

DOCUMENTO DE TRABAJO DEL BID N° IDB-WP-1230

De 70 a 700 a 70.000: Lecciones del Estudio de Jamaica

M. Caridad Araujo
Marta Rubio-Codina
Norbert Schady

Banco Interamericano de Desarrollo
División de Protección Social y Salud

Abril 2021

De 70 a 700 a 70.000: Lecciones del Estudio de Jamaica

M. Caridad Araujo
Marta Rubio-Codina
Norbert Schady

Catalogación en la fuente proporcionada por la
Biblioteca Felipe Herrera del
Banco Interamericano de Desarrollo
Araujo, Maria Caridad.

De 70 a 700 a 70.000: lecciones del estudio de Jamaica / M. Caridad Araujo, Marta
Rubio-Codina, Norbert Schady.

p. cm. — (Documento de trabajo del BID ; 1230)

Incluye referencias bibliográficas.

1. Home-based family services-Jamaica. 2. Home-based family services-Colombia. 3.
Home-based family services-Peru. 4. Child development-Government policy-Jamaica.
5. Child development-Government policy-Colombia. 6. Child development-Government
policy-Peru. I. Rubio-Codina, Marta. II. Schady, Norbert Rüdiger, 1967- III. Banco
Interamericano de Desarrollo. División de Protección Social y Salud. IV. Título. V.
Serie.

IDB-WP-1230

<http://www.iadb.org>

Copyright © 2021 Banco Interamericano de Desarrollo. Esta obra se encuentra sujeta a una licencia Creative Commons IGO 3.0 Reconocimiento-NoComercial-SinObrasDerivadas (CC-IGO 3.0 BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) y puede ser reproducida para cualquier uso no-comercial otorgando el reconocimiento respectivo al BID. No se permiten obras derivadas.

Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la CNUDMI (UNCITRAL). El uso del nombre del BID para cualquier fin distinto al reconocimiento respectivo y el uso del logotipo del BID, no están autorizados por esta licencia CC-IGO y requieren de un acuerdo de licencia adicional.

Después de un proceso de revisión por pares, y con el consentimiento previo y por escrito del BID, una versión revisada de esta obra podrá reproducirse en cualquier revista académica, incluyendo aquellas referenciadas por la Asociación Americana de Economía a través de EconLit, siempre y cuando se otorgue el reconocimiento respectivo al BID, y el autor o autores no obtengan ingresos de la publicación. Por lo tanto, la restricción a obtener ingresos de dicha publicación sólo se extenderá al autor o autores de la publicación. Con respecto a dicha restricción, en caso de cualquier inconsistencia entre la licencia Creative Commons IGO 3.0 Reconocimiento-NoComercial-SinObrasDerivadas y estas declaraciones, prevalecerán estas últimas.

Note que el enlace URL incluye términos y condiciones adicionales de esta licencia.

Las opiniones expresadas en esta publicación son de los autores y no necesariamente reflejan el punto de vista del Banco Interamericano de Desarrollo, de su Directorio Ejecutivo ni de los países que representa.



scl-sph@iadb.org

www.iadb.org/es/proteccionsocial

De 70 a 700 a 70.000: Lecciones del estudio de Jamaica

M. Caridad Araujo
Marta Rubio-Codina
Norbert Schady



2021

De 70 a 700 a 70.000: Lecciones del estudio de Jamaica

M. Caridad Araujo

Marta Rubio-Codina

Norbert Schady

Resumen[†]

Este documento compara tres versiones del mismo modelo de visitas domiciliarias, el conocido modelo de Jamaica, en su proceso de ser escalado de forma gradual desde un primer estudio de eficacia ('prueba de concepto') en Jamaica, a una prueba piloto en Colombia, a un programa a escala en Perú. Primero se describen el diseño, la implementación y el impacto de los tres programas. Luego se analizan los riesgos de ampliar la escala de implementación de la intervención en cada uno de los tres estudios y se examina cómo estos riesgos habrían podido afectar los resultados obtenidos, centrándose en tres de los elementos del modelo económico de escalamiento en Al-Ubaydly, et al. (disponible próximamente): la inferencia estadística adecuada, características de la población y características de la situación. El documento reflexiona sobre las lecciones aprendidas a raíz de estas experiencias con relación a cómo mejor alinear los objetivos de implementación, investigación y evaluación para facilitar y apoyar el proceso de llevar a escala intervenciones de desarrollo en la primera infancia. Se identifican aquellos atributos no negociables de las intervenciones que hay que mantener para preservar su efectividad a escala. En esta línea, se señala también como indispensable el apoyo de los gobiernos.

Palabras clave: visitas domiciliarias, calidad, desarrollo infantil, monitoreo, escalar una intervención o programa.

Códigos JEL: I38, J24, O15

[†]Agradecemos a Julieth Parra por su excepcional asistencia de investigación en la preparación de este capítulo.

Contenidos

1. Introducción	3
2. Las experiencias de Jamaica, Colombia y Perú	5
3. Los riesgos de escalar intervenciones	8
3.1 De 70 a 700: Jamaica y Colombia	8
3.2 De 700 a 70.000: Colombia y Perú	11
4. Lecciones aprendidas y preguntas adicionales	14
4.1 Lecciones aprendidas: Mitigando los riesgos de escalar	14
4.2 Lecciones aprendidas: Investigación y evaluación	16
5. Conclusión y cuestiones pendientes	17
Bibliografía	18
Figuras y tablas	21

1. Introducción

Este documento describe tres esfuerzos distintos para implementar y evaluar variantes de la misma intervención de visitas domiciliarias, orientada a mejorar las interacciones entre padres e hijos y el desarrollo infantil, en Jamaica, Colombia y Perú. Se trata de un análisis descriptivo. Si bien cada uno de estos esfuerzos se evaluó independientemente mediante experimentos aleatorios, es difícil hacer cualquier comparación estadística de las características y de los impactos entre los tres programas con los datos disponibles. Más bien, el objetivo de este ejercicio es documentar el proceso por el cual la intervención se escaló gradualmente, desde un estudio de eficacia en Jamaica ('prueba de concepto'), a un piloto en Colombia diseñado para ser escalado, hasta un programa gubernamental a escala en Perú.¹

El primero de estos programas se implementó en Kingston, Jamaica, a mediados de los años 1980. Un equipo de investigación elaboró un conjunto de materiales y actividades de juego para mejorar los resultados del desarrollo de niños en estado de malnutrición, mediante la estimulación psicosocial y el juego. La implementación estuvo a cargo de promotoras comunitarias de salud que, durante visitas domiciliarias semanales a sesenta y cuatro niños, modelaron actividades de juego e interacciones receptivas entre adultos y niños. Los hallazgos de un ensayo controlado aleatorizado (*Randomized Control Trial*, RCT) mostraron resultados importantes (0,88 desviaciones estándar [DE] en desarrollo global) después de 24 meses de visitas (Grantham-McGregor et al., 1991).² Hasta la fecha, se ha dado seguimiento a esos niños durante más de 30 años. Veinte años después de la finalización de la intervención, aquellos que habían sido asignados aleatoriamente al grupo de tratamiento habían completado más años de escolarización y tenían mejor rendimiento académico, más probabilidades de contar con un trabajo y salarios un 25% más altos que los asignados al grupo de control (Gertler et al., 2014). También obtuvieron mejores resultados en desarrollo cognitivo, desarrollo personal-social y de salud mental (Walker et al., 2011).

El segundo programa que describimos es la adaptación del modelo de Jamaica implementado a escala piloto en Colombia en 2010-2011. El objetivo principal de este esfuerzo era probar si la intervención de visitas domiciliarias podía integrarse en la infraestructura de un programa nacional de protección social como estrategia para escalar su implementación. Las visitas domiciliarias fueron realizadas por lideresas locales, conocidas como *madres líderes*, que habían sido elegidas por sus comunidades para hacer de enlace con los administradores del programa nacional de transferencias monetarias condicionadas del país, *Familia en Acción*. Un segundo objetivo de la evaluación fue identificar a través de qué mecanismos se canalizaban los impactos en el desarrollo infantil.

¹ La intervención se diseñó, desarrolló e implementó por primera vez en varios estudios en Jamaica, incluido el que describimos en detalle en este documento. Posteriormente, y hasta la fecha, se ha adaptado mejorado y reproducido en 18 ocasiones en 14 países diferentes. Todas estas réplicas han sido evaluadas. Este proceso ha conllevado el desarrollo del paquete *Reach Up and Learn* (Grantham-McGregor y Smith, 2016), que incluye materiales mejorados para la adaptación, implementación y capacitación del modelo en entornos de bajos recursos (Walker et al., 2018). Nos centramos en las experiencias de Jamaica, Colombia y Perú, ya que representan hitos en el proceso de llevar el modelo a escala.

² El desarrollo global (*Developmental Quotient*, DQ) incluyó las áreas siguientes: las subescalas de habilidad locomotora (motricidad gruesa), coordinación de manos y ojos (motricidad fina), audición y habla (lenguaje), y desempeño/rendimiento (cognición) de las Escalas de Desarrollo Mental de Griffiths.

Alrededor de 720 niños recibieron visitas domiciliarias durante 18 meses. Un ensayo controlado aleatorizado por conglomerados (*Cluster Randomized Control Trial*, CRCT) encontró resultados en el desarrollo infantil a corto plazo: 0,26 DE en la cognición, 0,22 DE en el lenguaje receptivo, y 0,18 DE en un índice agregado que combina la cognición, el lenguaje receptivo y expresivo, y la motricidad fina (Attanasio et al., 2014). Análisis adicionales con modelos estadísticos mostraron que estos resultados se debían principalmente al aumento de las inversiones de los padres en recursos materiales y de tiempo para sus hijos –es decir, a un aumento del número de materiales de juego y de la variedad de actividades de juego que el niño realizaba con un adulto (Attanasio et al., 2020). Sin embargo, no se encontró evidencia de que los efectos persistieran dos años después de que terminara el programa (Andrew et al., 2018).

