitive to early stimulation programs: children who participated in the Romanian foster care program had significantly higher verbal abilities at age eight years, and gains were observed among participants of the Jamaican stimulation program at ages 11–12, 17–18, and 22 years. The stimulation component of the Jamaica program also produced lasting gains in reading and vocabulary. A second stimulation program in Mozambique, however, did not find measurable effects on beneficiaries' vocabulary.

Only two of the five nutritional interventions that measured long-term language outcomes found significant effects, although a single nutritional intervention – breastfeeding promotion in Belarus – improved outcomes for 6.5 year old children in all three language subdomains. A supplementation program in Guatemala also improved reading and literacy. However, the supplementation-only arm of the Jamaica intervention did not have a measurable effect on reading, and a nutritional intervention in The Gambia did not significantly improve children's vocabulary. These results suggest that supplementation alone might not be enough to produce sustained linguistic effects, despite solid rationale for expecting that various types of nutritional interventions aimed at

children who are at risk for or suffering from deficiencies such as chronic malnutrition and iron deficiency could positively impact cognitive and language outcomes.

Social protection programs appear to have inconsistent effects on language outcomes. Although Mexico's Progresa CCT increased verbal abilities among beneficiaries and a CCT in Nicaragua produced measurable gains in vocabulary among boys, none of the three social protection interventions that measured reading and literacy produced significant improvements.

Non-traditional ECD interventions also proved effective in improving language outcomes, although the evidence is thin and mixed. In India, both a sanitation intervention and a governance campaign improved reading and literacy. Although the governance campaign did not have measurable effects on children's vocabulary, a deworming program in Kenya did significantly improve vocabulary among 15 year olds.

SOCIOEMOTIONAL DEVELOPMENT

Delayed improvement occurs in externalizing behavior, but little improvement is seen in internalizing behavior.

Social and emotional functioning involves the acquisition of the skills and knowledge required by a person to understand and manage emotions, set and achieve goals, empathize with

others, establish and maintain positive relationships, and make responsible decisions. It encompasses a broad range of internalizing and externalizing behaviors, which indicate how people view themselves and how they react to the world around them, respectively. These skills are important in learning to cope with difficulties and succeed in various endeavors.

Early childhood interventions appear to have a delayed effect on externalizing behavior; the review found no consistent effects on young children, but benefits from different early stimulation programs seem to show up more as children age into adolescence and beyond. No early intervention has yet been able to demonstrate later improvements in a child's ability to pay attention, notwithstanding the efforts of the early stimulation intervention in Jamaica and the micronutrients intervention in Thailand. However, the Environmental Enrichment Program in Mauritius and the early stimulation program in Jamaica did have a positive effect on some elements of externalizing behavior for teenagers including violent conduct.

The early stimulation program in Jamaica also had a sustained effect on internalizing behavior, while the nutrition arm of that intervention, foster care in Romania, and the early enrichment intervention in Mauritius did not. Those who benefited from early stimulation in Jamaica exhibited lower anxiety and depression and greater self-

OVERVIEW

esteem at 17 years old, a finding that persisted at age 20, although by that time it was largely driven by improvements among women. When participants reached the age of 22 years, the effect on anxiety had disappeared, but there was still a significant decrease in depression among beneficiaries.

The Strength and Difficulties Questionnaire, which is a brief screening test, is an additional measure of both internalized and externalized socioemotional behavior. Its results can change due to early childhood interventions. Mexico's conditional cash transfer program, an early stimulation program for low birthweight infants in Jamaica, and the early stimulation foster care program in Romania significantly reduced post-early childhood behavior problems as measured by the questionnaire.

SCHOOLING OUTCOMES

Early stimulation, preschool, and cash transfers appear to be most effective in improving educational outcomes.

Early childhood interventions could affect schooling through a number of possible pathways. For instance, improved cognitive development could result in increased scholastic achievement, while healthier children are better able to attend classes. Indeed, there is evidence that early childhood interventions can improve various educational outcomes.

Preschool programs and cash transfers appear effective in promoting on-time primary school enrollment. Of the four interventions evaluated, Mexico's Progresa and the Mozambican preschool program had a significant beneficial effect, while South Africa's Child Support Grant had no overall effect but did decrease the probability of delayed enrollment for girls and children whose mothers were less educated. A micronutrient supplementation program in Thailand, however, had no effect on on-time enrollment.

While the heterogeneity of interventions evaluated make it difficult to find a clear pattern in differences in the years of schooling completed, some general trends emerge to suggest that early stimulation and cash transfers (though not nutrition programs) may be effective in promoting more education. For example, early stimulation in Jamaica and preschool attendance in Uruguay increased schooling among participants. Similar results were seen among some beneficiaries of the Honduran, Mexican, and South African cash transfer. For nutrition, the INCAP supplementary feeding program in Guatemala had a large effect, but no effect was detectable from either the Jamaican supplementary feeding intervention or the maternal supplementation program in The Gambia. Remarkably, the largest improvement in schooling came from a clean water program in China.

Nutrition programs also did not have a detectable effect on school performance, but participants in early stimulation programs may be more likely to perform well and attend post-primary school. Only two programs were evaluated for their effect on post-primary attendance – *Hogares Comunitarios* in Colombia and the early stimulation intervention in Jamaica – both of which caused a large increase in the probability of attending. Furthermore, beneficiaries of the early stimulation in Jamaica, as well as the children who participated in a preschool program in Argentina and Chile’s Early Childhood Care and Education programs did significantly better on their subject matter and standardized achievement tests. Conversely, in three different nutrition interventions, participants’ test scores were not significantly different than their peers.

