Guía metodológica

Todos y todas se quedan en la escuela

Modelo de indicadores de alerta temprana sobre deserción escolar

Save the Children

ICEFI
INSTITUTO CENTROAMERICANO DE ESTUDIOS FISCALES

Guatemala, Centroamérica, septiembre de 2022

Guía metodológica

Todos y todas se quedan en la escuela

Modelo de indicadores de alerta temprana sobre deserción escolar

Créditos

Supervisión Jonathan Menkos Zeissig ▪ director ejecutivo, Icefi

Coordinación Carlos Gossmann Zarazúa ▪ economista sénior, Icefi

Investigación y redacción Julio César Hernández ▪ economista investigador, Icefi
Sucely Marleny Donis Bran ▪ economista investigadora, Icefi

Colaboración Paolo Marsicovetere ▪ investigador, consultor Icefi
Barbara Stone ▪ pasante, Icefi

Diseño, diagramación y proceso de publicación Mónica Juárez Balcárcel
▪ Asistente de comunicación, Icefi

Difusión por medios electrónicos José Ochoa Arévalo
▪ Asistente de comunicación, Icefi

Administración Débora Alvarado ▪ coordinadora de Desarrollo Institucional

© Instituto Centroamericano de Estudios Fiscales (Icefi)
7ª avenida 5-45 zona 4, edificio XPO1
Oficinas 505 y 506
Ciudad de Guatemala, Guatemala

Septiembre 2022

Insumo elaborado para el proyecto «*Sin Dejar a la Niñez Atrás*»
con el auspicio de Save the Children

Este documento ha sido elaborado por el Instituto Centroamerican.o de Estudios Fiscales (Icefi). Su contenido es responsabilidad exclusiva del Icefi y no necesariamente refleja los puntos de vista de las instituciones cooperantes antes mencionadas. En el Icefi consideramos que el conocimiento siempre está en construcción, por lo que cualquier comentario u observación es bienvenido en el correo electrónico info@icefi.org.

El Icefi promueve el uso del lenguaje inclusivo que no discrimine ni marque diferencias entre las personas según su sexo o identidad de género. No obstante, por economía expresiva, este documento podría contener expresiones en masculino genérico para referirse por igual a mujeres y hombres, uso que en ninguna forma debe entenderse como una intención discriminatoria por parte del Instituto.

Índice

Presentación 9

1. El derecho a la educación en Guatemala 12

1.1 Definición y características/elementos del derecho a la educación 12

1.2 Obligaciones del Estado de Guatemala respecto al derecho a la educación 13

1.2.1 Obligaciones a nivel internacional sobre el derecho a la educación 13

1.2.2 Obligaciones a nivel nacional que garantizan el derecho a la educación 18

2. Indicadores de educación en Guatemala 23

2.1 Contexto demográfico y algunas condiciones socioeconómicas 23

2.2 Principales indicadores de educación de la población guatemalteca 23

2.2.1 Alfabetismo y años de escolaridad total 23

2.2.2 Alfabetismo y años de escolaridad juvenil 24

2.2.3 Matricula neta estudiantil 26

2.2.4 Establecimientos educativos y estudiantes inscritos por nivel y sector 26

2.2.5 Población fuera de la escuela 28

2.2.6 Niñez y adolescencia fuera de la escuela 32

2.2.7 El acceso a herramientas tecnológicas de las y los estudiantes 34

2.3 Impactos en educación derivado de la pandemia del COVID-19 37

2.4 Inversión pública en educación por parte del Gobierno central 38

3. Metodología: predicción de abandono escolar 44

3.1 Antecedentes sobre el modelo que predice el abandono escolar 45

3.2 Explicación de los modelos predictivos y de aprendizaje de máquina (machine learning) 48

3.3 Construcción del modelo para medir el riesgo de abandono escolar 50

3.3.1 Generar base de datos 55

3.3.2. Aplicación del modelo de medición de riesgo de abandono escolar 61

4. Conclusiones 67

5. Recomendaciones 70

6. Referencias bibliográficas 73

Índice de Figuras

Figura 1. Características del derecho a la educación 13
Figura 2. Metas ODS 4 – Educación de Calidad 17
Figura 3. Eje, prioridades, metas y resultados del Plan Nacional de Desarrollo 2032, relacionados con el derecho a la educación 19
Figura 4. Sub-ajuste, ajuste correcto y sobreajuste de un conjunto de datos con tendencia cuadrática 50
Figura 5. Tabla de Eventos 51
Figura 6. Elaboración de base de datos para modelo de deserción escolar 55
Figura 7. Ejemplo de base de datos consolidada de registros administrativos 59
Figura 8. Ejemplo de base de datos consolidada integrando variables *proxys* 60
Figura 2. Curva ROC 63
Figura 2. Curva ROC con el punto de sensibilidad 64