El tercer programa que analizamos, el programa Cuna Más en Perú, fue creado tomando como modelo el programa de Colombia.³ Fue una verdadera intervención a escala, implementada por el Ministerio de Desarrollo e Inclusión Social. Cuna Más tenía como objetivo dar servicio a todos los niños de 0 a 3 años en localidades rurales con altos niveles de pobreza y desnutrición crónica, y estaba geográficamente focalizado en las zonas donde operaba el programa nacional de transferencias monetarias condicionadas, *Juntos*. Durante el período de su evaluación, Cuna Más amplió la cobertura de 5.338 niños en abril de 2013 a 67.332 niños en diciembre de 2015. Actualmente atiende a más de 115.000 niños (Figura 1).

La evaluación, realizada por CRCT, incluyó a 3.530 niños en el grupo de tratamiento y mostró mejoras modestas pero significativas (0,10 DE) en un índice agregado de desarrollo infantil que combina el desarrollo cognitivo, de lenguaje, de motricidad fina, de motricidad gruesa y de las habilidades personales y sociales (Araujo et al., 2019). Los impactos fueron mayores (0,14 DE) en los niños asignados al programa que recibieron al menos una visita domiciliaria.⁴ También se constató que los cuidadores del grupo de tratamiento jugaban con sus hijos más a menudo y usaban menos estrategias de disciplina punitiva que los del grupo de control. Además, entre los niños de 36 meses o más, edad en la que dejan de recibir visitas, los niños asignados aleatoriamente al grupo de tratamiento tenían una probabilidad 9 puntos porcentuales mayor de estar inscritos en educación preescolar, con respecto a unas tasas de inscripción del 50% en el grupo de control.

Si bien el espíritu de la intervención es común en los tres países, su diseño y aplicación difirieron de varias maneras. El proyecto de Jamaica fue un estudio de eficacia, altamente controlado. La experiencia en Colombia trató de adaptar y aplicar el modelo de manera que pudiera reproducirse a escala, descansando para ello en los recursos comunitarios existentes y una estructura de supervisión de baja intensidad. Ambos experimentos se llevaron a cabo en el contexto de proyectos de investigación. En Perú, en cambio, se hizo una implementación a

³ Además del servicio de visitas domiciliarias, Cuna Más opera jardines de cuidado infantil en zonas urbanas de ingresos bajos. Por practicidad, en este documento nos referimos al servicio de visitas domiciliarias como Cuna Más, sin hacer explícito que solo nos referimos a uno de los dos servicios proporcionados por el programa.

⁴ Una tercera parte de los hogares en el grupo de tratamiento en la muestra de evaluación no recibió visitas domiciliarias. Los motivos se describen más adelante en el documento.

escala del modelo por parte de una agencia de gobierno, acompañada de una muy rápida expansión de cobertura.

Enmarcamos nuestra comparación de los tres programas usando elementos del modelo de escalar intervenciones presentado en Al-Ubaydli, et al. (disponible próximamente): (a) la inferencia estadística adecuada, (b) las características de la población y (c) las características de la situación.⁵ Organizamos nuestro análisis del diseño, la implementación y los impactos de los programas de Jamaica, Colombia y Perú en torno a estos tres elementos.

El resto del documento está organizado de la siguiente manera. La próxima sección describe las tres intervenciones. En la siguiente sección, se comparan los programas utilizando el marco del modelo de Al-Ubaydli et al. (disponible próximamente) para discutir la gran reducción observada en el tamaño de los impactos de los programas a medida que se amplía la escala de la intervención. En la sección final se reflexiona sobre las lecciones aprendidas.

2. Las experiencias de Jamaica, Colombia y Perú

Comenzamos con una descripción más detallada de las principales características del diseño de las tres intervenciones de visitas domiciliarias analizadas, presentadas en la Tabla 1.

Los tres programas compartían el mismo (a) modo de implementación –visitas semanales a cargo de promotoras comunitarias (facilitadoras⁶) con niveles educativos bajos en promedio; (b) metodología –visitas basadas en el juego que buscaban promover el desarrollo infantil y las interacciones entre cuidadores y niños; y (c) un currículo básico de actividades estructuradas y materiales de juego –principalmente juguetes caseros de bajo costo, bloques, libros ilustrados, rompecabezas, actividades de clasificación y emparejamiento– con adaptaciones al contexto local.

Las actividades del currículo se organizaron por orden de dificultad para facilitar el proceso de andamiaje (es decir, la práctica de enseñar a un niño algo nuevo a través de y construyendo sobre aquellas cosas que ya sabe), así como en torno a las rutinas diarias de la familia. El objetivo era trabajar a través de los cuidadores, construyendo una relación positiva con ellos, a fin de fortalecer sus habilidades y su disfrute al promover el desarrollo infantil y ayudar a sus niños a aprender (Walker et al., 2018). Con ese objetivo, durante cada visita, la facilitadora modelaba actividades que introducían nuevos conceptos y retos.

Específicamente, la facilitadora hacía demostraciones de acciones para alentar a los cuidadores a involucrarse en el juego con sus hijos, a conversar con ellos y a responder a sus vocalizaciones y acciones. También les alentaba a elogiar al niño y celebrar sus esfuerzos y

⁵ El modelo presenta un cuarto elemento, los efectos colaterales y de equilibrio general, el cual no se analiza en este documento.

⁶ A lo largo del documento usamos el término “facilitadora” para referirnos a la persona que realiza la visita domiciliaria (*home visitor*) independientemente de su género, formación y experiencia –aspectos que difieren entre programas, tal y como se describe en la Tabla 1. Asimismo, usamos el término “supervisora” independientemente de su género. La Tabla 1 muestra el perfil para facilitadoras y supervisoras de cada programa.

logros para promover su autoestima y desarrollo socioemocional. Se hacía especial hincapié en escuchar la opinión del cuidador y darle ánimos. En última instancia, el programa buscaba generar cambios de comportamiento en las prácticas de crianza que perduraran más allá de la intervención.⁷

A pesar de estas similitudes, hubo diferencias de implementación importantes entre los programas. En Jamaica, participaron niños de 9 a 24 meses de edad y el programa duró 24 meses, por lo que atendió a algunos niños hasta los 4 años. En Colombia, los niños recibieron visitas durante un período más corto, 18 meses, desde los 12-24 meses de edad hasta los 30-42 meses de edad. El programa peruano fue el más ambicioso en cuanto a la duración prevista y la edad de inicio; en él participaron las madres embarazadas y los niños desde su nacimiento hasta los 24 meses de edad, con visitas que comenzaban en el embarazo y se prolongaban hasta que el niño cumplía 3 años.

La población seleccionada también difirió entre estas intervenciones. En promedio, los niños que recibieron la intervención en Jamaica y Perú vivían en comunidades con mayores niveles de pobreza –en barrios marginales urbanos en Jamaica y en comunidades rurales pobres y remotas en Perú– y experimentaban tasas más altas de desnutrición crónica⁸ que los niños que recibieron la intervención en Colombia. Esto puede ser importante en términos del potencial de la intervención para mejorar los indicadores de desarrollo con respecto a los niveles iniciales (en la línea de base), así como para mejorar la cantidad y calidad de la inversión de los padres en sus hijos. Estudios anteriores han demostrado que, durante la primera infancia, el impacto de las intervenciones suele ser mayor entre los niños de condición socioeconómica más desventajada (Bitler et al., 2014 en Estados Unidos; y Havnes y Mogstad, 2015 en Noruega), aunque no siempre es así (Nores et al., 2019 en Colombia).

En la Tabla 2 se presenta información sobre la implementación de los programas, principalmente a partir de los registros administrativos. La escala difiere sustancialmente entre las tres experiencias, lo que da lugar a diferencias importantes.

Algunas de las variables de implementación, como el número de visitas o la rotación de facilitadoras, pueden ser consideradas como *proxies* de la calidad estructural. A diferencia de Jamaica y Colombia, Perú se enfrentó a desafíos sustanciales para atender a todas las familias y fomentar su participación en el programa (tan solo el 66% de los niños del grupo de tratamiento de la evaluación recibió al menos una visita domiciliaria), garantizar el régimen de visitas (solo ocurrieron la mitad de las visitas previstas) y retener a las facilitadoras y a las supervisoras (que, en promedio, permanecieron en el programa durante 17-18 meses).

⁷ Si bien los cuidadores son fundamentales para el éxito de la intervención –al igual que la mejora de su autoestima y autoeficacia– el modelo no se centra específicamente en la salud o el bienestar económico de los cuidadores, ni les proporciona atenciones directamente; tampoco proporciona servicios de cuidado infantil a las familias. En este sentido, este modelo es diferente de *Nurse Family Partnership* (Olds, 2010) y *Early Head Start* (Love et al., 2013), dos de los programas de visitas domiciliarias más investigados en los Estados Unidos.

⁸ Se trata de niños que tenían un puntaje de talla para la edad de más de dos DE por debajo de la mediana de la talla para la edad de la población de referencia de niños sin problemas de malnutrición de la OMS.

Estos desafíos se debieron posiblemente al gran número de beneficiarios del programa, que además se encontraban repartidos por una vasta zona geográfica.