EMPLOYMENT AND LABOR MARKET OUTCOMES

Early stimulation can help.

The goal of many early childhood interventions is to improve human capital, and labor market outcomes offer an important measure of its fulfillment. Early childhood services devoted to enriching an individual’s environment by increasing inputs in education, health, and nutrition can determine the nature of these outcomes.

Indeed, there is evidence that early stimulation as well as good nutrition can positively influence participants’ subsequent labor market outcomes.

Psychosocial stimulation in Jamaica dramatically increased earnings, especially among women and people with full-time jobs. Interestingly, additional (though flawed) evidence of an early stimulation training provided to mothers through the Early Enrichment Program in Turkey did not cause its participants to be better employed but did delay the starting age of employment – commonly associated with lifetime earnings – likely because it also increased the probability of attending college.

Heterogeneous Effects

Later-life effects are generally gender-neutral, but girls and those from poor families or more educated parents do tend to have better schooling outcomes.

Evidence for heterogeneous effects – or differences in outcomes due to individual characteristics – is reported in fewer than half of the studies. Although evidence is thin for specific outcomes, a few noteworthy trends emerge for the broader outcome domains.

Based on the available evidence, the later-life effects of early childhood intervention appear to be mostly gender-neutral, especially in the physical and socioemotional domains and for nutrition and early learning or childcare interventions; in other words, there is usually no significant difference in the benefits accrued to girls versus boys. However, girls are much more

likely than not to benefit from interventions that affect schooling outcomes, and neither gender is likely to enjoy lasting physical benefits from interventions occurring in early childhood.

Conversely, for socioeconomic status, there are some groups that are significantly more likely to benefit than others. Interventions affecting physical outcomes appear to benefit the rich and the poor equally when they affect them at all, but the poor and those with better-educated parents benefit significantly more from interventions that improve schooling than do children from richer families or those with less-educated parents.

Time Effects

The persistence of effects over time varies by outcome domain. Interventions lasting at least the full first 1,000 days may be more effective, and additional exposure to some programs can be beneficial.

Three dimensions of time are evaluated for their effect on post-early childhood outcomes: temporal trajectories, age at exposure, and length of exposure. These elements are critical to consider when designing an intervention, but few studies examine these important elements, making it impossible to draw firm conclusions. Nevertheless, the preliminary findings drawn from the available evidence can help target future research.

In the first dimension—temporal trajectories—outcomes within a given intervention are traced across time to determine whether and how they change as a child ages. It appears that socioemotional benefits, particularly those that pertain to internalizing behavior, may fade over time, but cognitive benefits from an early stimulation program in Jamaica not only remained significant from 11 to 22 years old but actually increased in magnitude.

The evidence regarding the temporal trajectory of physical outcomes suggests that they tend to remain constant over time. Evaluations of the six different programs with IE results at multiple points in time found null effects for most estimates for height and weight—both in early childhood and later in life. In a few cases, however, initial anthropometric benefits during early childhood disappeared by the post-early childhood evaluations, and in one instance, six- to eight-year-old Nepali children whose mothers had received prenatal micronutrients were taller than their peers despite no detectable difference in length at birth.

To evaluate age at exposure, this review examines the first 1,000 days, a period that is believed to be critical for a child's development. Four impact evaluations specifically isolated the effect of treatment during this period compared to later periods, but because the interventions are so varied, it is difficult to draw firm conclusions. Children who began to receive South Africa's Child

Support Grant before they were two years old were not significantly taller or less stunted than those who began the program between two and five years old. In China, however, children who had access to clean water during the first 1,000 days completed more schooling than those who gained access later in life. Indonesian children enjoyed higher cognitive, math and schooling outcomes if the Safe Motherhood Program was in place over their first 1,000 days compared to children who benefitted after age two. Finally, a review of four evaluations of the Bucharest Early Intervention Program was inconclusive; age at placement influenced some cognitive development outcomes but not others.

Impact evaluations of six nutrition programs starting at various ages and lasting for various lengths of time – but always starting and ending during the first 1,000 days – demonstrated few later-life effects. However, a seventh program – providing supplemental feeding in Guatemala – demonstrated that continuous exposure from pregnancy through the first two to three years of age was more important than at three to six years of age and caused larger and significant results for highest grade completed, reading comprehension, and nonverbal cognition. Taken together, this suggests that effective interventions may need to not only start early but also continue throughout and perhaps beyond the first 1,000 days.

In the final dimension – length of exposure – much of the evidence comes from dose response estimates for cash transfer programs or from evaluations of preschool interventions. Given how important length of exposure is to determining the optimal timing of an intervention, very little evidence is available to guide policy makers on the effect of longer participation in any given intervention. What evidence does exist, however, highlights two important areas in which longer exposure times may be helpful in producing benefits. For preschool or childcare programs, it appears that longer exposure can lead to higher school enrollment rates, while additional involvement in a cash transfer program during the early childhood period could help reduce behavioral problems through adolescence.

Knowledge Gaps

More causal evidence is needed to fill gaps and corroborate findings.

International attention around early childhood development is fairly new. Much of the scientific evidence supporting the need for ECD comes from work within the United States and other developed countries that has recently been able to thoroughly explore the post-early childhood outcomes of interventions. Many low- and middle-income countries, confronted by different development challenges, have

OVERVIEW

focused their efforts primarily on child survival and growth, subsequently limiting their ability to invest in interventions such as preschool and other stimulation programs.