Índice de Gráficas

Gráfica 1. Guatemala: Pirámide poblacional 2022 y 2050 21
Gráfica 2. Guatemala: niñas, niños y adolescentes de 7 a 17 años según lectura o escritura en un idioma indígena y el idioma en el que se le enseña en un centro educativo (2014) 23
Gráfica 3. Guatemala: Tasa neta de cobertura – En porcentajes, periodo 2000 a 2021 26
Gráfica 4. Guatemala: Porcentaje de establecimientos educativos por nivel y sector (2019-2021) 27
Gráfica 5. Guatemala: Porcentaje de estudiantes por nivel educativo y sector (2019-2021) 28
Gráfica 6. Guatemala: Estimación de la Población de 0 a 18 años fuera de la escuela en 2021 30
Gráfica 7. Guatemala: Deserción según nivel educativo, en porcentaje del total matriculado (2015 a 2021) 31
Gráfica 8. Guatemala: Deserción por nivel educativo según sexo (2015 a 2021) 32
Gráfica 9. Guatemala: Causa principal por la que no asiste a un establecimiento educativo en 2018 niños, niñas y adolescentes entre 4 y 17 años. 33
Gráfica 10. Guatemala: Acceso a computadora en niños, niñas y adolescentes de 7 a 17 años que asistieron a un establecimiento educativo en 2018 36

Gráfica 11. Guatemala: Acceso a servicios de internet en niños, niñas y adolescentes de 7 a 17 años que asistieron a un establecimiento educativo en 2018 36
Gráfica 12. Guatemala: gasto ejecutado por el Ministerio de Educación, como proporción de los egresos totales del Gobierno central y del PIB, 2015 – 2021 ejecutado, presupuesto 2022 39
Gráfica 13. Guatemala: Entidades con mejor nivel de ejecución de la inversión pública en niñez y adolescencia (IPNA) en el presupuesto vigente del Gobierno Central en 2022. 40
Gráfica 14. Ejemplo de una regresión lineal 47
Gráfica 15. Ejemplo de una regresión lineal con poca capacidad predictiva 49
Gráfica 16. Ilustración de la curva ROC 53

Índice de Tablas

Tabla 1. Principales obligaciones del Estado de Guatemala a nivel internacional en cuanto al derecho a la educación 14
Tabla 2. Principales recomendaciones al Estado de Guatemala por exámenes realizados por Órganos de Tratados, en materia de derecho a la educación 15
Tabla 3. Leyes que reconocen el derecho a la educación en Guatemala 18
Tabla 4. Guatemala: Características lingüísticas de la niñez y adolescencia en Guatemala (2018) 22
Tabla 5. Guatemala: Población alfabeta y número de años de escolaridad. De 15 y más años de edad (2018) 24
Tabla 6. Guatemala: Población alfabeta y número de años de escolaridad – de 15 a 24 años .. 25
Tabla 7. Guatemala: Causa principal por la que no asiste a un establecimiento educativo en 2018 adolescentes entre 13 y 17 años 34
Tabla 8. Guatemala: Acceso a servicios tecnológicos de niños, niñas y adolescentes de 7 a 17 años que asistieron a un establecimiento educativo en 2018 35
Tabla 9. Guatemala: Simulación de impactos en pérdida de años de escolaridad, puntaje de pruebas pisa y nivel de rendimiento mínimo, según el tiempo de cierre de establecimientos educativos 38
Tabla 10. Guatemala: Egresos del Ministerio de Educación según programa, devengados en 2016 a 2021. Cifras en millones de quetzales y como porcentaje del promedio total 41
Tabla 11. Listado de variables que puede incluir el modelo de estimación de riesgo de abandono escolar 56

Siglario

BM	Banco Mundial
CUI	Código único de identificación
CDESC	Comité de Derechos Económicos, Sociales y Culturales
CESCR	Comité de Derechos Económicos, Sociales y Culturales
CDN	Comité de los Derechos del Niño
CERD	Comité para la Eliminación de la Discriminación Racial
CEDAW	Comité para la Eliminación de la Discriminación Contra la Mujer
CRPD	Comité sobre Derechos de Personas con Discapacidad
CPR	Constitución Política de la República
ROC	Curva de la Característica Operativa del Receptor
Encovi	Encuesta Nacional de Condiciones de Vida
ENEI	Encuesta Nacional de Ingresos y Egresos
GTJ	Grupo de trabajo sobre juventud
Icefi	Instituto Centroamericano de Estudios Fiscales
IPNA	Inversión Pública en Niñez y Adolescencia
Mineduc	Ministerio de Educación
ODS	Objetivos de Desarrollo Sostenible
OEACT	Obligaciones del Estado a Cargo del Tesoro
ONU	Organización de las Naciones Unidas
Unesco	Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura
PNDK	*Plan Nacional de Desarrollo K'atun Nuestra Guatemala 2032*
PISA	Programa Internacional para la Evaluación de Estudiantes

Presentación

El derecho a la educación permite la inclusión de las personas en la sociedad, lo cual genera efectos multiplicadores que mejoran las condiciones de vida y permiten el desarrollo humano. Entre ellos se pueden mencionar: facilita el acceso a empleos decentes, coadyuva a la superación de las condiciones de pobreza y pobreza extrema, permite acceder a la información y la adquisición de habilidades para el ejercicio crítico e informado de la ciudadanía. Como bien lo menciona el Comité de Derechos Económicos, Sociales y Culturales (DESC), «La educación es una de las mejores inversiones financieras que los Estados pueden hacer, pero su importancia no es únicamente práctica pues dispone de una mente instruida, inteligente y activa, con libertad y amplitud de pensamiento, es uno de los placeres y recompensas de la existencia humana».