Los tres programas fueron evaluados mediante RCT, cuyas características se comparan en la Tabla 3. En Jamaica y Colombia, las evaluaciones fueron diseñadas y dirigidas por investigadores. El caso de Perú fue algo distinto. El Ministerio de Desarrollo e Inclusión Social, que operaba Cuna Más, tenía un mandato técnico de evaluar sus intervenciones. Este estaba estrechamente alineado con un proceso de presupuestación basada en resultados, dirigido por el Ministerio de Economía y Finanzas. El decreto mediante el cual se creó Cuna Más requería que se realizara una evaluación de impacto experimental y, solo una vez finalizada dicha evaluación, el programa podría convertirse en una partida permanente en el presupuesto del gobierno.

En Jamaica y Colombia se utilizaron pruebas diagnósticas para evaluar los resultados de los programas en el desarrollo de los niños: las Escalas de Desarrollo Mental de Griffiths en Jamaica y las Escalas de Desarrollo Infantil de Bayley en Colombia. Ambas escalas eran largas y complejas de aplicar –tenían muchos ítems y debían ser administradas por profesionales. Su administración también es costosa ya que requiere del pago de licencias por su uso (por niño evaluado) y de la compra de kits de materiales muy caros. En Perú, en cambio, se utilizó una prueba de tamizaje, los Cuestionarios de Edades y Etapas (*Ages and Stages Questionnaires*, ASQ), porque la administración de pruebas diagnósticas completas (como las pruebas Griffiths o Bayley) se consideró demasiado compleja y costosa, sobre todo teniendo en cuenta que la muestra de la evaluación era de mayor tamaño. Las pruebas de tamizaje están diseñadas para identificar a los niños con retrasos en el desarrollo y, por lo tanto, se centran en la parte inferior de la distribución de habilidades, lo que puede reducir su capacidad para detectar impactos.⁹ Además, en Perú algunos de los ítems se administraron vía reporte del cuidador, lo que podría introducir sesgos, ya que los cuidadores del grupo de tratamiento podrían: a) tener más conocimiento sobre el desarrollo infantil y ser más conscientes de sus logros (sesgo de observación) y/o b) tener una mayor tendencia a sobreestimar estos logros para dar cuenta del éxito de la intervención (sesgo de conveniencia). Las evaluaciones de Perú y Colombia tuvieron tasas de desgaste de muestra similares entre las mediciones de línea de base y de seguimiento.

⁹ La comparabilidad de la magnitud de los efectos entre estudios es limitada no solo porque cada uno de ellos usa diferentes instrumentos para medir los resultados, sino también porque estos resultados se escalan con la DE de cada muestra. Si se hubiera utilizado la misma medida de resultado para las tres evaluaciones, pero dicha medida no tuviera una norma o escala comparable, cualquier comparación de la magnitud de los efectos seguiría siendo limitada debido a las diferencias en la distribución de habilidades (desarrollo) entre las poblaciones (y, por lo tanto, a las diferencias en sus varianzas). Si se hubiera utilizado la misma medida y si esta medida tuviera una muestra normativa, incluso si fuera de una población externa, las comparaciones relativas serían más limpias, aunque todavía podrían verse afectadas por las diferencias entre las distribuciones de la muestra de estudio y de la muestra normativa. El uso de una medida común, globalmente normalizada y culturalmente neutral, como la versión larga de la Escala Global para el Desarrollo Temprano (*Global Scale for Early Development*, GSED; Cavallera et al., 2019), permitiría realizar comparaciones significativas de la magnitud de los efectos entre las evaluaciones. Sin embargo, el desarrollo del GSED o de medidas globales similares para la evaluación de impacto de programas aún está en desarrollo.

3. Los riesgos de escalar intervenciones

Esta sección utiliza el modelo económico de escalamiento de Al-Ubaydli et al. (disponible próximamente), a su vez basado en Al-Ubaydli et al. (2019), para analizar la intervención de acompañamiento a familias en Jamaica y los esfuerzos posteriores para adaptarla y ampliar su escala de implementación en Colombia y Perú. Discute si los riesgos de llevar a escala una intervención propuestos en el modelo estaban presentes en cada uno de estos esfuerzos y reflexiona sobre cómo podrían haber afectado los resultados. Lo hace, en primer lugar, considerando las diferencias entre Jamaica y Colombia ("de 70 a 700"), suplementando el análisis con información de otros pilotos similares en Bangladesh, India y China. Luego, se discuten las diferencias entre Colombia y Perú ("de 700 a 70.000").

3.1 De 70 a 700: Jamaica y Colombia

Como se muestra en la Tabla 3, los efectos a corto plazo del estudio de eficacia en Jamaica (0,88 DE) son sustancialmente mayores en magnitud que los estimados para el piloto en Colombia (0,18 DE). Esto implica que, al pasar "de 70 a 700", el impacto se redujo alrededor de un 80%. ¿Qué podría explicar una diferencia tan dramática?

Problemas de inferencia estadística

Una posibilidad que se analiza en el modelo de Al-Ubaydli et al. (disponible próximamente) es la de los problemas de inferencia estadística. Podrían surgir dudas en torno a la inferencia estadística si el estudio de Jamaica reflejara simplemente un "golpe de suerte" de entre todas las posibles muestras extraídas de una población determinada –específicamente, que se obtuviera un impacto en la cola derecha de la distribución de todos los posibles impactos. Sin embargo, no parece haber sido así. Además del experimento más conocido de Jamaica, que tuvo un impacto de 0,88 DE, hubo al menos otras tres réplicas evaluadas a través de RCTs en Jamaica, todas listadas en la Tabla 4. Los impactos en dos de estos experimentos fueron de mayor tamaño (0,94 y 1,18 DE, respectivamente) y solo en uno de ellos el impacto fue mucho menor (0,28 DE) y no significativo. Bajo esta perspectiva, no parece haber nada inusual en el tamaño de los resultados del "estudio de Jamaica".¹⁰

Dicho esto, sin embargo, sí hay algo inusual en los resultados de Jamaica, no solo en relación con los resultados de Colombia, sino también con los de las réplicas en (a) Bangladesh, donde se han llevado a cabo tres estudios de eficacia con impactos de entre 0,25 y 0,38 DE en el índice de desarrollo mental de las escalas de Bayley (segunda edición); (b) India, donde un experimento reciente tuvo un impacto de 0,21 DE en un índice que combina el desarrollo cognitivo, el lenguaje receptivo, el lenguaje expresivo y la motricidad fina evaluado por las escalas de Bayley (tercera edición); y (c) China, donde una intervención piloto vagamente

¹⁰ Ponemos el término entre comillas porque el experimento de Jamaica más conocido y que describimos aquí no fue el único experimento de la intervención en Jamaica. Ni tan siquiera fue el primero, por lo que tampoco podemos describirlo como el experimento de Jamaica "original". Es simplemente el más conocido, en gran parte, debido al seguimiento de largo plazo.

basada en el modelo de Jamaica encontró impactos de 0,17 DE en un índice de las escalas de Bayley que combina el desarrollo cognitivo, el lenguaje receptivo, el lenguaje expresivo, la motricidad fina, la motricidad gruesa y el desarrollo socioemocional.¹¹

Esto se ve claramente en la Tabla 4 y la Figura 2. La figura muestra la relación entre la magnitud del efecto promedio para cada evaluación (en el eje vertical) y el logaritmo de la población cubierta por el programa (en el eje horizontal). El color de los círculos varía según el país y un círculo vacío indica que el impacto determinado no es estadísticamente significativo. Como se observa en la figura, la magnitud del efecto promedio disminuye a medida que aumenta la escala, o el número de niños que reciben la intervención. Relacionado con lo anterior, los impactos encontrados en Jamaica (primeras evaluaciones), aunque variables, son en promedio sustancialmente mayores que los observados en otros lugares.

Por lo tanto, la Figura 2 plantea varias explicaciones posibles para la disminución de la magnitud de los impactos a medida que el programa de Jamaica fue escalado gradualmente. Una de las opciones es un problema específico de inferencia estadística: el instrumento utilizado en las evaluaciones de Jamaica –o algunos de sus elementos– puede haber sido más sensible a los tipos de cambios en el desarrollo infantil producidos por las visitas. Las pruebas de diagnóstico del desarrollo, como las Escalas de Griffiths o de Bayley, comparten ítems similares, ya que todas buscan medir el mismo constructo –el desarrollo infantil global. Sin embargo, el énfasis relativo en cada dominio del desarrollo, así como los ítems específicos que se incluyen y su redacción, sí varían de una prueba a otra. A futuro, un análisis que estudie cuidadosamente las puntuaciones de rendimiento a nivel individual en cada evaluación sería útil para comprender si esto puede explicar, en parte, las diferencias en el tamaño de los efectos entre Jamaica y los demás lugares.

Representatividad de la población

Otra cuestión planteada en Al-Ubaydli et al. (2019), es la representatividad de la población y, de manera más general, si los participantes del programa fueron similares en todas las intervenciones. En el estudio de Jamaica, los participantes procedían de un censo que identificó a niños de 9 a 24 meses de edad que sufrían desnutrición crónica, con un peso al nacer superior a 1,8 kg, sin hermanos gemelos, sin una discapacidad detectable, con madres con bajos niveles educativos y que habitaban en viviendas precarias de los barrios marginales de Kingston. Es posible que el hecho de que los niños en Jamaica fueran, en promedio, más pobres que los de Colombia explique, en parte, que la magnitud de los impactos fuera menor en Colombia. Sin embargo, las familias que se beneficiaron de la intervención en India y Bangladesh eran, en todo caso, aún más desfavorecidas que las de Jamaica, por lo que ésta no parece ser una explicación suficiente para los impactos menores obtenidos en las réplicas del modelo de Jamaica en otros lugares.