Although this picture is slowly changing, very little is known about the effectiveness of ECD programs across the full range of outcome domains in developing countries, particularly in the post-early childhood period. While many early studies in developing countries, such as the INCAP supplementary feeding program in Guatemala and the maternal biscuit intervention in The Gambia, made important discoveries about the short-term effects of nutritional interventions during the first 1,000 days, the ability of researchers to assess the longer-term effects of these programs is limited by the design of the initial studies, which was not intended for follow-up.

At the time these evaluations were implemented, it was not yet widely understood that early childhood interventions could have sustained effects on ECD outcome domains, and it will be particularly important for future evaluations of ECD programs to facilitate long-term follow-up. The logistical difficulty of conducting long-term follow-up studies, particularly in the developing country context, has further contributed to a dearth of research.

One goal of this review is to further clarify the existing knowledge base of

long-term effects of early childhood interventions and help inform future evaluations. Early childhood interventions can impact a variety of cognitive, linguistic, socioemotional, physical, educational, and employment outcomes; however, this review identified just 55 impact evaluations across all possible intervention types and outcome domains that passed the quality check. The evidence in several domains was particularly thin. For example, only one study measured the effects of ECD programs on employment outcomes, and only one intervention type – stimulation – had long-term effects measured in each of the six outcome domains. Future research should aim not only to provide more evidence across the full range of possible outcomes throughout an individual's lifespan, but also to expand the scope of interventions evaluated for their effects.

Challenges

Long-term follow-up evaluations face logistical challenges that contribute to the knowledge gap.

Evaluations aiming to estimate impacts of early childhood programs after a prolonged period face four challenges: confounding variables, attrition, designing for follow-up, and external validity. These challenges are not unique to evaluations of early childhood interventions or to longer-term evaluations, but they may be

compounded here. In particular, issues of attrition and confounding variables are primarily responsible for the exclusion of more than half of the impact evaluation studies otherwise eligible for this review.

The evaluations that today constitute the evidence base of later-life effects of ECD interventions were often not designed with that objective. Many study designs were implemented prior to the existence of strong evidence of effects across a range of outcomes in the post-early childhood years and were not designed to track participants into adolescence and adulthood.

Additionally, universal, standardized measurement across a range of outcomes over the lifespan – cognition and socioemotional development in particular – are not yet established, making it difficult to know how best to assess these constructs longitudinally. Although in some cases researchers are able to apply econometric methods based on analysis of past performance to tease out lasting effects that can be attributed to the original intervention, the absence of prior planning for long-term follow-up at the implementation stage has complicated causal inference from these studies.

Finally, evaluations of all types, including impact evaluations, have challenges of external validity – the ability to apply results found in one study to a different scale, context, or time. Most of the interventions evaluated here are somewhat small;

scaling up to a national level may present administrative or other challenges. Furthermore, the longer-term nature of the interventions means that interventions included here occurred in an era – sometimes 30 years ago – that may have influenced interactions with the project in very different ways than would be expected in the contemporary context, even if in the exact same location. Potential variation across location underscores the need to fill in Regional gaps.

Implications

While much has been written on early childhood development and the near-term benefits to children selected to participate in interventions, few studies look at the sustained impacts of these programs. At this point in the development of the literature, this systematic review aims to contribute to the field's progress by collecting those studies that offer high-quality, causal estimates, providing analysis on interventions shown to have sustained effects across a range of outcomes, and identifying research gaps to help guide future studies.

Design Challenges

Evaluation design is critical in yielding valid causal estimates. Design is a major determining factor in how similar the comparison group is to the treated group, which in turn is the basis for calculating attributable program effects. Problems in design can be exacerbated over time as they interact with other factors, compromising the comparability of the two groups.

Two well-known interventions, the program of the Instituto de Nutrición de Centroamérica y Panamá (INCAP) in Guatemala and the Turkish Early Enrichment Project, suffered from this problem. Because of weaknesses in the initial randomization, many of the INCAP studies rely on a comparison group that is not statistically equivalent to the treated group, while in Turkey, group comparability suffered from selection bias in one wing of the study and extremely high attrition rates overall. Such design challenges can undermine causal inference.

The results of this outcome-based systematic review imply that some domains may be easier to affect than others. In cognitive, linguistic, socioemotional, and employment domains, the evidence suggests that early stimulation interventions can result in sustained benefits to children, and various interventions were successful in improving subsequent schooling outcomes. Conversely, there was little evidence of a strong post-early childhood effect in physical outcomes across the range of evaluated interventions.

Despite these promising results, much work remains to be able to draw firm conclusions on the post-early childhood effects of ECD interventions. With 55 impact evaluation studies on 25 projects of 20 intervention types across 22 countries, the results presented are still indicative rather than conclusive. High-quality evaluations of interventions that could significantly impact a child's development, including nontraditional interventions such as clean water and sanitation or agriculture, are hard to find but are necessary to determine the most effective method of intervening. Furthermore, evaluations across Regions are important to capture context-specific variables and improve general external validity. The analysis of heterogeneous effects, especially by gender and socioeconomic status, can improve intervention targeting.

Early childhood interventions, like the children they serve, have transformative potential if properly supported. As the old proverb says, the best time to plant a tree is 20 years ago, but the second best time is today. That applies to investments in children as well as it applies to investments in evaluations that can track their progress.

