**Evaluación de Impacto del Programa FCC
Plan de Análisis de Resultados**

# Introducción

La práctica de actividades deportivas por parte de niños y jóvenes como mecanismo eficaz para la promoción de la inclusión social es reconocida desde hace algunas décadas por distintas organizaciones gubernamentales y no gubernamentales y por la sociedad en general. Sin embargo, la evidencia rigurosa de un efecto causal de la realización de actividades deportivas sobre el desarrollo social y económico de las personas es relativamente escasa.

A lo fines de cubrir este vacío, CAF, Banco de Desarrollo de América Latina, está realizando un esfuerzo por promover evaluaciones experimentales del impacto de los programas de fútbol sobre el proceso de acumulación de capital humano en niños y jóvenes.

En este documento se presenta el plan de análisis de resultados de la evaluación de impacto de un programa de futbol para niños y jóvenes de entre 5 y 17 años de edad, implementado por la Fundación Fútbol con Corazón (FCC) en diversas ciudades de Colombia. La evaluación se basó en la asignación aleatoria a grupos de tratamiento y control de una muestra de 1.649 niños que manifestaron interés por entrar al programa. Los niños asignados al tratamiento tuvieron la oportunidad de participar en las actividades regulares de FCC entre marzo de 2015 y Julio de 2016, fecha a partir de la cual comenzó el levantamiento de información sobre distintas variables de resultado por medio de encuestas presenciales a los niños y sus padres[[1]](#footnote-1).

El documento describe las principales características del programa FCC y presenta el plan de análisis de los datos de la línea de seguimiento. Al momento de escribir este documento, se cuenta con dos informes cualitativos realizados por un consultor externo al equipo encargado de la evaluación cuantitativa, una guía metodológica del modo de funcionamiento de FCC, su modelo pedagógico, y un esquema de las variables e instrumentos de medición empleados. En la sección 2 se presentan las principales características de la intervención, en la sección 3 se describe el diseño metodológico de la evaluación, mientras que en la sección 4 se presenta el plan de análisis de los datos de la línea de seguimiento.

# Contexto: el programa FCC y su modelo pedagógico

La Fundación Fútbol con Corazón (FCC) es una organización privada sin fines de lucro, creada en 2010 en Barranquilla y que actualmente desarrolla sus actividades en varias ciudades de Colombia. Entre sus objetivos se incluyen la promoción del desarrollo de habilidades para la vida a través de la lúdica y el deporte; la promoción de la resolución pacífica de conflictos, generando una cultura de paz y mejorando la convivencia familiar y ciudadana; el fomento de la equidad de género y el empoderamiento de la mujer; y la permanencia en la educación formal y el uso adecuado del tiempo libre.

Para lograr estos objetivos, desarrolla un programa de actividades basadas en la práctica de fútbol dirigido a niños y adolescentes. Las actividades se desarrollan en dos sesiones de trabajo por semana de 120 minutos cada una, en las que los participantes reciben un taller de convivencia basado en la metodología lúdico-experiencial y un adiestramiento deportivo basado en técnicas de entrenamiento en fútbol. Asimismo, semanal o quincenalmente, los participantes asisten a encuentros deportivos y/o campeonatos con la metodología de Fútbol por la Paz. Participar en FCC no tiene ningún costo directo para los beneficiarios. Las sesiones incluyen una merienda para cada niño, por cada día de entrenamiento.

La selección de los beneficiarios se realiza a partir de convocatorias periódicas y abiertas a toda la comunidad, en las que se invita a participar todos aquellos interesados en formar parte de las actividades regulares de la institución. Como requisitos de elegibilidad se incluyen: (i) residir en las cercanías de alguna de las sedes de FCC, (ii) tener entre 5 y 17 años de edad y (iii) estar matriculado en la educación formal. Los seleccionados se incorporan al programa y se agrupan en cuatro categorías en función de su edad: Creadores (5 a 7 años), Exploradores (8 a 10 años), Constructores (11 a 13 años) y Promotores (14 a 17 años).

El modelo pedagógico de FCC se destaca por considerar la práctica del fútbol como la herramienta idónea para el logro de objetivos vinculados a la formación de habilidades para la vida, lo que lo distingue del enfoque de otras instituciones cuyos objetivos apuntan más al desarrollo de habilidades físicas y técnicas para el éxito deportivo. Esto lleva, entre otras cosas, a que la competitividad inherente a la actividad deportiva quede en un segundo plano y se priorice la generación de un entorno favorable a la formación de valores y capacidades para la vida. Estas diferencias de enfoque se manifiestan en cuatro elementos característicos del programa de la fundación FCC: el rol del asesor, el contenido programático, los encuentros de fútbol (torneos realizados a través de la metodología de fútbol por la paz) y la escuela para padres.