¹¹ Para el análisis de un programa en China que refleja más de cerca el modelo de *Reach Up*, ver Heckman, J., Liu, B. y Zhou, J. (disponible próximamente).

Otra explicación, quizás, es una posible selección de la población incluida en el estudio en base a los beneficios esperados. En evaluaciones posteriores –en particular las de Colombia y Perú– se cuantifica cuidadosamente y se reporta el número de familias que tuvo la opción de recibir visitas domiciliarias y la rechazó. Las estimaciones de impacto que se reportan son de intención de tratamiento (*Intent-to-Treat* o ITT por sus siglas en inglés), lo que significa que incluyen a todas las familias que se asignaron aleatoriamente al grupo de tratamiento, independientemente de que efectivamente recibieran tratamiento. Este no es el caso en los estudios de Jamaica. En ninguno de ellos se informa si hubo familias que rechazaron las visitas y, en caso afirmativo, si se seleccionaron familias de reemplazo de algún modo. Parece probable que, en los tres países, las familias que esperaran beneficiarse más del programa fueran más propensas a participar. Si hubo familias en Jamaica que rechazaron las visitas, la selección en base a las ganancias esperadas podría ser parte de la explicación de que se obtuvieran resultados mayores allí que en otros lugares, ya que las familias que potencialmente rechazaron la intervención no fueron incluidas en el cálculo de los impactos de la intervención.

Representatividad de la situación

A continuación, pasamos a una discusión sobre la representatividad de la situación. El proyecto de Jamaica no pretendía ser representativo de una situación del mundo real, sino más bien una prueba de concepto. El proyecto piloto de Colombia fue un intento de realizar visitas de manera escalable –es decir, de forma que se pudieran replicar a escala. Varios desafíos durante la implementación podrían haber comprometido la fidelidad al escalar el modelo jamaicano en Colombia.

Los datos presentados en las Tablas 1 y 2 muestran diferencias en la naturaleza y calidad de la supervisión que recibieron las facilitadoras. Mientras que en Jamaica sólo había tres facilitadoras y dos supervisoras, en Colombia participaron 144 facilitadoras (164, teniendo en cuenta la rotación del personal), y seis supervisoras. En Colombia no sólo fue mayor el número de personal involucrado, sino también la tasa de facilitadoras por supervisora. La tasa de facilitadoras por supervisora fue de 1,5 en Jamaica, en comparación con 24 en Colombia. Además, en Jamaica las supervisoras realizaron sesiones de acompañamiento semanales mientras que éstas ocurrieron cada 7-10 semanas en Colombia. En Jamaica, las investigadoras del proyecto también fueron sus supervisoras con lo que sus calificaciones y conocimientos especializados acerca de la implementación de la intervención eran sustancialmente mejores que los de las supervisoras en Colombia, quienes eran graduadas universitarias con diversos tipos de experiencia previa. La participación directa de las investigadoras en la implementación de la intervención y su supervisión en Jamaica también puede ser parte de la explicación detrás de los impactos de mayor tamaño. Las investigadoras que participaron en el estudio habían elaborado el currículo y la metodología de las visitas domiciliarias, pensando cuidadosamente en la forma de asegurar que fueran apropiadas al entorno donde se aplicaban. También sabían exactamente cómo debían realizarse las visitas domiciliarias. Es probable que en las demás réplicas –incluidas las de Bangladesh, Colombia, India y China– todas estas condiciones se hayan reproducido en menor grado que en el “estudio de Jamaica”.

En resumen, hay varias explicaciones posibles para la gran disminución en la magnitud de los impactos obtenidos al pasar "de 70 a 700". En primer lugar, hubo diferencias en los instrumentos utilizados para medir el desarrollo infantil, y algunas de las diferencias observadas en la magnitud de los impactos pueden ser simplemente un artefacto estadístico. En segundo lugar, puede haber diferencias en la población que recibió visitas domiciliarias en los distintos marcos de estudio. No creemos probable que las diferencias en el estatus socioeconómico o la salud infantil sean la razón principal detrás de los impactos de mayor tamaño en Jamaica ya que los niños atendidos en otros marcos de estudio, como Bangladesh o India, provenían de entornos igualmente vulnerables sino más. La posible selección (no documentada) en base a los beneficios esperados, quizás más en Jamaica que en otros lugares, podría ser parte de la explicación. En términos más generales, podría ser que haya algo inherente al modelo, su diseño o su forma de implementación, que sea más apropiado para Jamaica (o para las brechas de los niños en Jamaica), y que esto sea parte de la explicación. En tercer lugar, hubo diferencias subyacentes en cómo se implementó el programa en diferentes lugares. Parte de esto tal vez no sea sorprendente: las evaluaciones de Jamaica fueron 'pruebas de concepto', mientras que la intervención en Colombia buscó implementar y evaluar una intervención que, al menos en principio y por diseño, era escalable. La tasa de facilitadoras por supervisora fue mayor en Colombia que en Jamaica, y además, en Colombia, la frecuencia de la supervisión y la participación de los investigadores en la implementación de la intervención fueron menores.

3.2 De 700 a 70.000: Colombia y Perú

A continuación, comparamos los programas de Colombia y Perú destacando si pueden haber existido riesgos de ampliar la escala adicionales durante el rápido aumento de cobertura del programa en Perú, en donde el modelo fue implementado a escala nacional.

El programa Cuna Más en Perú fue una intervención a gran escala, en condiciones reales para este tipo de intervenciones, dirigida por un Ministerio recién creado. Durante la campaña electoral de 2011, el candidato Ollanta Humala anunció que, si salía elegido, empezaría un programa para la primera infancia. Después de su elección, Humala creó un ministerio nuevo con un programa insignia: Cuna Más. Este Ministerio contaba con personal altamente técnico que se valió activamente de la experiencia internacional (incluida la de los autores de este documento y la de los investigadores involucrados en las evaluaciones de Jamaica y Colombia) para diseñar un programa apropiado para las zonas rurales de Perú. Las autoridades tenían cinco años para mostrar resultados, la duración del mandato presidencial.

Los programas de visitas domiciliarias requieren de mucho personal, siendo su selección y las tasas consideraciones clave para la calidad, los costos y el proceso de escalar las intervenciones. En Perú no había suficientes personas en las comunidades participantes que cumplieran con las calificaciones técnicas requeridas originalmente para ser facilitadoras. Como consecuencia, Cuna Más tuvo que adecuar estos requisitos, lo que supuso la contratación de facilitadoras con niveles educativos inferiores a los inicialmente deseados. Esto pudo afectar negativamente la calidad, y ciertamente retrasó el proceso de adaptación de la intervención y sus lineamientos, así como la selección de las facilitadoras. Asimismo, no

todas las supervisoras en Perú tenían educación terciaria, como se requería inicialmente y como fue el caso en Colombia.

Es importante señalar que, a pesar de su escala, el programa peruano mantuvo proporciones relativamente bajas de familias por facilitadora –10, en promedio– y de facilitadoras por supervisora –también 10, en promedio. Esta fue una decisión deliberada que tomaron los líderes de Cuna Más, en un esfuerzo por proteger la calidad y operar en localidades dispersas y mal conectadas. Ambas proporciones fueron sustancialmente menores que las de Colombia.

La implementación a escala de los programas de visitas domiciliarias también requiere repensar las estrategias de capacitación inicial y de formación continua (acompañamiento o mentoría). La capacitación inicial (pre-servicio) fue mucho más corta en Perú que en Colombia: en Colombia, las facilitadoras tuvieron dos semanas de capacitación previa al servicio, incluyendo prácticas de las visitas, además de otra semana de formación uno o dos meses después de empezar el programa. En Perú, en cambio, las facilitadoras tuvieron cuatro días de capacitación previa al servicio, sin prácticas. En cuanto a las supervisoras, en Colombia recibieron seis semanas de formación, comparado con solo nueve días en Perú. Además, es posible que hubiera una mayor pérdida de fidelidad en la capacitación en Perú debido a que el número de personas que se tuvieron que formar (y volver a formar por la alta rotación de personal) en un período de tiempo más corto fue mucho mayor.¹²

Tanto en Colombia como en Perú (así como en Jamaica), la formación de las facilitadoras durante la prestación del servicio –el acompañamiento continuo– se realizó en forma de encuentros individuales con la supervisora, que actuó como mentora. Cabe destacar que la frecuencia de estos encuentros era mayor en Perú –una reunión cada dos semanas– que en Colombia –una reunión planificada cada seis semanas, pero que en realidad se produjo cada 7-10 semanas. En Colombia, se implementaron otras soluciones de bajo costo para compensar el contacto menos frecuente en persona entre las facilitadoras y las supervisoras, incluyendo boletines con contenido importante a incluir durante las sesiones de acompañamiento y recordatorios semanales a las facilitadoras a través de mensajes de texto.

Por último, los gobiernos a menudo se encuentran con restricciones administrativas durante la implementación (requisitos para las compras públicas, procesos de planeación presupuestal) y deben tener en cuenta consideraciones políticas que influyen tanto en el contenido como en el cronograma de implementación de la intervención –que a menudo es más rápido de lo técnicamente deseable. Todas estas cuestiones estuvieron presentes en cierta medida en la implementación de Cuna Más.