Introduction: Review Questions and Strategy

While the later-life effects of interventions during the early childhood period are generally well documented in the developed world, far less evidence exists in developing countries. To address this knowledge gap, the review uses a comprehensive search strategy and a rigorous screening method to compile the causally robust evidence on the post-early childhood effects of early childhood interventions in developing countries. The evidence is organized around six outcome domains to determine which intervention types can effectively and consistently produce sustained effects in particular areas of child development.

In the *World Development Report 2006: Equity and Development*, the World Bank singled out early childhood development (ECD) interventions as a promising policy area to achieve both equity and efficiency objectives (World Bank 2005). More recently, in the *World Development Report 2015: Mind, Society, and Behavior*, the Bank again chose child development as a key facet of inequality, noting that children from developing nations have systematically lower socioemotional and cognitive stimulation in their early years, which together with the high stress of growing up poor can impair decision-making and deliberative abilities (World Bank Group 2015).

These flagship reports perceive that ECD can play a key role in achieving the Bank's twin goals to reduce extreme poverty and promote shared prosperity while encouraging economic growth. The realization of these aims is predicated on the ability of countries to "build human capital and increase long term productivity" through "access to early childhood development, health, nutrition, education, and basic infrastructure," which enhances "mobility on the economic and social ladder within and across generations" (World Bank 2013, p. 28). Programs targeted at ECD do just that – build human capital – by intervening during a critical period of development when it is suggested that interventions can improve both the starting point and trajectory of a child's life path and provide a longer time horizon over which the benefit stream is realized (Carneiro and Heckman 2003).

Early childhood development is an integrated construct influenced by many factors, such as nutrition, health, hygiene, early learning, and stimulation. For example, good nutrition during the first 1,000 days, from conception to the child's second birthday, is essential for normative linear growth (Victora and others 2008; Black and others 2013) and health brain formation (Couperus and Nelson 2006; Georgieff and Rao 1999). The plasticity of the young brain (that is, its capacity to change) allows young children to benefit from positive inputs such as stimulation and nutrition but also makes them vulnerable to negative external shocks including deprivation and abuse (Shonkoff and

others 2012a; Fox and others 2010). Cognitive, language, fine motor, and socioemotional skills important for educational and social success emerge during these years (Heckman 2008; NRC and IOM 2000; Shonkoff 2011). Risk factors related to poverty (for example, undernutrition, poor sanitation, insensitive parenting) in early childhood are associated with delays in these skills as well as in school progress (Georgieff 2007; Grantham-McGregor and others 2007; Walker and others 2007; Glewwe, Jacoby, and King 2001).

Experimental evidence suggests that nutrition, health, early learning, and other factors can play a major role in shaping young children's subsequent school attainment, performance, and earnings (Heckman 2008; Naudeau and others 2011; Barnett 2011; Duncan and others 2007). Indeed, the benefits of a variety of early childhood interventions are well documented in developed countries.¹ Long-term evaluations of children who received these interventions in the United States found positive life outcomes in education, health, fertility, risky behaviors, and the labor market (Smith 2009; Cunha and Heckman 2009; Schweinhart 2007; Campbell and others 2002; Camilli and others 2010; Reynolds and others 2007; Anderson and others 2003; Bitler, Hoynes, and Domina 2014; Olds, Sadler, and Kitzman 2007; Sweet and Appelbaum 2004). This evidence led many in the international development community – including the World Bank – to promote ECD and to focus on interventions as the first step in a sequenced strategy to build the skills needed for productivity and economic growth (World Bank 2010).

In developing countries, numerous studies and reviews establish that early childhood interventions can improve early childhood outcomes (Maternal and Child Nutrition Study Group 2013). Yet the ability of ECD interventions to improve later outcomes – those occurring after the early childhood period² – is less well studied. Impact evaluations (IEs) that examine the post-early childhood effects from early childhood interventions in middle- and low-income countries are scarce, and it may be problematic to extend the findings from the United States to developing countries. While the challenges of developing countries differ from place to place, on average they face challenges that Organisation for Economic Co-operation and Development countries do not, such as weaker institutions and lower economic development. It follows that interventions that are effective in wealthy nations may not have the same results in the resource-constrained contexts of low- and middle-income countries, especially given that basic needs in these countries are often less well met.

The converse may also be true – interventions that are effective in developing countries may not yield the same results in wealthier contexts where basic needs are met.³ This dynamic may explain differences in the set of evaluated interventions by national income.⁴ Nevertheless, evidence from developed countries can help establish the physiological pathways between particular inputs and ECD outcomes. However, as was

discovered in IEG's systematic review of maternal and child mortality (IEG 2013), even where what to do is known (and even this is not always the case with ECD), knowing how to do it under such different circumstances is a persistent challenge.

This review by the Independent Evaluation Group (IEG) tries to address that challenge by gathering and analyzing the available impact evaluation evidence on the post-early childhood effects of interventions conducted during the early childhood years in the developing world. The purpose of doing so is not to supplant existing evidence, but rather to help practitioners understand how evidence from IEs supports or challenges beliefs about ECD interventions and how this evidence can be used to inform development policy. Impact evaluations aim to overcome confounding factors inherent in other forms of evaluation to determine the causal impact of an intervention. This highly rigorous form of evaluation is particularly well suited to evaluate the claims of lasting effects from early childhood interventions, considering the number and scope of confounding factors that arise over time.

By taking such a specific approach—focusing exclusively on the post-early childhood effects of any ECD intervention—this review goes beyond the existing work done in this field, an anthology of which is found in appendix E.⁵ For instance, a 2011 review published in *The Lancet* series on ECD in low- and middle-income countries included some evidence on post-early childhood effects, but these outcomes were a relatively minor part of the review (Engle and others 2011).