## El rol del asesor

Las clases son dirigidas por instructores, a los que se les denomina asesores. Su objetivo central consiste en enseñar, orientar y facilitar el proceso de aprendizaje en valores y habilidades para la vida. Las jornadas normales se dividen en tres tiempos: (i) calentamiento, preparación e indicaciones, (ii) juego, y (iii) conclusiones, consensos y acuerdos. Esto implica que los asesores deben tener una gama más amplia de atributos que los entrenadores tradicionales de fútbol, lo cual se puede evidenciar con su formación académica complementaria en ramas de psicología o carreras afines. Estos factores permiten alejar el elemento competitivo inherente a la práctica del fútbol que, de acuerdo a otros estudios, es uno de los aspectos que reduce las bondades que podría tener el fútbol de afectar positivamente algunas habilidades socioemocionales de los niños (CAF, 2017).

## Contenido programático

Estructuralmente el entrenamiento tiene una programación temática de 12 meses. Esta programación temática está agrupada en dos subprogramas, uno de talleres de convivencia y uno de entrenamiento deportivo, los cuales se trabajan de manera paralela. El programa de convivencia se desarrolla en cuatro módulos temáticos: formación en valores, habilidades para la convivencia, promoción y prevención de la salud, y conformación del proyecto de vida. El programa de entrenamiento deportivo consiste en tres módulos: capacidades físicas, capacidades técnicas y capacidades tácticas. Cada uno de estos módulos tiene objetivos específicos que se desarrollan por medio de una serie actividades que se llevan a cabo durante un período de un mes.

## Encuentros deportivos (torneos)

Los encuentros deportivos están basados en la metodología Fútbol por la Paz. Esta metodología posee varias reglas diferentes a las empleadas en un partido tradicional de fútbol. Mientras que en una competencia tradicional el equipo ganador es aquel que marca el mayor número de goles, en la metodología Fútbol para la Paz los partidos se ganan con corazones, los cuales se pueden ganar no solo anotando goles[[2]](#footnote-2), sino también demostrando una buena conducta durante la realización del juego. Los equipos están conformados entre 5 u 8 jugadores dependiendo del tamaño de la cancha de la sede y de la edad de los participantes. Los equipos tienen una composición mixta en lo que respecta al género de los jugadores: cada equipo debe estar conformado por al menos dos niñas y dos niños, y para que los goles de un equipo sean válidos, el primer gol debe ser marcado por una mujer. En la etapa del juego, el papel del asesor es secundario, limitándose solamente a solucionar conflictos en el caso de que los niños no puedan llegar a un mutuo acuerdo, o corrigiendo alguna acción en la que no se aplique correctamente las reglas de juego. Cabe destacar que no existe la figura del árbitro, puesto que precisamente se busca que sean los mismos niños quienes resuelvan las disputas que se puedan derivar del juego, ayudándoles de este modo a desarrollar su capacidad de resolver conflictos.

## Escuela para padres

De forma complementaria y voluntaria, los padres de los beneficiarios reciben una sesión de trabajo mensual llamada Escuela para padres, en la que se les forma en los mismos temas en los que son formados los participantes o sus hijos.

# El diseño de la evaluación

## Objetivo y metodología de la evaluación

El estudio busca determinar, a través de una evaluación experimental, la relevancia de la práctica de fútbol bajo la metodología de Fútbol para la Paz descrita antes, durante horas extra escolares, sobre diversas medidas de habilidades cognitivas y no cognitivas y bienestar general, en niños y adolescentes de entre 5 y 17 años de edad.

Además de la evaluación cuantitativa, que pretende medir los impactos de la intervención, se realizó un estudio cualitativo con el fin de indagar más sobre dimensiones de impacto del programa, así como mecanismos o canales a través de los cuales podrían operar esos impactos.

## Diseño de la evaluación cuantitativa

El experimento realizado consistió en crear un grupo de tratamiento seleccionado al azar de una población de niños elegibles que manifestaron interés de practicar fútbol en FCC. Ante el exceso de demanda por la entrada al programa FCC, se seleccionó el ingreso de los aspirantes de manera aleatoria. Aquellos niños elegibles que no resultaron seleccionados para participar en el programa fueron considerados parte del grupo de control. De esta manera, se obtuvo un grupo de control estadísticamente comparable con el grupo que recibe el tratamiento (programa FCC). La aleatorización fue realizada por el equipo de evaluación sobre un listado de niños interesados en participar provisto por FCC.

La medición se realizó a través de la aplicación de una encuesta de seguimiento en hogares desde septiembre de 2016, en la cual participó el niño/joven y su cuidador principal (padre o representante). Por razones logísticas no se realizó encuesta de línea base antes de comenzar el programa, pero dado que los grupos fueron seleccionados aleatoriamente, cualquier diferencia estadísticamente significativa que se observe entre los grupos podrá atribuirse al programa FCC.