¹² En ambos países, la capacitación siguió un modelo en cascada, donde cada nivel formó al siguiente nivel en la escala de implementación. En Colombia, dos psicólogos, que también apoyaron la adaptación del programa, formaron a las 6 supervisoras, quienes a su vez formaron a las facilitadoras en grupos de cuatro (el total de facilitadoras que había en cada localidad más una sustituta). En Perú, el personal de la oficina central en Lima formó aproximadamente a 1,000 supervisoras en tandas, quienes a su vez formaron a sus facilitadoras (10 facilitadoras cada una). La capacitación en cascada corre el riesgo de perder fidelidad a medida que la formación desciende por la cascada, particularmente si el número de personas a formar es grande, ya que esto generalmente conduce a mayores grupos de capacitación y a un tiempo de formación más corto.

Inicialmente, el personal de Cuna Más tenía poca experiencia en los procesos de compras públicas. El programa promovió principalmente en la producción de juguetes y manuales, pero calculó mal el tiempo y los costes de transacción asociados con su adquisición y distribución en las localidades rurales. En consecuencia, la intervención se inició antes de que se entregaran los materiales en todas las comunidades, lo que probablemente afectó a la calidad de las visitas domiciliarias.

Además, aunque Cuna Más era un programa público, muchas tareas administrativas clave a nivel local dependían del trabajo voluntario de miembros de la comunidad. Esto provocó ineficiencias y retrasos en los pagos y la disponibilidad de recursos, y afectó la motivación del personal. Por último, las presiones políticas dieron lugar a objetivos de expansión muy rápidos (a menudo a costa de la calidad) o a decisiones de asignación de recursos que no eran óptimas (por ejemplo, priorizar la distribución de los uniformes del personal sobre la de los manuales y los juguetes).

Así, el contexto en el que se implementó el programa en Perú fue muy diferente al de Colombia. En Colombia, lo que se implementó fue una versión (posiblemente escalable) del programa original de Jamaica. En Perú, por otro lado, la implementación del programa fue una política pública real –una política fundamentada en el resultado de investigaciones anteriores, pero, aun así, una política pública. Las diferencias de escala fueron enormes: en el momento en que se realizaron las evaluaciones de los dos programas, el número de familias atendidas en Perú era más de 100 veces mayor que en Colombia.

¿Qué es lo que explica entonces las distintas magnitudes de los impactos de la intervención entre Colombia y Perú? Una variedad de factores, entre ellos las diferencias en los instrumentos utilizados para la evaluación del desarrollo, en la población que recibió la intervención y en la implementación, incluida la estrategia de capacitación. Todos estos factores pueden ser parte de la explicación. Sin embargo, la mayor diferencia entre las muestras de evaluación en Colombia y Perú fue posiblemente en el número efectivo de visitas recibidas por los niños tratados.

Las diferencias en el número de visitas surgieron por dos motivos principales. En primer lugar, la participación en el programa entre los hogares asignados aleatoriamente al grupo de tratamiento fue mayor en Colombia (97%) que en Perú (66%).¹³ En segundo lugar, entre los

¹³ En Colombia, se logró una participación muy alta porque el equipo de investigación hizo un gran esfuerzo para garantizar que todos los hogares del grupo de tratamiento recibieran visitas. En este sentido, la evaluación en Colombia es algo parecida a un experimento de laboratorio –con una estrategia de intervención que podría ser replicada a escala– donde todas las personas asignadas a un grupo de tratamiento reciben el tratamiento en la práctica. En Perú, varios factores explican la menor participación entre la muestra de evaluación. Algunos de estos factores son específicos de la muestra de evaluación, otros están relacionados con la participación en el programa (adopción) de manera más amplia. El marco de la muestra para la evaluación de Cuna Más tenía como objetivo excluir los municipios donde había conflictos sociales, así como aquellos en los que Cuna Más operaba servicios de cuidado infantil. Sin embargo, debido a errores en los datos administrativos, hubo algunas comunidades en la muestra de tratamiento que en realidad sí tenían conflictos sociales o servicios de cuidado infantil, un hecho que solo se hizo evidente después de la asignación aleatoria de municipios a los grupos de tratamiento y control y de la recopilación de datos en la línea de base. De manera similar, debido a las demoras en la implementación del programa después de la selección de la muestra, algunos niños en las comunidades de tratamiento eran mayores que la edad límite establecida para la inscripción al programa (24 meses) cuando éste se implementó. En ambos casos se mantuvo a estos niños en la muestra de evaluación –tanto los que estaban en

que recibieron al menos una visita, el número de visitas recibidas fue mayor en Colombia –81% de las visitas planificadas– que en Perú –50% de las visitas planificadas. En otras palabras, las visitas efectivas fueron el 79% en Colombia y el 32% en Perú. Por lo tanto, tal vez no sea sorprendente que la magnitud del impacto estimado en Perú sea sustancialmente menor al de Colombia. De hecho, la razón de las magnitudes del impacto de las intervenciones (0,18 DE en Colombia; 0,10 DE para la estimación del ITT en Perú) es menor que la razón de las visitas efectivas, lo que sugiere que, en todo caso, el impacto de un número dado de visitas fue mayor en Perú que en Colombia.

En resumen, a medida que el programa de visitas domiciliarias pasó de ser un estudio de eficacia en Jamaica, a un piloto en Colombia, a una intervención a escala en Perú, la magnitud de los impactos disminuyó sustancialmente. Probablemente esta disminución se debe a una variedad de razones, y no podemos precisar de manera concluyente la importancia relativa de los diferentes factores. Desde nuestro punto de vista, lo más importante son las posibles disminuciones de la calidad de la implementación manifestadas en formas que son difíciles de medir y cuantificar. Es más complicado garantizar que un programa se implemente exactamente según lo previsto cuando el número total de niños que reciben la intervención es 720 y el número total de facilitadoras es 164, como en Colombia, que cuando el número de niños es 64 y el número de facilitadoras es 3, como en Jamaica. Y es órdenes de magnitud más difícil todavía preservar la fidelidad y asegurar un número de visitas adecuado cuando el número de niños se acerca a 70.000 y el número de facilitadoras a 7.000, como en Perú.

4. Lecciones aprendidas y preguntas adicionales

Este análisis comparativo del proceso de escalar gradualmente el modelo de visitas domiciliarias de Jamaica, que pasó de ser un estudio de eficacia a un programa piloto en Colombia y a un programa nacional en Perú, nos permite extraer algunas lecciones relacionadas con el cómo llevar a escala este tipo de intervención. Nos centramos en dos tipos de lecciones aprendidas: (a) mitigar los riesgos de escalar una intervención examinados anteriormente e (b) investigación y evaluación.

4.1 Lecciones aprendidas: Mitigando los riesgos de escalar

Lección 1: Algunos componentes y atributos clave del proyecto no son negociables.

Los componentes y atributos no negociables del proyecto son los cimientos de cómo se busca cambiar el comportamiento de los padres mediante la intervención. Por lo tanto, cualquier

áreas donde el programa no tenía la intención de operar como los que sobrepasaban la edad de inscripción—para evitar sesgos. No obstante, ninguno de ellos recibió visitas. En conjunto, estos niños representan el 19% de los niños del grupo de tratamiento. Otro 15% de los niños en el grupo de tratamiento no recibió visitas porque el programa no pudo contratar facilitadoras en sus comunidades o porque sus familias rechazaron participar en el programa, un valor que es notablemente superior al de la evaluación en Colombia, que fue del 3%.

estrategia de implementación debe mantenerlos, incluso cuando se adapta a diferentes entornos. Destacamos dos.

Capacitación y mentoría. Las facilitadoras necesitan una formación adecuada antes y durante el servicio, así como orientación y supervisión periódica. La capacitación debe ser principalmente práctica, más que teórica, y la orientación y supervisión deben centrarse en cosas concretas que la facilitadora hace (o no hace) durante la visita. El perfil y la experiencia previa de las facilitadoras no son tan relevantes, siempre y cuando cuenten con el apoyo que necesitan. La actitud y motivación de las facilitadoras y las supervisoras son clave. Muchos programas experimentan una gran rotación de personal, lo que generalmente es perjudicial y reduce la calidad de la implementación. Los programas deben poner énfasis en retener al personal eficaz.

Contenidos. La intervención debe tener un contenido claro y estructurado que guíe a la facilitadora durante la visita. Cuanto menor sea el nivel educativo promedio de las facilitadoras y mayor sea la rotación, mayor será la necesidad de estructura y andamiaje para guiar cada visita. El programa también necesita tener los recursos adecuados para desarrollar actividades y juguetes que sean apropiados para el nivel de desarrollo de cada niño y que promuevan interacciones positivas. El trabajo central de la facilitadora se hace a través de la madre o cuidador y la demostración es clave durante la visita. Los juguetes deben permanecer en las casas para que las familias puedan continuar con las actividades durante la semana.

Lección 2: El apoyo político es indispensable. La participación y la apropiación del gobierno son condiciones necesarias para la implementación a escala. Es beneficioso involucrar a los funcionarios del gobierno en las primeras etapas del proceso para obtener su aceptación e identificar las limitaciones administrativas y políticas que probablemente surjan a medida que el programa se expande a escala. Si bien se deben conservar los componentes clave del proyecto, el objetivo no debe ser diseñar una intervención ideal, sino una que un gobierno pueda y quiera implementar a escala.

Construir un programa en la vida real implica negociar *trade-offs* –esto es, hacer ciertas concesiones. Una decisión (*trade-off*) importante e inevitable es escoger entre escalar rápidamente o hacerlo con calidad. La ventana de oportunidad del apoyo político puede ser estrecha y puede requerir escalar un programa a un ritmo más rápido de lo que sería deseable desde un punto de vista puramente técnico y, por lo tanto, resulta en un sacrificio en términos de calidad y fidelidad. Una cuestión importante es preguntarse hasta qué punto es factible hacer correcciones a un programa que se ha escalado muy rápidamente.