The present report differs from *The Lancet* piece in the breadth of the interventions and outcomes included, the specificity of the age at evaluation, and the types of studies accepted for inclusion. While Engle and others use an intervention-based approach to focus on psychosocial and educational interventions and the resulting range of outcomes, this review includes IEs of any type of intervention occurring during the early childhood period and presents them through an outcome-based framework (Waddington and others 2012). Additionally, unlike *The Lancet* review, which included outcomes measured during the early childhood period for both children and parents, this review only reports outcomes in the post-early childhood period for children. Finally, while *The Lancet* includes both efficacy⁶ and effectiveness studies implemented using local or foreign capacity, this review is restricted to those interventions that use local capacity so as to provide evidence on interventions that could actually be replicated by low- and middle-income countries.

A systematic review published recently in the *Annals of the New York Academy of Sciences* examined the effect of integrated interventions (Grantham-McGregor and others 2014) in low- and middle-income countries, but again its scope differs in significant ways from this review. As with *The Lancet* review, Grantham-McGregor and others focused

on a narrow group of interventions (stimulation and nutrition), included a broader range of outcomes and ages at evaluation (both parent and child outcomes measured during the early childhood period and after), and drew evidence from interventions implemented using local or foreign capacity. It should also be noted that their use of evidence from the post-early childhood period was limited.

Despite having received relatively little attention, post-early childhood benefits are quite important as they comprise a major economic argument for investing in young children based on the assumption that returns continue over time. Budget and time constraints, together with the estimation challenges inherent in follow-up IEs, may be responsible for the relatively few IEs that address later-stage outcomes of early childhood interventions. However, it is important to take stock of the causal evidence that does exist. With that in mind, this review uses evidence from all impact evaluations with credible causal estimates for interventions occurring before primary school age on outcomes occurring at or after primary school enrollment. It aims to answer the following questions:

- What is the evidence of attributable effects on outcomes in primary school and beyond from interventions in low- and middle-income countries that occur during the early childhood period?
- How do the post-early childhood effects of early childhood interventions vary by socioeconomic status, gender, age at intervention, and age at evaluation, particularly during the first 1,000 days from conception to the child's second birthday and from age three to primary school enrollment at age five to six?

For the purposes of this review, the defined intervention period for early childhood runs from conception to primary school enrollment of the child. Accordingly, all early childhood interventions are considered that are either directly provided to the child or to the parent, caregiver, or pregnant or lactating mother. This includes a range of interventions such as preschool, conditional cash transfers, and behavior change interventions such as early stimulation by caregivers, promotion of exclusive breastfeeding, complementary feeding practices, and health and hygiene practices.

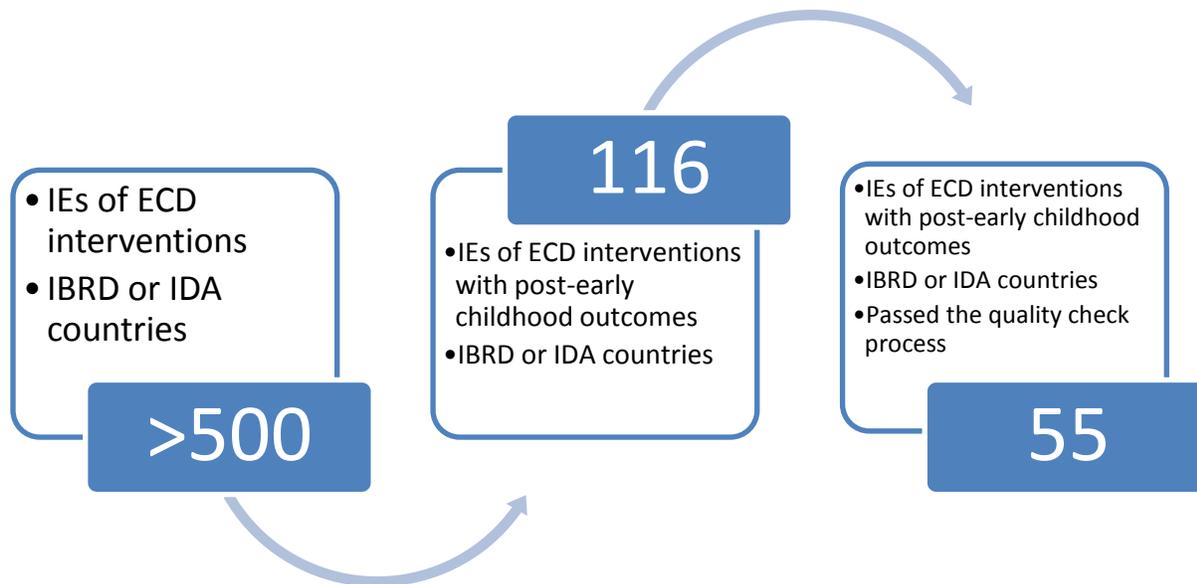
Search Strategy for Identification of Relevant Studies

Studies were identified through a detailed search strategy based on an approach used by a previous systematic review in IEG's Maternal and Child Health series (IEG 2013). Using search terms that encompassed the outcomes, methods, and the definition of low- and middle-income countries employed by the International Bank for Reconstruction and Development (IBRD) and the International Development Association (IDA), studies

between 1990 and 2014 were collected from relevant bibliographic databases and the World Bank’s impact evaluation database. This was supplemented by searches of relevant institutions and a hand search of top ECD authors, followed by the snowballing of references from included studies and related systematic reviews identified during the electronic search. The search included both published impact evaluations and unpublished grey literature, defined as working papers and studies soliciting review from the research community, in so far as they were identifiable by the established search procedure. (See appendix C for more information on the search protocol.)