Se evaluaron 1.649 niños del estudio (tanto los niños de tratamiento como los del grupo de control) después de un año de haber iniciado las prácticas de fútbol. Se prevé realizar levantamiento adicional de información entre 5 y 10 años después que finalicen las prácticas de fútbol, con el propósito de determinar los impactos de mediano y largo plazo de la intervención (especialmente en términos de éxito educativo y laboral), sin embargo, este último levantamiento de información no está previsto todavía por CAF.

## Muestra

El tamaño total inicial de la muestra fue de 1.649 niños, distribuidos en nueve localidades de Barranquilla. La cantidad de participantes en cada estrato (sede, categoría de edad y género), dependió de la cantidad de cupos que estuvieron disponibles al momento de la evaluación.

## Asignación al tratamiento

La asignación al tratamiento fue aleatoria y se estratificó por sede, sexo y categoría de edad[[3]](#footnote-3). La muestra inicialmente fue de 1.649 niños clasificados en grupos de control, tratamiento y lista de espera[[4]](#footnote-4).

La lista de espera sirvió para reemplazar a los niños asignados al tratamiento que dejaban de asistir con regularidad. Cuando el niño acumulaba más de tres inasistencias, se contactaba telefónicamente y/o se visitaba para verificar el interés o la posibilidad de que continuara su participación en FCC. En caso de declinar su participación, el representante debía completar un formulario en el cual renunciaba formalmente a participar en los entrenamientos y, por lo tanto, liberaba el espacio para la inclusión de otro niño. Si en esa sede, categoría y género, se contaba con lista de espera, la misma era utilizaba para reemplazar al niño que había declinado. Cabe destacar que el ingreso al tratamiento desde la lista de espera también se determinaba en forma aleatoria cuando el número de niños en la lista superaba al número de niños a reemplazar. Si, por el contrario, no había lista de espera disponible para reemplazar al niño, el equipo de FCC era el responsable de conseguir el doble de formularios que de cupos disponibles para que de esa forma se pudiera realizar una selección al azar. Por último, los niños en lista de espera que nunca fueron utilizados como reemplazos, se consideran para el análisis como parte del grupo de control. Si bien al momento de elaborar este informe no se cuenta con información sobre la fecha de incorporación a la intervención de los niños de la lista de espera, se evaluará la posibilidad de explotar esa información para indagar posibles impactos diferenciales en función de la exposición al tratamiento. En la Tabla 1 se muestra la asignación al tratamiento por sede.

**Tabla 1.-** Asignación al tratamiento.

|  |  |  |  |
| --- | --- | --- | --- |
| Sede | Control | Tratamiento | Total |
| Bosque | 136 | 198 | 334 |
| Juan Mina | 65 | 65 | 130 |
| Minisedes | 43 | 52 | 95 |
| Nueva Colombia | 89 | 153 | 242 |
| Paz | 165 | 181 | 346 |
| Playa | 30 | 31 | 61 |
| Siape | 26 | 38 | 64 |
| Soledad | 66 | 105 | 171 |
| Villa San Pedro | 81 | 125 | 206 |
| Total | 701 | 948 | 1,649 |

## Tamaño de la muestra y Efecto Mínimo Detectable (EMD)

De acuerdo a experiencias similares, el impacto que puede generar un programa de las características mencionadas es modesto. Por ejemplo, Felfe et al. (2011) encontraron que la participación de niños de entre 3 y 10 años en actividades deportivas tuvo un impacto de 0,13 desviaciones estándar tanto en habilidades cognitivas como socioemocionales.

La selección de los participantes del programa fue llevada a cabo mediante una aleatorización estratificada según sede, sexo y categoría de edad. El cálculo del EMD se realizó considerando este diseño y con el uso del software Optimal Design.

La cantidad de participantes encuestados en la línea de seguimiento fue de 1.490[[5]](#footnote-5). No obstante, se debe considerar que algunas variables no fueron recolectadas para todos los rangos de edad debido al diseño y alcance de los instrumentos de medición empleados: para las variables que son respondidas por niños de 8 años en adelante se cuenta con 1.362 observaciones, y de 12 en adelante con 833 casos.

La tasa de cumplimiento, o tasa efectiva de asistencia al programa, se define como el número de clases a las que un participante asiste, en relación al total de clases que debería haber asistido desde el día que entró a participar. Para el cálculo del EMD se asume un cumplimiento del 70%.

El número total de estratos fue de 54 con un tamaño promedio de 31. Para el cálculo del poder estadístico del experimento, la varianza de los tamaños de los efectos se fijó en 0, lo cual supone que el impacto del programa sería el mismo en todos los estratos. Además, se asume que la proporción de la varianza explicada por las variables de estratificación es del 10%. Finalmente, considerando parámetros convencionales como un poder del 80% y un nivel de significancia del 5%, se determinó que para las dimensiones medidas en la muestra total, el EMD es de 0,22 desvíos estándar[[6]](#footnote-6). Por su parte, para los grupos de niños de 8 y 12 años en adelante, el EMD es de 0,23 y 0,30, respectivamente.