Otro *trade-off* es elegir entre un programa que se centre estrechamente en el desarrollo infantil a través del juego y un programa más amplio –por ejemplo, uno que también incluya aspectos de salud y nutrición infantil, higiene o violencia intrafamiliar. Los programas pueden convertirse en víctimas de su propio éxito: un programa que es eficaz cuando se centra en el desarrollo infantil puede volverse ineficaz cuando cubre demasiados aspectos del bienestar del niño y del hogar, aunque todos estos aspectos sean importantes por sí mismos. Es fundamental evitar

sobrecargar las intervenciones, sobre todo teniendo en cuenta el perfil de las facilitadoras (por lo general, con baja calificación técnica), la duración de la visita (una hora a lo sumo) y los tiempos disponibles para capacitación.

4.2 Lecciones aprendidas: Investigación y evaluación

Lección 1: El proceso de escalar una intervención como parte del plan de investigación.

Un proceso de investigación diseñado para lograr escalar una intervención con calidad incluiría fases de estudio de eficacia, de réplicas rápidas e implementación a escala que sean flexibles y que permitan que la intervención se ajuste a los hallazgos de las etapas anteriores. Los elementos de un marco de mejora continua de la calidad (actuar-planificar-hacer-estudiar-actuar...) pueden incorporarse a este razonamiento –ver Davis et. al (disponible próximamente), Stuart (disponible próximamente) y Chambers y Norton (disponible próximamente) para obtener más información sobre el diseño de estudios de investigación que tienen en mente entender y fortalecer el proceso de escalar una intervención.

Los estudios de eficacia son útiles para identificar los componentes no negociables de una intervención y para comprender los mecanismos mediante los cuales se logran los resultados deseados. También pueden permitir experimentar con la flexibilización de ciertos aspectos de la intervención, especialmente los que son más costosos de aplicar a escala, como la capacitación y la formación continua. Para pasar a la implementación a escala es necesario que estos aspectos se puedan replicar rápidamente en contextos diferentes. En las fases de estudio de eficacia y de prueba piloto también es fundamental hacer un seguimiento de las familias que abandonan el programa y de las que deciden no participar y comprender las razones que motivan estas decisiones.

Lección 2: Poder estadístico. Los estudios deben diseñarse con suficiente poder estadístico para identificar efectos heterogéneos en las dimensiones que se consideren relevantes. La heterogeneidad puede considerarse no solo en función de características del individuo, sino también de características del lugar—por ejemplo, estudios de múltiples lugares y diferentes entornos culturales y geográficos que puedan requerir adaptaciones de la intervención.

Lección 3: Métrica. El uso de una medida común para los resultados de desarrollo infantil clave de la intervención (tanto intermedios como finales) que esté normalizada (idealmente utilizando normas globales) facilitará la comparabilidad entre estudios.

Lección 4: Patrocinadores y publicación en revistas científicas. Los donantes que financian investigaciones y las revistas científicas pueden desempeñar un papel importante alineando los incentivos necesarios para promover investigaciones que informen sobre los esfuerzos de escalar intervenciones, apoyando la publicación de: a) estudios de réplicas; b) estudios de implementación; c) estudios con resultados nulos; d) estudios sobre métricas y mediciones; y e) seguimientos a medio y largo plazo, los cuales son necesarios para calcular adecuadamente los beneficios de las intervenciones.

5. Conclusión y cuestiones pendientes

Aunque los estudios examinados en el presente documento arrojan luz sobre el proceso de llevar una intervención a escala, todavía hay cuestiones importantes relacionadas con la intervención y los esfuerzos examinados que permanecen sin respuesta y que podrían reducir el coste de la implementación a escala. Entre ellas figuran las siguientes: a) a qué edad del niño se debe iniciar una intervención como la aquí analizada; b) la eficacia y rentabilidad de las variaciones en la duración, la intensidad y el modo de implementación (por ejemplo, visitas grupales o individuales), y la forma en que esta elección puede variar según la edad, la geografía y la cultura del niño; y (c) cuáles son los efectos indirectos del tratamiento (en la familia, en la comunidad). Responder a estas preguntas, además de aplicar las lecciones aprendidas del proceso de llevar a escala el programa de Jamaica y la teoría presentada por el modelo económico de escalamiento, ayudará a todas las partes interesadas a estar mejor preparadas para escalar sus propios programas.

Bibliografía

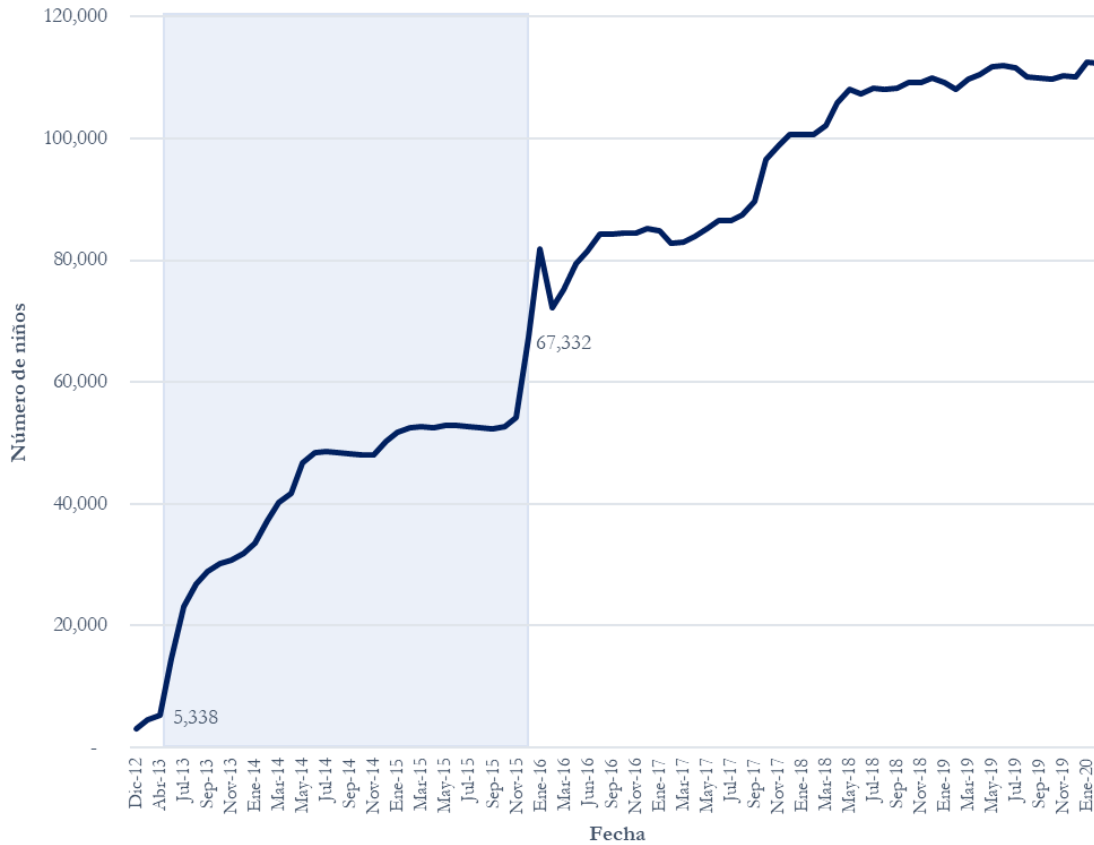
- Al-Ubaydli, O., List, J., & Suskind, D. (2019). The Science of Using Science: Towards an Understanding of the Threats to Scaling Experiments. *NBER Working Paper, 25848*. <https://doi.org/10.3386/w25848>
- Al-Ubaydli, O., Lee, M., List, J., & Suskind, D. (disponible próximamente). The Science of Using Science: A New Framework for Understanding the Threats to Scaling Evidence-Based Policies. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.
- Andrew, A., Attanasio, O., Augsberg, B., et al. (2019). Effects of a scalable home-visiting intervention on child development in slums of urban India: evidence from a randomised controlled trial. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. <https://doi.org/10.1111/jcpp.13171>
- Andrew, A., Attanasio, O., Fitzsimons, E., et al. (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. <https://doi.org/10.1371/journal.pmed.1002556>
- Araujo, M., Dormal, M., Grantham-McGregor, S., et al. (2019). *Home Visiting at Scale and Child Development*.
- Attanasio, O., Cattan, S., Fitzsimons, E., et al. (2020). Estimating the Production Function for Human Capital : Results from a Randomized Controlled Trial in Colombia. *American Economic Review, 110*(1), 48–85. <https://doi.org/https://doi.org/10.1257/aer.20150183>
- Attanasio, O., Fernandez, C., Fitzsimons, E., et al. (2014). Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial. *BMJ, 349*(sep29 5), g5785–g5785. <https://doi.org/10.1136/bmj.g5785>
- Bitler, M., Hoynes, H., & Domina, T. (2014). Experimental Evidence on the Distributional Effects of Head Start. *NBER Working Paper, 20434*.
- Cavallera, V., Black, M., Bromley, K., et al. (2019). The Global Scale for Early Development (GSED). *Early Childhood Matters, 80–84*.
- Chambers, D.&Norton, W. (disponible próximamente). Sustaining Impact after Scaling Using Data and Continuous Feedback. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.
- Davis, J., Guryan, K., Hallberg, K.& Ludving, J. (disponible próximamente). Studying Properties of the Population: Designing Studies that Mirror Real World Scenarios. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.
- Duncan, G., & Magnuson, K. (2013). Investing in preschool programs. *Journal of Economic Perspectives, 27*(2), 109–132. <https://doi.org/10.1257/jep.27.2.109>
- Gertler, P., Heckman, J., Pinto, R., et al. (2014). Labor market returns to an early childhood stimulation intervention in Jamaica. *Science, 344*(6187), 998–1001. <https://doi.org/10.1126/science.1251178>