Initial searches produced more than 500 potentially-relevant evaluations of ECD interventions in IBRD or IDA countries; 116 provided estimates for post-early childhood outcomes. A rigorous quality check against the inclusion criteria yielded 55 relevant studies of sufficient quality to include in the analysis (figure 1). These studies represent 20 intervention types (see figure 2) across 25 projects in 22 low- and middle-income countries.

Figure 1. Flow of Search Results



Note: ECD = early childhood development; IBRD = International Bank for Reconstruction and Development; IDA = International Development Association; IEs = impact evaluations.

Despite this relatively thin evidence base, these 116 studies constitute the complete body of IE knowledge on the post-early childhood effects of ECD programs. The number is expected to grow. It should be remembered that the field of early childhood development is still relatively new, especially for developing country contexts. The pioneering IEs that offer evidence of later-life effects of early childhood development interventions were designed decades ago. Considerable evolution of both ECD theory

and empirical standards and design has occurred since the initiation of some of the earliest studies reviewed (for example, the supplementary feeding program of the Instituto de Nutrición de Centroamérica y Panamá was begun in the 1970s), and most were not designed to serve as platforms for understanding long-term effects. This is not to critique the garden of child development evidence for being newly sown. Rather, in light of this dynamic evidence base, this review seeks to take stock of what is known and point out where further cultivation might bear fruit.

Box 1. Analyzed Outcomes: The Rule of Three

This review reports outcomes across six different domains: physical development, cognitive development, language development, socioemotional development, schooling outcomes, and employment and labor market outcomes. Outcomes were selected for inclusion based on how often they were measured and reported and the extent to which they provided unique insight into measurements within the domain.

Outcomes that were reported in three or more independent studies were included. The high frequency of these outcomes was indicative of their usefulness in measuring and understanding early childhood development, and enabled comparison across contexts and intervention types.

In some cases, however, an outcome may have been reported in more than three studies, but was not included due to high overlap with other indicators. For example, both fat mass index and body mass index (BMI) were reported in many studies, but this review only included BMI measurements. Both indicators capture a similar measurement of a child's physical development, and including both would have provided little unique insight.

An exception to the "rule of three" was made for outcomes in employment and fine motor skills, both of which have high intrinsic value. Fine motor skills provide a useful measure of the development of an individual's executive function capabilities – an outcome domain in which little post-early childhood evidence exists. Similarly, given the focus on long-term outcomes in this review, employment outcomes are analyzed even though they are infrequently reported.

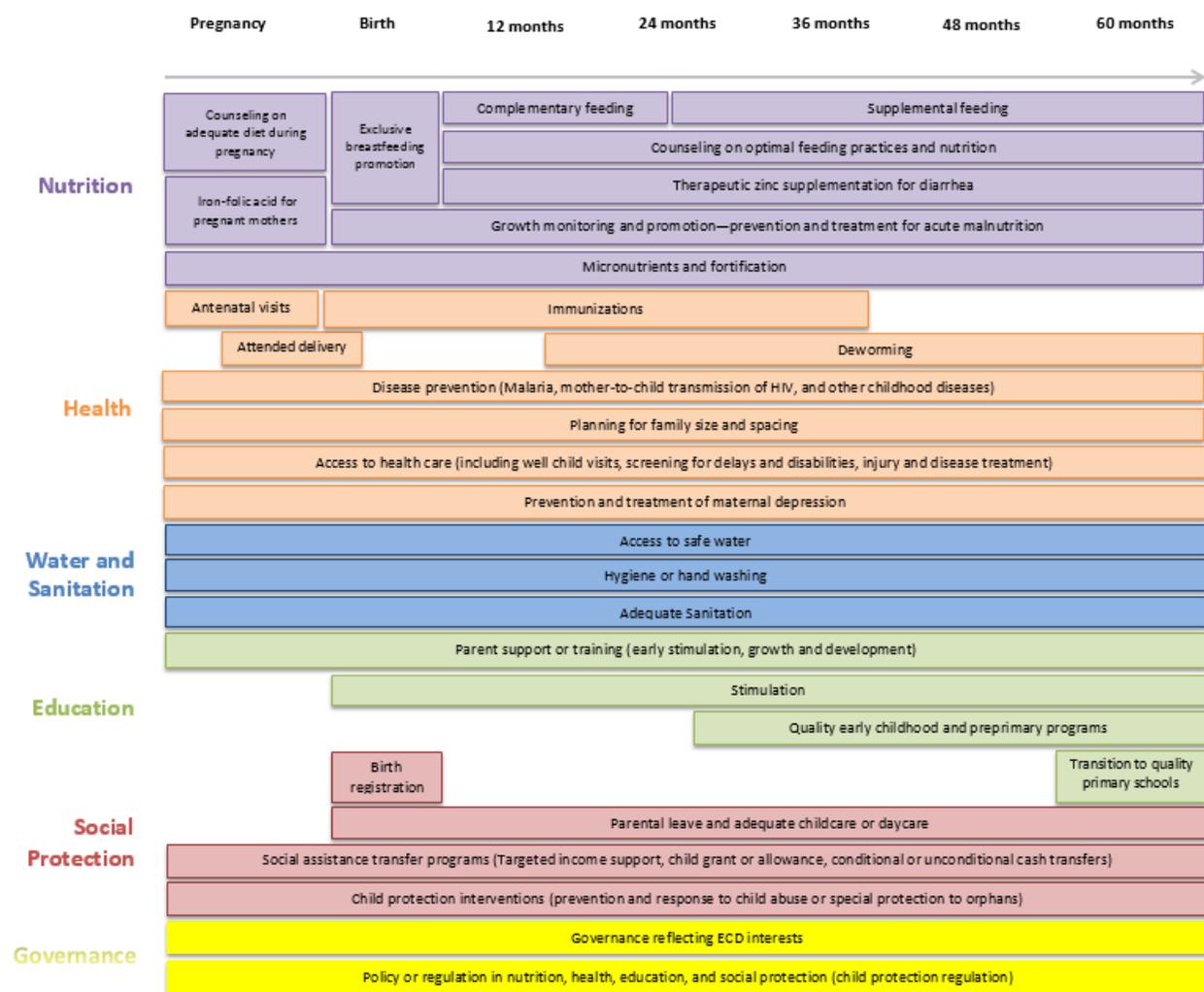
Criteria for Inclusion and Exclusion of Studies