## Resultados de interés / otros resultados

La selección de las dimensiones en las que se espera encontrar impactos significativos del programa se basó en los resultados del estudio cualitativo, en conversaciones del equipo de evaluación con especialistas de la institución implementadora sobre características de la intervención, y en los resultados de la (escasa) literatura empírica sobre programas de deportes para el desarrollo de habilidades en niños y jóvenes. El informe cualitativo, realizado por un consultor externo al equipo de evaluación, se nutrió de un trabajo de campo conformado por visitas y observación delas actividades del programa y entrevistas en profundidad con actores claves (profesionales involucrados en el trabajo de la institución, antiguos alumnos y sus familiares, y familiares de los niños que actualmente participan del programa). En función de esto se priorizó las dimensiones que se creían podían ser afectadas en el corto plazo.

Sobre la base de estos insumos, las dimensiones sobre las que se espera hallar un efecto son la disciplina, respeto y responsabilidad. Además, en niños que poseen problemas previos se espera hallar un impacto sobre autocontrol, autoestima y asertividad. No obstante, el estudio extiende el análisis para otras dimensiones socioemocionales, cognitivas, uso de tiempo, autoconcepto físico, percepciones sobre la equidad de género, entre otras.

En la Tabla 2 se detallan las dimensiones medidas, junto al instrumento empleado y el rango de edad en el que pudo ser aplicado[[7]](#footnote-7). La elaboración del instrumento se llevó a cabo en base a experiencias previas y con el asesoramiento de profesionales en psicología, con el propósito de ser adaptables al entorno de la intervención, adecuarse a la edad de la población objetivo y ser aplicables en el largo plazo.

**Tabla 2.** Principales variables e instrumentos de medición utilizados.

|  |  |  |  |
| --- | --- | --- | --- |
| Variables | Instrumento[[8]](#footnote-8) | Cuestionario | Edades |
| Actitud hacia el colegio | BASC | niños | 5 a 17 años |
| Relaciones interpersonales | BASC | niños | 5 a 17 años |
| Uso del tiempo libre | Elaboración propia | padres y niños | 5 a 17 años |
| Hábitos saludables | Elaboración propia | padres y niños | 5 a 17 años |
| Rol de modelo y redes de apoyo | Elaboración propia | niños | 5 a 17 años |
| Habilidades sociales | BASC | padres | 5 a 17 años |
| Liderazgo | BASC | padres | 5 a 17 años |
| Autoestima | BASC | niños | 8 a 17 años |
| Autoconcepto físico | AF-5 | niños | 8 a 17 años |
| Agresividad | AESC | niños | 8 a 17 años |
| Integración y entorno | Elaboración propia | niños | 8 a 17 años |
| Expectativas de logro | Elaboración propia | niños | 8 a 17 años |
| Autocontrol/Hiperactividad | BASC | padres y niños | 8 a 17 años |
| Autoconfianza | BASC | niños | 12 a 17 años |
| Ayuda y colaboración | AECS | niños | 12 a 17 años |
| Asertividad/seguridad y firmeza | AESC | niños | 12 a 17 años |
| Conformidad con lo socialmente correcto | AESC | niños | 12 a 17 años |
| Equidad de género | Elaboración propia | padres y niños | 12 a 17 años |
| Prácticas de riesgo | Elaboración propia | niños | 12 a 17 años |
| Rendimiento escolar | Elaboración propia | padres  | 12 a 17 años |

## Efectos heterogéneos

Además de los resultados globales del programa, se considera posible que algunos resultados varíen en función de características de la intervención, de los niños o de sus familias. Entre las posibles dimensiones de heterogeneidad en los resultados, se identifican las siguientes:

* Grupos de Edad: de una edad a otra las prioridades en cuanto a la formación de valores que busca lograr FCC cambian. Por ejemplo, en jóvenes pueden ser importantes los aspectos de sexualidad, y en los niños el modo de comportamiento.
* Sexo: en algunos casos las niñas pueden ser más receptivas con las enseñanzas que se les está impartiendo. Por ejemplo, podría ocurrir que en las variables de equidad de género sean las mismas niñas quienes en un comienzo tengan una percepción de que este problema sea socialmente aceptado y que lo vean como algo normal. Sin embargo, esto no es del todo claro, ya que el problema también podría estar más arraigado en los niños.
* Sede: no hay hipótesis claras en cuanto a las heterogeneidades por sede, a pesar de su distinto modo de financiamiento y de sus características de infraestructura como por ejemplo el material de las canchas. No obstante, se planea realizar una clasificación que permita estudiar alguna posible heterogeneidad.
* Nivel socioeconómico: se plantea indagar los impactos del programa en función del nivel socioeconómico de los hogares. El índice de nivel socioeconómico puede ser obtenido a partir de la tenencia de bienes como computadora, lavadora, microondas y nevera.