- Grantham-McGregor, S., Powell, C., Walker, S., et al. (1991). Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: The Jamaican study. *The Lancet*, 338(8758), 1–5. [https://doi.org/10.1016/0140-6736\(91\)90001-6](https://doi.org/10.1016/0140-6736(91)90001-6)
- Grantham-McGregor, S., & Smith, J. (2016). Extending The Jamaican Early Childhood Development Intervention. *Journal of Applied Research on Children: Informing Policy for Children at Risk*, 7(2), Article 4.
- Hamadani, J., Huda, S., Khatun, F., et al. (2006). Psychosocial stimulation improves the development of undernourished children in rural Bangladesh. *The Journal of Nutrition*, 136(10), 2645–2652. <https://doi.org/10.1093/ajph/136/10/2645> [pii]
- Havnes, T., & Mogstad, M. (2015). Is universal child care leveling the playing field? *Journal of Public Economics*, 127, 100–114.
- Heckman, J., Liu, B. & Zhou, J. (disponible próximamente). The Economics of Investing in Child Development: Applying Home Visiting at Scale. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.
- Love, J., Chazan-Cohen, R., Raikes, H., et al. (2013). What Makes a Difference: Early Head Start Evaluation Findings in a Developmental Context. *Monographs of the Society for Research in Child Development*, 78(1).
- Luo, R., Emmers, D., Warrinnier, N., et al. (2019). Using community health workers to deliver a scalable integrated parenting program in rural China: A cluster-randomized controlled trial. *Social Science and Medicine*, 239(July), 112545. <https://doi.org/10.1016/j.socscimed.2019.112545>
- Nahar, B., Hossain, M., Hamadani, J., et al. (2012). Effects of a community-based approach of food and psychosocial stimulation on growth and development of severely malnourished children in Bangladesh: a randomised trial. *European Journal of Clinical Nutrition*, 66(6), 701–709. <https://doi.org/10.1038/ejcn.2012.13>
- Nores, M., Bernal, R., & Barnett, W. (2019). Center-based care for infants and toddlers: The aeioTU randomized trial. *Economics of Education Review*, 72(May), 30–43. <https://doi.org/10.1016/j.econedurev.2019.05.004>
- Olds, D. (2010). The Nurse-Family Partnership: From Trials to Practice. In A. Reynolds, A. Rolnick, M Englund, & J. Temple (Eds.), *Childhood Programs and Practices in the First Decade of Life* (pp. 49–75). <https://doi.org/10.1017/CBO9780511762666.004>
- Powell, C., Baker-Henningham, H., Walker, S., et al. (2004). Feasibility of integrating early stimulation into primary care for undernourished Jamaican children: cluster randomised controlled trial. *BMJ*, 329(7457), 89–91. <https://doi.org/10.1136/bmj.38132.503472.7C>
- Powell, C., & Grantham-McGregor, S. (1989). Home visiting of varying frequency and child development. *Pediatrics*, 84(1), 157–164.
- Stuart, A. (disponible próximamente). Accounting for Differences in Population: Predicting Intervention Impact at Scale. In J. List, D. Suskind, & L. Supplee (Eds), *The scale-up effect in early childhood & public policy: Why interventions lose impact at scale and what we can do about it*. Routledge.
- Tofail, F., Hamadani, J., Mehrin, F., et al. (2013). Psychosocial Stimulation Benefits Development in Nonanemic Children but Not in Anemic, Iron-deficient Children. *The Journal of Nutrition*, 143(6), 885–893. <https://doi.org/10.3945/jn.112.160473>.Three

- Walker, S., Chang, S., Smith, J., et al. (2018). The Reach Up Early Childhood Parenting Program. Origins, Content, and Implementation. *Zero to Three Journal*, 38(4), 37.
- Walker, S., Chang, S., Powell, C., et al. (2004). Psychosocial Intervention Improves the Development of Term Low-Birth-Weight Infants. *The Journal of Nutrition*, 134(6), 1417-1423. <https://doi.org/10.1093/jn/134.6.1417>
- Walker, S., Chang, S., Vera-Hernández, M., et al. (2011). Early childhood stimulation benefits adult competence and reduces violent behavior. *Pediatrics*, 127(5), 849–857. <https://doi.org/10.1542/peds.2010-2231>

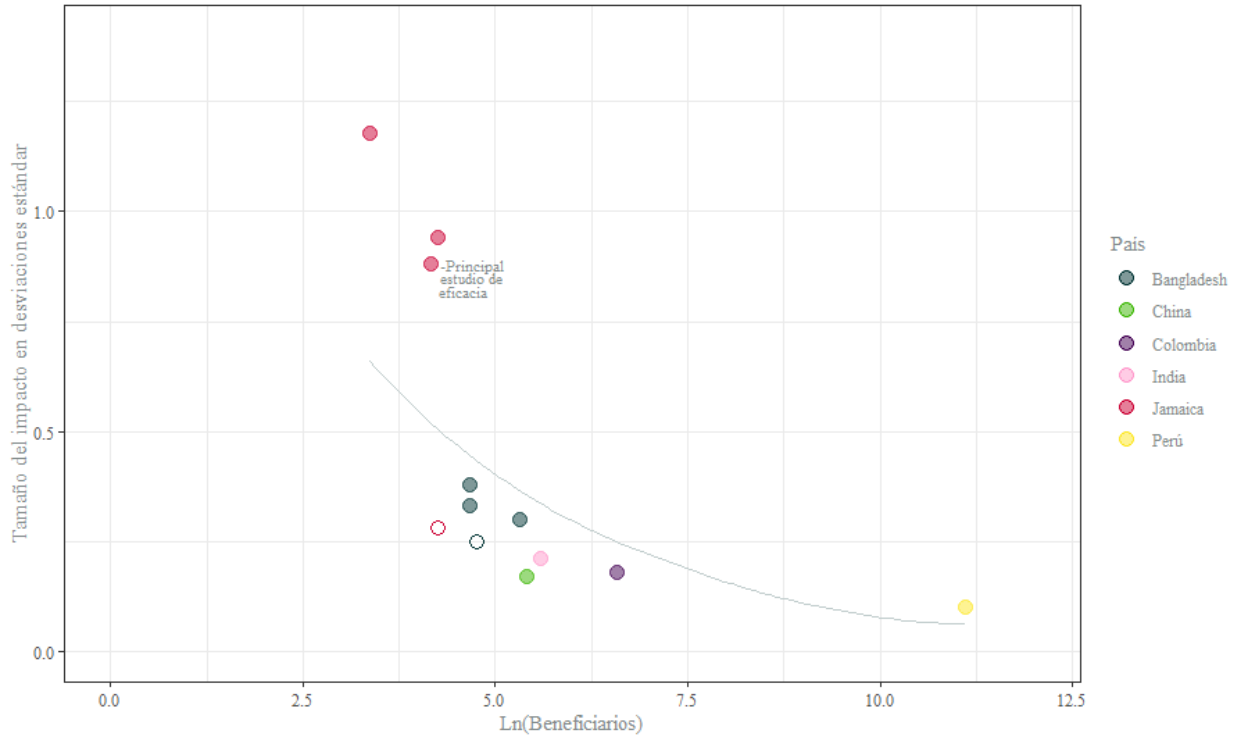
Figuras y tablas

Figura 1. Ampliación de cobertura de Cuna Más, 2012-2020



Fuente: Elaborado por los autores a partir de datos históricos de la cobertura del programa del Ministerio de Desarrollo e Inclusión Social – Perú, disponible en: <http://sdv.midis.gob.pe/Infomidis/#/>

Figura2. Tamaño del impacto de las diferentes réplicas del modelo de Jamaica



Fuente: Elaborado por los autores a partir de la Tabla 4.