The inclusion criteria required that the studies (i) employ a quantitative impact evaluation methodology using experimental or quasi-experimental identification strategies with a credible counterfactual⁷ to identify causal attribution; (ii) evaluate any intervention occurring in a low or middle-income country;⁸(iii) limit the population of analysis to children exposed to the intervention during the early childhood period, defined here as conception to the age of entry into primary school (or age six, when information on primary school age is not available);⁹ (iv) assess post-early childhood outcomes occurring from primary school enrollment through adulthood; (v) be

published between 1990 and the present (February 2015) and use end-line data occurring no earlier than 1990; and (vi) were subject to some form of peer review. To reduce the risk of publication bias, grey literature was included for consideration if a full-text version is publicly available and the study passes other quality and inclusion criteria. These inclusion criteria were selected to best answer the evaluation questions above. These criteria limited the body of admissible evidence; different evaluation questions may result in the inclusion of otherwise valid studies not incorporated into this review.¹⁰

A taxonomy of ECD interventions is presented in figure 2, categorized by intervention type and the ideal age at intervention. To be included in this review, a study could be on any type of intervention, whether a typical ECD intervention or not, as long as it was implemented through local capacity channels (for example, governments, nongovernmental organizations, or private sector firms) and would therefore be more easily replicated by local implementers.

Figure 2. Essential Interventions for Young Children and Families



Source: Denboba and others (2014), adapted by IEG.

All included studies had a sample size of 50 or more and an attrition rate lower than 40 percent.¹¹ The unit of analysis could be at the individual, household, facility (for example, school, center, or clinic), or community level.¹² Studies using aggregated national or regional data, as in cross-country or national interrupted time series analyses with few observations over multiple periods are excluded from this review.

All studies meeting the inclusion criteria were subjected to a double-coded quality review that assessed the strength of the internal, external, and construct validity. The primary criteria for this assessment was the extent to which the identifying assumptions of the identification strategy used were met, whether the data were representative of a definable and policy-relevant population, and whether the key indicators and outcomes in the report were measured in an unbiased and reliable way. (See appendix D for information about the rating process and coding protocol.)

Based on these criteria, studies (or in some cases specific estimation strategies within a study) were assigned an overall quality rating: A, AA, or AAA. As in past IEG systematic reviews, only AA and AAA studies were included in the review. Studies with an AA rating have a credible counterfactual with identification assumptions plausibly fulfilled; those with an AAA rating have a credible counterfactual with identification assumptions clearly fulfilled. Studies graded A leave significant doubts about the validity of the counterfactual and the likelihood that its identifying assumptions have been violated; for completeness these A-quality studies are listed in appendix F.

This process identified 55 sufficiently high quality impact evaluations used for synthesis in this review. Independent of the rigorous standards applied in the search and coding process, challenges inherent to systematic reviews remain (see box 2).

Relatively few impact evaluations can answer the important questions about long-term effects of early childhood interventions posed by this review. Moreover, because the evidence that is available is spread across a broad range of intervention and outcome combinations, the evidence for any particular combination can be quite thin, or completely nonexistent. Thus, although the review contains all causally robust evaluations, results should be viewed as comprehensive but nascent and prone to change as new results are published.

In light of the challenges endemic to reviews (see box 2), rather than assuming that evidence gaps indicate interventions that do not work, this report focuses on interventions shown to be effective. One test of robustness of a review is whether the main messages would change if the cutoff date for inclusion were altered. As explained in appendix C, this review completed an initial search and analysis ending in 2013 and then refreshed that search prior to publication to include material released in 2014. Although there was a substantial uptick in the number of studies from which to draw evidence, the conclusions remained almost entirely unchanged. This finding supports the view that the conclusions included here are likely to be broadly stable as still more evaluative, causal evidence is produced.

Box 2. Challenges and Cautions for Systematic Reviews

Notwithstanding the thoroughness of the search strategy, challenges remain to representativeness and interpretation of results that are common to all systematic reviews. This review meets or exceeds standard practice, such as it exists, in every instance.

Challenges to representativeness of the interventions arise from the fact that the process of selecting interventions to be evaluated by an impact evaluation is purposeful rather than random.

- Some types of projects are less amenable to impact evaluation methods and will be underrepresented.
- Interventions that report on intermediate rather than final outcomes are excluded.
- Importantly, the lack of existing impact evaluations for a family of interventions indicates a need for evaluations in that area, not that the interventions are ineffective.