Varios de estos efectos heterogéneos pretenden ser estudiados para ver si existe alguna complementariedad entre la participación dentro del programa y alguna característica del niño o del hogar. Sin embargo, los resultados esperados en estos aspectos no son definitivos, por lo cual una vez que se obtengan los resultados, deberían analizarse los mecanismos que actúan en favor de estos.

## Variables de control

De la información recolectada en la línea de seguimiento, se puede emplear como variable de control a la educación del jefe de hogar y a la identificación de si el jefe es la madre o el padre, asumiendo que el hecho de haber brindado la posibilidad de formar parte del programa no influyó en las decisiones del nivel de educación máximo alcanzado por los padres. De igual forma, existe la posibilidad de incluir el nivel socioeconómico construido en función de la tenencia de bienes como computadora, lavadora, microondas y nevera.

Las variables de estratificación que formaron parte de la aleatorización como la sede, el género y la categoría de edad del niño también serán incluidas como variables de control.

## Levantamiento de campo:

El trabajo de campo se realizó entre Septiembre de 2016 y Abril de 2017. Se realizaron entrevistas tanto días de semana como durante fines de semana, debido a la cantidad de niños que participaban en el colegio todo el día durante los días de semana. Como condición se estableció que estuviese presente el cuidador principal del niño para realizarle ambas entrevistas y que éste firmara un consentimiento informado. La metodología utilizada para contactar al cuidador del niño(a) fue a través de dos opciones complementarias: 1) citas preestablecidas, 2) Visitas directas al hogar.

Se estableció un protocolo de contacto para agotar todas las alternativas posibles antes de considerar que un contacto estaba perdido o que rechazó la encuesta. Al momento de elaborar este documento se contó con el 90% de recontacto de la muestra (es decir, un desgaste de 10%), lo que equivale a un total de encuestas efectivas de 1.490 (850 correspondientes al grupo de tratamiento y 640 al control) de las 1.649 posibles.

# Plan de análisis de resultados

## Especificación principal

Con el fin de medir los efectos del programa, un primer paso consistirá en evaluar los efectos de **haber ofrecido** la participación en FCC a través del *Intention to Treat* (ITT). Dado que la población no cumplidora es mínima - al considerar que un niño participa del programa cuando asiste al menos a una clase - resultaría trivial estimar los efectos del tratamiento sobre los tratados (TOT) a través del *Local Average Treatment Effect* (LATE) utilizando como instrumento la asignación al tratamiento.

Debido a la ausencia de línea de base, no es posible emplear un modelo de Diferencias en Diferencias (DID) o ANCOVA para medir los impactos del programa. Por esta razón, se empleará el estimador de diferencias simples de medias post-tratamiento. Formalmente, el efecto de la asignación al tratamiento (ITT) se va a calcular en base al siguiente modelo:

$Y\_{ija, POST}=θASIGN\_{i}+X\_{ija,PRE}β+γ\_{j}+λ\_{a}+ε\_{ija}$ (1)

donde $Y\_{ija, t}$ es el valor de la variable de resultados para el niño $i$, de edad $a$ y que practica la actividad en la sede $j$; $ASIGN\_{i}$ es una variable binaria que es igual a 1 si el niño $i$ resultó ganador de la beca, y 0 en caso contrario;$ X\_{ija}$ es un vector de características del niño y su hogar medidas en la línea de base; $ε\_{ija}$ representa el término de error; y los parámetros $γ\_{j}$ y $ λ\_{a}$ denotan, respectivamente, efectos fijos por sede y grupo de edad del participante. El coeficiente de interés es $θ$, que representa el efecto de haber sido asignado al tratamiento sobre la variable de resultado.

Por otra parte, el TOT se puede obtener a través del LATE, estimando el siguiente modelo:

$Y\_{ija, POST}=θTRATAM\_{i}+X\_{ija,PRE}β+γ\_{j}+λ\_{a}+ε\_{ija}$ (2)

donde $TRATAM\_{i}$ en principio puede ser una variable binaria que asume el valor 1 si el niño $i$ efectivamente asistió a los entrenamientos de FCC, y 0 en caso contrario. Sin embargo, debido a que en este tipo de intervención, no resulta claro definir cuándo un niño recibe el tratamiento -es decir, definir un porcentaje de clases a las que un niño debió haber asistido como mínimo para considerar que recibió el tratamiento- una alternativa es definir a $TRATAM\_{i}$ como el porcentaje efectivo de asistencias que obtuvo cada niño. De este modo, para evaluar el TOT, por un lado pueden definirse cortes en los niveles de asistencia para definir un tratamiento discreto del 10%, 30% y 50%, y por otro lado se puede asumir un tratamiento de naturaleza continua, por ejemplo, usando la tasa de asistencia definida como el cociente entre la cantidad de veces que el niño asistió y la cantidad total de clases a lo largo de los 16 meses de intervención.