Tabla 1. Diseño de los programas en Jamaica, Colombia y Perú

	Jamaica	Colombia	Perú
Intervención	Visitas domiciliarias		
Periodicidad y duración de las visitas previstas	Una hora, semanalmente		
Duración máxima (meses)	24	18	36
Edad en meses al inicio del programa	9-24	12-24	1-24
Edad en meses al final del programa	33-48	30-42	36
Duración de la capacitación de las facilitadoras	n. d.	2 semanas antes del servicio (incluyendo prácticas), 1 semana durante el servicio	4 días antes del servicio (sin prácticas)
Duración de la capacitación de las supervisoras	No aplica	6 semanas (incluyendo prácticas)	9 días (sin prácticas)
Selección de las facilitadoras	Seleccionadas por las investigadoras	Madres líderes. ¹ Respondieron a un test de comprensión lectora y fueron entrevistadas sobre su disponibilidad e interés por los investigadores	Candidatas nominadas por su comunidad y seleccionadas por las supervisoras
Características			
Perfil	Trabajadoras comunitarias de la salud	63% madres líderes ¹ 37% señoras recomendadas por madres líderes ¹	Requisito de alfabetización
Mujeres	100%	100%	85%
Escolarización (años)	n. d.	8-9	10
Área de procedencia	De Kingston	De las comunidades	De las comunidades
Remuneración de las facilitadoras (USD al mes)	n. d.	~ 40 (19,5% del salario mínimo concurrente)	100 durante la evaluación (36% del salario mínimo concurrente). Aumentó a 115 en 2017.
Selección de supervisoras	No aplica	A cargo de los investigadores	Convocatoria abierta regional
Perfil de las supervisoras			
Educación	Fueron las mismas investigadoras	Todas universitarias. Títulos en psicología o trabajo social o, en su ausencia, experiencia trabajando con familias	Requerimiento de educación terciaria. En la práctica, que al menos hubieran seguido con los estudios después de la educación secundaria
Mujeres	100%	100%	~72%
Área de procedencia		De Bogotá	De la comunidad o el municipio correspondiente

	Jamaica	Colombia	Perú
Remuneración de las supervisoras (USD al mes)	n. d.	~ 520-600, con incrementos cada 6 meses entre 2010 y 2011	830-900, con salarios más altos en las zonas más remotas
Periodicidad de la supervisión	Semanal	Cada seis semanas	Cada dos semanas
Familias por facilitadora	20-21	5	10
Facilitadoras por supervisoras	3/2	24	10
Criterio de focalización	Niños con desnutrición crónica en barrios marginales urbanos de Kingston	Niños entre 12-24 meses de familias beneficiarias del programa de transferencias condicionadas (que se focaliza en el 20% más bajo de la distribución de ingresos del país) en 96 localidades semi-urbanas (entre 2.000-42.000 habitantes)	Todos los niños entre 1-24 meses en pueblos rurales (hasta 2.000 habitantes o 400 viviendas) con índices de pobreza del 50% o más e índices de desnutrición crónica del 30% o más
Características de la línea de base y destinatarios del programa	n. d.		
Desnutrición crónica		14%	37%
Educación materna (años)		7-8	7
Casas con alcantarillado		90%	25%
Casas con suelo de tierra		6%	73%

Notas: n.d. información no disponible. ¹Las madres líderes son líderes locales elegidas por sus comunidades para hacer de enlace con los administradores del programa nacional de transferencias monetarias condicionadas.

Tabla 2. Estado de implementación del programa según la evaluación de seguimiento*

	Jamaica	Colombia	Perú
Agencia a cargo	Investigadoras	Investigadores	Ministerio de Desarrollo e Inclusión Social
Visitas domiciliarias efectivas (% de las visitas previstas)	~100%	81%	50%
Tasa efectiva: Familias por facilitadora	20-21	~5,4	9,5
Tasa efectiva: Facilitadoras por supervisora	3/2	~27,3	9
Frecuencia efectiva de supervisión	Semanal	Cada 7-10 semanas	Cada dos semanas
Meses de permanencia en el programa de las facilitadoras¹	24	15 (mediana 17,3)	17,1
Meses de permanencia en el programa de las supervisoras¹	24	18	18,6
Número total de niños atendidos	64	720	~67.332
Número de facilitadoras	3	Inicialmente 144; 164 al final de la implementación del piloto	~10.000
Número de supervisoras	2	6	~1.000
Participación	100%	97%	66% ²
Costos (USD por niño por año)	n. d.	~500	~300
Intervención descontinuada	Sí	Sí	No

Notas: n. d. información no disponible. *Datos cuantitativos de la encuesta de seguimiento. ¹Promedio de meses en el programa según la encuesta de seguimiento. ²Calculado con la muestra de la evaluación del impacto.

Tabla 3. Evaluación del programa en Jamaica, Colombia y Perú

	Jamaica	Colombia	Perú
Diseño de la investigación	RCT, aleatorizado a nivel de niño	CRCT, aleatorizado a nivel de localidad	CRCT, aleatorizado a nivel de municipio
Grupos de tratamiento	T1 = Visitas T2 = Suplementos alimenticios T3 = Ambos	T1 = Visitas T2 = Suplementos alimenticios T3 = Ambos	T1 = Visitas T2 = Visitas + reuniones grupales dos veces por semana <i>T2 no se implementó, todas las comunidades T2 recibieron T1</i>
Representatividad de la población	No diseñados para ser representativos de la población elegida		
Mediciones	Escalas de Desarrollo Mental de Griffiths, HOME, altura y peso	Escalas Bayley-III de Desarrollo Infantil, Indicadores de Cuidado Familiar (FCI), altura, peso y hemoglobina	Cuestionarios de Edades y Etapas (ASQ3), FCI, altura y peso
Grupo a cargo del diseño de evaluación	Investigadoras	Investigadores	Investigadores, con aportes de los equipos técnicos del Ministerio de Economía y Finanzas y del Ministerio de Desarrollo e Inclusión Social.
Grupo a cargo de la recolección de datos	Investigadoras	Empresa contratada especializada en encuestas, formada y supervisada por los investigadores	Instituto Nacional de Estadística e Informática de Perú (agencia gubernamental), contratado y supervisado por el Ministerio de Economía y Finanzas. Los investigadores participaron en la formación y la supervisión de los encuestadores.
Tamaño de la muestra (Tratamientos de visitas domiciliarias)	129 niños (32 en T1 + 32 en T3) 127 niños analizados (30 en T1 + 32 en T3)	1440 niños (360 en T1 + 360 en T3) 1263 niños analizados (318 en T1 + 319 en T3)	5.339 niños (3.530 en T1 + T2) 4.685 niños analizados (3.192 en T1 + T2)
Pérdida de muestra	0%	9,4% en el grupo de control 11,7% en T1 13,9% en T2 11,4% en T3	8,6% en el grupo de control 9,3% en T1 + T2
Impacto (en Desviaciones Estándar, DE)	<u>Intención de tratamiento</u> 0,88 DE en cociente de desarrollo (<i>Developmental Quotient, DQ</i>) ¹	<u>Intención de tratamiento (T1)</u> 0,26 DE en desarrollo cognitivo 0,22 DE en lenguaje receptivo 0,18 DE en un índice agregado (cognición, lenguaje receptivo y expresivo, y motricidad fina) T3: estadísticamente no significativos respecto a T1.	<u>Intención de tratamiento (T1 + T2)</u> 0,08 DE en desarrollo cognitivo 0,10 DE en lenguaje 0,10 DE en índice total ² <u>Tratamiento en los tratados (T1 + T2)</u> 0,15 DE en desarrollo cognitivo 0,15 DE en lenguaje 0,14 DE en índice total ²

Notas ¹El cociente de desarrollo (DQ) incluye las subescalas locomotoras (motricidad gruesa), coordinación de manos y ojos (motricidad fina), habla y escucha (lenguaje), y desempeño/rendimiento (cognición). ²El índice total es una suma de desarrollo cognitivo (resolución de problemas), lenguaje, motricidad fina, motricidad gruesa y desarrollo personal-social.

Tabla 4. Réplicas del modelo jamaicano y tamaños del impacto, en desviaciones estándar (DE)

País	Referencia	n (T)	Resultado	Impacto
Jamaica	Powell & Grantham-McGregor (1989)	58 (29)	Escala de Desarrollo Mental de Griffiths, primera edición; Cociente de desarrollo (<i>Developmental Quotient</i> , DQ)	1,18 DE
	Grantham-McGregor et al., (1991)	127 (30)		0,88 DE
	Walker et al., (2004)	130 (63)		0,28 DE (no significativo)
	Powell et al., (2004)	129 (65)		0,94 DE
Bangladesh	Hamadani et al., (2006)	193 (92)	Escala Bayley de Desarrollo Infantil, segunda edición; Índice de desarrollo mental (<i>Mental Development Index</i> , MDI)	0,33 DE
	Nahar et al., (2012)	322 (59)		0,30 DE
	Tofail et al., (2013)	Niños sin anemia: 196 (104)		0,38 DE
		Niños anémicos (deficiencia de hierro): 216 (110)		0,25 DE (no significativo)
Colombia	Attanasio et al., (2014)	1.263 (318)	Escala Bayley de Desarrollo Infantil, tercera edición; índice agregado de desarrollo cognitivo, lenguaje y motricidad fina	0,18 DE
Perú	Araujo et al., (2019)	4.685 (3.192)	Cuestionarios de Edades y Etapas (ASQ3); índice agregado de desarrollo cognitivo, lenguaje, motricidad fina, motricidad gruesa y desarrollo personal-social	0,10 DE
India	Andrew et al., (2019)	378 (191)	Escala Bayley de Desarrollo Infantil, tercera edición; índice agregado de desarrollo cognitivo, lenguaje receptivo y expresivo y motricidad fina	0,21 DE
China	Luo et al., (2019)	390 (190)	Escala Bayley de Desarrollo Infantil, tercera edición; índice agregado de desarrollo cognitivo, lenguaje receptivo y expresivo, motricidad fina, motricidad gruesa y desarrollo socioemocional	0,17 DE

Nota: Tabla basada en la Tabla 1 de Grantham-McGregor y Smith (2016). Añade los estudios publicados desde entonces y excluye aquellos sin evaluaciones aleatorizadas (evaluaciones experimentales) o en los que las visitas domiciliarias se combinaron con otra intervención. Reportamos la magnitud del impacto en los índices agregados de desarrollo, independientemente de si hay impactos en sus subescalas. En todos los casos, reportamos el tamaño de la muestra y la magnitud del impacto para el grupo de tratamiento que recibió solo visitas (comparado con el grupo de control). Se reporta el impacto ajustado por covariables de la línea de base. Convertimos la magnitud del impacto en DE usando la DE del grupo de control en la línea de seguimiento. Los detalles específicos de los estudios citados están disponibles vía petición a los autores.