Challenges to representativeness of the impact evaluations are twofold.

- The review includes only concluded studies; it cannot use impact evaluations that are planned or in process.
- As with all reviews, the sample may suffer from file drawer bias or publication bias wherein studies that yield null results are not completed. Alternately, it has been hypothesized that only studies with experimental designs can be published with statistically insignificant results because of stronger internal validity; this may lead to a false conclusion that randomized trials are more likely to return null results.

Challenges to interpretation of results imply a need for thoughtful application of findings.

- Impact evaluations of projects funded by foreign aid likely underestimate the true effect of the intervention because they measure partial (or local) equilibrium effects rather than the general equilibrium effects resulting from the fungibility of government budgets, which allows countries to reallocate health funding away from the foreign-funded activities (Wagstaff 2011).
- Null results must be interpreted carefully: they do not necessarily mean there is no effect. They may occur where there is measurement error, insufficient sample size (power) to detect an effect, spillover from treatment to the control group, differential attrition, insufficient behavioral incentives, or implementation challenges. Distinguishing the causes of a null result is often untenable.
- External validity is a persistent challenge. Applicability of results to a different context—time, place, or scale—is likely a function of project complexity (Woolcock 2013), administrative capacity, political supportability, and alignment with the most important barriers of the target environment. The ideas and processes of an intervention may have greater external validity than the intervention itself.

Consensus analysis was carried out across the studies comparing statistical significance and magnitude in forest plots. Where sufficient numbers of studies exist with outcomes in a common construct, meta-analysis was conducted to explore the presence of an overall effect.

The report presents evidence for post-early childhood outcomes of early childhood-age interventions over six important domains: physical, cognitive, language, socioemotional, schooling, and employment and labor market outcomes. Chapter 7 breaks down heterogeneous effects and the differential outcomes observed both between subgroups and over time. Chapter 8 discusses the challenges of evaluating long-term effects of interventions conducted during the early childhood years, and chapter 10 identifies the remaining gaps in knowledge about the post-early childhood effects of ECD interventions.

¹ In developing countries, preschool participation contributed to an increase in lifetime earnings by 5–10 percent (Engle and others 2007, 2011). Belfield and others (2000) computed the benefits of the Perry Preschool Program in the United States to be \$150,000 (in 2000 dollars) per child through age 40 because of crime reduction.

² The post-early childhood period is determined to start on enrollment in primary school or, when that information is unavailable, at age six years.

³ The effectiveness of a particular intervention is a function of, among other things, the complementary slackness of the constraint in the outcome's production function that that intervention is designed to address. For an example of a production function model of a range of inputs on a range of children's outcomes, see Tanner 2012 and Becker 1993.

⁴ This presents a mechanism for the endogeneity of the evidence base by national income. For example, in contexts where a particular input (say nutrition) has a lower shadow price on the production of a particular output, the effect size of an intervention designed to alleviate that (nutritional) constraint will be low. Thus there may be comparatively fewer interventions of that input in that context, and so there may be fewer evaluations of that (nutrition) input.

⁵ For a full review of the existing systematic review literature, see IEG (2015).

⁶ The World Bank's handbook *Impact Evaluation in Practice* (World Bank 2011) defines efficacy trials as having "heavy technical involvement from the researchers during the implementation of the program" and which do not use "regular implementation channels." For the purposes of the inclusion criteria of this review, this is codified such that studies are excluded which do not rely on local capacity for implementation of the intervention.

⁷ Constructing a valid comparison group for use as a counterfactual – the outcome that would have been observed for a participant in the absence of the program – is the defining feature of impact evaluation methods.

⁸ Despite being recently classified as high-income countries by the World Bank in July 2013, studies from both Chile and Uruguay are both included in this review due to the fact that end-line data for those studies was collected during the period when they were still categorized as middle-income countries. All countries included in this report are currently World Bank client countries.

⁹ The early childhood period is defined as the time between conception and a child's entry to primary school. If age at the time of entry to primary school was not mentioned, 59 months was used as a cut off. Therefore, these impact evaluations must include outcomes measured after

entry to primary school or age six years (post-early childhood) resulting from interventions that occurred before entry to primary school or age six years.

¹⁰ For example, impact evaluations of outcomes measured within the early childhood period are excluded, vacating a fairly substantial literature of nutrition and feeding interventions that demonstrate health benefits, sometimes even a year or more after the end of the intervention but still before primary school.

Similarly, in order to more cleanly demarcate and identify early childhood interventions and post-early childhood effects, evaluations whose intervention or evaluation ages straddle the age cutoff of primary school (or 59 months) are excluded. For example, follow-up studies of a deworming campaign in Uganda that included some seven-year olds and potentially students in the early primary grades are not considered, though they do show robust village-level effects (see Alderman and others 2006; Croke 2014)

¹¹ In the professional judgment of the team, sample sizes below 50 engender serious doubts pertaining to their external validity; lower sample sizes are also more likely to be under-powered, complicating the interpretation of null results. There was one exception: a single outcome estimate from one of the studies of the Jamaica supplementation and stimulation project had a sample size of 48. Similarly, attrition rates over 40 percent (many would say 20 percent or even less) are judged to be seriously challenged by selection bias. See Chapter 8 for a more detailed discussion on challenges to follow-up impact evaluations, including problems of attrition.

¹² For example, if all students at a given primary school had participated in the intervention, attendance or school performance could be measured at the school level. In practice, almost all of the outcomes were measured on the individual level.