El parámetro de interés de la especificación anterior es $θ$, pero su interpretación varía al asumir tratamientos de distinta naturaleza. Para el caso discreto, $θ$ representa el efecto de participar en las prácticas de la Academia para los niños que efectivamente participaron, definiendo a priori un nivel mínimo de asistencias. Para el caso continuo, el parámetro mediría, en promedio, el efecto de asistir a una clase adicional en la Academia sobre la variable de interés.

Debido a que la asistencia al tratamiento es endógena, en principio la estimación del TOT se puede efectuar bajo el método de variables instrumentales, utilizando como instrumento la asignación al tratamiento $Z\_{i}$.

## Análisis de funcionamiento de las escalas

Como parte del análisis de la validez de los instrumentos, se realizaron análisis de confiabilidad de las escalas y factorial exploratorio.

La confiabilidad de una escala refleja la medida en que los ítems de una misma escala se relacionan entre ellos: permite suponer que las preguntas sí se refieren y evalúan una sola y misma variable. Para analizar la confiabilidad de las diferentes escalas de la encuesta, se utilizó el alfa de Cronbach, que asume valores entre -1 y 1. Un alfa por debajo de 0,5 es malo: indica que los distintos ítems de una misma escala no miden lo mismo y que por ende no es una escala confiable. Un alfa por encima de 0,6 es aceptable y un alfa por encima de 0,8 es bueno: los resultados se pueden utilizar para realizar los análisis estadísticos deseados.

Por su lado, el análisis factorial exploratorio establece si los ítems de una escala se agrupan en una misma variable o en distintas variables. Esto permite analizar las distintas variables o conceptos que puede estar midiendo una misma escala. Este análisis se realizó utilizando el método de componentes principales para analizar el comportamiento de cada ítem en la solución a 1 solo factor. Se esperaba que para la solución a 1 factor cada ítem cargara en el factor con al menos 0.35 en valor absoluto.

Estos dos tipos de análisis permiten concluir si los datos recogidos pueden ser utilizados para los análisis que se quieren hacer con estas variables o si es necesario excluir previamente algún ítem de las escalas para mejorar la medición de la variable[[9]](#footnote-9).

El plan de análisis contempla tres tipos de tratamiento de los datos, de acuerdo al resultado de confiabilidad y análisis factorial:

1. **Escalas con buen funcionamiento:** en aquellas variables que hayan presentado un funcionamiento adecuado, es decir, un alpha de Cronbach mayor a 0,6 y que hayan arrojado un solo factor en el análisis factorial, se utilizarán sus datos para el análisis de resultados sin ningún tipo de modificación.
2. **Escalas con funcionamiento adecuado después de eliminar algún ítem:** las variables que presenten un alpha de Cronbach mayor a 0,6, y que hayan arrojado un solo factor en el análisis factorial luego de eliminar uno o más ítems, se utilizarán sus datos para el análisis de resultados con la modificación de los ítems eliminados.
3. **Escalas con mal funcionamiento (ligero):** las variables que presenten un alpha de Cronbach entre 0,4 y 0,6, y que hayan arrojado más de un factor en el análisis factorial, se analizarán con mucho cuidado.
4. **Escalas con mal funcionamiento (fuerte):** las variables que presenten un alpha de Cronbach por debajo de 0,4, y que hayan arrojado más de un factor en el análisis factorial, se descartarán y no se utilizarán para los resultados.

## Análisis del desgaste

Cuando la tasa de desgaste es de naturaleza aleatoria, esto solamente representa un problema en cuanto al poder estadístico. No obstante, cuando la pérdida de unidades elegibles de la muestra inicial no es aleatoria se introduce un sesgo de selección: la asignación al tratamiento pierde su aleatoriedad inicial.

Por esta razón, una vez que se determinen con precisión las unidades perdidas, se analizará la naturaleza de dicho desgaste. En caso de que se encuentre que no fue aleatoria (tasas de desgaste distintas entre grupos de tratamiento y control, y características distintas entre los que salen versus los que siguen en la muestra final), se seguirán algunas estrategias como la paramétrica basada en el modelo de selección de Heckman (1979) y no paramétricas como las propuestas por Horowitz y Manski (2000) y Lee (2009).

Además, otro aspecto clave para que el problema de desgaste sea menos relevante es que su porcentaje debe ser pequeño; en el programa de FCC, la tasa de desgaste prevista es del 10%

## Estructura de los errores

Además de presentar los resultados con errores robustos a la heteroscedasticidad, se presentarán dos tipos de correcciones alternativas, una para tener en cuenta la posible correlación al interior de ciertos grupos (clusters) y otra para tener en cuenta la presencia de hipótesis múltiples.

### Corrección por múltiples hipótesis

En principio existe la posibilidad de estimar los efectos del programa sobre cada *outcome* medido. No obstante, dado que se evalúan varias dimensiones a la vez, lo correcto es emplear pruebas de hipótesis múltiples como una medida preventiva para no caer en el problema de sobre-rechazo de hipótesis nulas. De la misma manera, estas pruebas son útiles cuando se analicen efectos heterogéneos en distintos grupos de la población.

En este sentido, se pueden estimar test basados en *Family Error Rate* (FWER), en *False Discovery Rate* (FDR) o con índices agregados de *outcomes*. El primer enfoque considera la probabilidad de que al menos una hipótesis nula sea rechazada cuando todas las hipótesis nulas son verdaderas, el segundo se concentra en la proporción esperada del total de hipótesis nulas que se rechazan erróneamente, y el tercero evalúa la significatividad individual de un índice agregado de variables de interés. Si bien los test usuales de FWER, como Bonferroni y Holm, tienen menor poder estadístico con respecto a los tests FDR, en el presente trabajo se realizará un ajuste bajo el método FWER de List et al. (2015), el cual, a diferencia de los tradicionales, tiene mayor poder debido a que considera una estructura de dependencia conjunta de los estadísticos de prueba; además es asintóticamente balanceado en el sentido de que las probabilidades marginales de cometer algún error de tipo I son aproximadamente iguales. Para el enfoque de FDR se puede emplear el algoritmo de dos etapas de Benjamini, Krieger, y Yekutieli (2006), el cual posee mayor poder que el algoritmo más básico de Benjamini y Hochberg. Finalmente, para la creación de índices agregados se puede seguir la estrategia descrita en Anderson (2008), en donde se requiere que las variables de interés estén estandarizadas, y cuyo objetivo es crear un índice en donde los ponderadores de cada variable se calculan para maximizar el monto de información capturado en el índice y dan menos importancia a las variables que están más correlacionadas.

De este modo, en primer lugar se pueden usar los test FWER y FDR para realizar la inferencia en tres grupos de variables (habilidades cognitivas, socioemocionales, y bienestar) considerando en cada caso las variables que a priori parecen ser las que más efectos poseen (ver sección 3.2); además la corrección bajo estos métodos permitirán determinar cuáles son los *outcomes* que determinan los resultados globales sobre las dimensiones estudiadas. Luego, se crearán índices agregados con variables que conceptualmente pueden ser agrupadas. Entre los índices a medir se encuentran los siguientes:

**Tabla 4.-** Índices agregados.

|  |  |  |  |
| --- | --- | --- | --- |
| **ÍNDICE** | **REPORTE** | **EDAD** | **VARIABLES** |
| Habilidades socioemocionales | padres y niños | 8 a 17 años | Relaciones interpersonales, autosuficiencia, autoestima, habilidades sociales, autoconcepto físico, conexión social, equidad de género según niños y padres, liderazgo, actitud del día a día, problemas de conducta, hiperactividad según niños y padres, comunicación funcional, actitud hacia el colegio  |
| Habilidades socioemocionales | padres y niños | 12 a 17 años | Relaciones interpersonales, autosuficiencia, autoestima, habilidades sociales, autoconcepto físico, respeto, solidaridad, honestidad, tolerancia, conexión social, equidad de género según niños y padres, liderazgo, actitud del día a día, problemas de conducta, hiperactividad según niños y padres, comunicación funcional, actitud hacia el colegio  |
| Habilidades socioemocionales | niños | 8 a 17 años | Relaciones interpersonales, autosuficiencia, autoestima, autoconcepto físico, conexión social, equidad de género según niños, hiperactividad según niños, actitud hacia el colegio  |
| Habilidades socioemocionales | niños | 12 a 17 años | Relaciones interpersonales, autosuficiencia, autoestima, autoconcepto físico, respeto, solidaridad, honestidad, tolerancia, conexión social, equidad de género según niños, hiperactividad según niños, actitud hacia el colegio  |
| Habilidades socioemocionales (BASC) | niños | 8 a 17 años | Relaciones interpersonales, autosuficiencia, autoestima, hiperactividad según niños, actitud hacia el colegio  |
| Habilidades socioemocionales (BASC) | padres  | 6 a 17 años | Habilidades sociales, actitud del día a día,hiperactividad según padres, comunicación funcional |
| Habilidades socioemocionales (AESC) | niños | 12 a 17 años | Respeto, solidaridad, honestidad, tolerancia |
| Habilidades cognitivas | padres | 6 a 17 años | Rendimiento en matemática y comunicación |
| Bienestar | padres | 6 a 17 años | Salud comparado con otros niños, salud comparado año pasado, dificultades en el niño comparado con otros niños, dificultades en el niño comparado año pasado |
| Entorno | niños | 8 a 17 años | Fumando, tomando alcohol, gente borracha, consumo de drogas, venta de drogas, uso de arma, agresión a otra persona, robo |

# Referencias

Barron, J. M., B. T. Ewing, and G. R. Waddell (2000): “The Effects Of High School Athletic Participation On Education And Labor Market Outcomes," The Review of Economics and Statistics, 82(3), 409-421.

Benjamini, Y., Krieger, A., y Yekutieli, D. (2006): “Adaptive Linear Step-Up Procedures That Control the False Discovery Rate” Biometrika, 93, 491–507.

Cabane, C., A. Hille, and M. Lechner (2015): “Mozart or Pelhe Effects of Teenagers' Participation in Music and Sports," Discussion paper.

CAF, 2017: “Práctica de fútbol y desarrollo de habilidades cognitivas y socioemocionales: Evidencia de un estudio experimental en Lima, Perú,” Versión preliminar para comentarios.

Ewing, B. T. (1998): “Athletes and work," Economics Letters, 59(1), 113-117.

Felfe, C., M. Lechner, and A. Steinmayr (2011): “Sports and Child Development," IZA Discussion Papers 6105, Institute for the Study of Labor (IZA).

Heckman, James (1979): “Sample Selection Bias as a Specification Error,” Econometrica, 47(1), 153-161.

Horowitz y Manski (2000): “Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data,” Journal of the American StatisticalAssociation, 95, 77-84.

Lechner, M. (2008): “Long-Run Labour Market E\_ects of Individual Sports Activities," IZA Discussion Papers 3559, Institute for the Study of Labor (IZA).

Lee, David (2009): “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects,” Review of Economic Studies, 76, 1071–1102

List, J., Shaikh, A. y Xu, Y. (2015): “Multiple Hypothesis Testing in Experimental Economics,” University of Chicago.

Long, J. E., and S. B. Caudill (1991): \The Impact of Participation in Intercollegiate Athletics on Income and Graduation," The Review of Economics and Statistics, 73(3), 525-31.

Pfeifer, C., and T. Cornelissen (2010): “The impact of participation in sports on educational attainment{New evidence from Germany," Economics of Education Review, 29(1), 94-103.

Stevenson, B. (2010): “Beyond the Classroom: Using Title IX to Measure the Return to High School Sports," The Review of Economics and Statistics, 92(2), 284-301.

1. Por razones logísticas, no se realizó encuesta de línea base, pero dado que los grupos fueron seleccionados aleatoriamente, cualquier diferencia estadísticamente significativa que se observe entre los grupos podrá atribuirse al programa FCC. [↑](#footnote-ref-1)
2. Al equipo que marca el mayor número de goles se le otorga 9 corazones, si hay empate 5 corazones y el que anote menos goles se les da 3 por participación. [↑](#footnote-ref-2)
3. Los participantes son clasificados en cuatro categorías en función de su edad: creadores (5 a 7), exploradores (8 a 10), constructores (11 a 13) y promotores (14 a 17). [↑](#footnote-ref-3)
4. Adicionalmente, fueron inscritos 52 niños, los cuales no fueron considerados para realizar la aleatorización. Esto debido a que en su estrato de sede, género y grupo de edad, no existían suficientes niños para llevar a cabo la aleatorización (falta de demanda) o a la falta de cupos disponibles (falta de oferta), lo cual implicó que fueran seleccionados sin un procedimiento aleatorio. Es por esto que no forman parte de la muestra. [↑](#footnote-ref-4)
5. Esta información proviene de un reporte previo del levantamiento de la información. [↑](#footnote-ref-5)
6. Si se considerase valores alternativos para ciertos parámetros, el EMD variaría: (i) si la varianza del tamaño del efecto asciende a 0,1, el EMD es de 0,28; (ii) si el cumplimiento es del 40%, el EMD sería de 0,38; (iii) finalmente, si se tomara la muestra final de 1.490 participantes sin ajustar por tasa de desgaste, el EMD sería de 0,21. [↑](#footnote-ref-6)
7. El cuestionario de niños posee diferentes versiones según el rango de edad, debido a la instrumentación de los tests utilizados. El cuestionario de padres es el mismo para todos. [↑](#footnote-ref-7)
8. Se utilizaron instrumentos estandarizados y en algunos casos se refrasearon preguntas para adaptarlos al lenguaje local, como por ejemplo: “Siento deseos de dejar la escuela” por “Siento deseos de dejar el colegio”. [↑](#footnote-ref-8)
9. Todos los análisis de esta sección se realizaron con el paquete estadístico R 3.3.2 [↑](#footnote-ref-9)