En Guatemala, el derecho a la educación está normado desde el marco nacional como el internacional. Desde el ámbito legal, la educación constituye un servicio que el Estado otorga a la ciudadanía, desde su primera infancia y niñez, sin ser excluyente y de forma gratuita. No obstante, la realidad de cientos de miles de menores actualmente es que se encuentran fuera del sistema educativo (tanto a nivel público como privado). Estimaciones realizadas a partir de los datos de la matrícula estudiantil del Ministerio de Educación (Mineduc) y las proyecciones de población presentadas por el Instituto Nacional de Estadística, muestran que durante 2021 más de 1.2 millones de niñas, niñas y adolescentes estuvieron fuera de la escuela, colegio o instituto.

Este problema afecta a varios niveles de educación y sus determinantes son diversas y transversales, que pueden afectar a un estudiante por uno a varios factores asociados, que derivan en abandono de los estudios al finalizar un ciclo y no continúe al siguiente grado, aún habiendo aprobado el actual, o incluso, deserte durante el ciclo lectivo que está cursando. Muchas de estas situaciones están asociadas con fenómenos socioeconómicos, pero hace falta desarrollar un análisis técnico, apegado a mediciones de datos administrativos y del contexto del país, para poder explicar de la manera más asertiva la condición de abandono o deserción.

De esta cuenta, el Icefi en asocio con Save the Children, dentro del proyecto *Sin Dejar a la Niñez Atrás,* han elaborado esta guía metodológica para elaborar un modelo predictivo de deserción escolar. La metodología se basa en un modelo desarrollado por Adelman, M., Haimovich, F. y Vazquez, E. que, desde el Banco Mundial,

publicaron el documento *Predicting school dropout with administrative data: new evidence from Guatemala and Honduras*. Este se basa en un modelo de determinantes que buscan medir el riesgo de abandono escolar a partir de una serie de variables que parten de los registros administrativos que las entidades rectoras de los servicios de educación puedan brindar. Los resultados del modelo permiten medir la sensibilidad, precisión, especificidad y proporción de falsos positivos, de estudiantes que tienen riesgo de abandonar los estudios, lo cual se puede generar a partir de la información procesada.

El objetivo del modelo es delimitar algunas determinantes para prevenir el riesgo de la deserción escolar. Con ello, las autoridades pueden implementar sistemas de alerta temprana que eviten que los factores de deserción impacten en los estudiantes, antes que suceda el abandono, o bien, que atiendan a los estudiantes que ya abandonaron sus estudios y los reincorporen dentro del sistema educativo. Al proteger y recuparar la matrícula estudiantil, se garantiza el derecho a la educación, y los réditos sociales y económicos para el país son enormes.

Así pues, en un contexto en el que los sistemas de educación han adoptado nuevas modalidades en las que la educación a distancia y dirigida fue necesario ajustarlas por el distanciamiento social provocado por la pandemia del COVID-19, el riesgo de deserción escolar se incrementó y seguramente impactó las tasas de matrícula neta y bruta, en todos lo niveles educativos. Atender esta problemática puede ser la diferencia para que un niño o niña tenga un futuro mejor y, por ello, se debe procurar que *Todos y todas se quedan en la escuela*.

Icefi
Guatemala, Centroamérica, agosto de 2022

El derecho a la educación en Guatemala

1

La educación es un derecho fundamental de todas las personas, su garantía facilita o permite el goce pleno de otros derechos humanos. Es considerada como una de las llaves maestras para la inclusión de las personas en la sociedad, con efectos en el acceso a empleos decentes, en la superación de la pobreza, en el acceso a la información, en la adquisición de habilidades para el ejercicio crítico e informado de la ciudadanía, etc. (GTJ; 2021: 29). A continuación, se presentan los compromisos que el Estado de Guatemala ha asumido respecto al derecho a la educación. Antes de ello se presenta una breve definición del derecho a la educación y sus características, de acuerdo con el Comité de Derechos Económicos, Sociales y Culturales (CDESC).

1.1 Definición y características/elementos del derecho a la educación

El Comité de Derechos Económicos, Sociales y Culturales (CDESC) en su Observación General número 13, indica que la educación es un derecho humano intrínseco y un instrumento indispensable para la realización de otros derechos humanos, que facilita a las personas a salir de la pobreza y a participar plenamente en sus comunidades.

La educación tiene un papel decisivo en la emancipación de la mujer, en la protección de los niños contra la explotación, en la promoción de los derechos humanos, en la democracia y en la protección del ambiente. «La educación es una de las mejores inversiones financieras que los Estados pueden hacer, pero su importancia no es únicamente práctica pues dispone de una mente instruida, inteligente y activa, con libertad y amplitud de pensamiento, es uno de los placeres y recompensas de la existencia humana» (CDESC; 1999: 1).

De acuerdo con la Declaración Universal de los Derechos Humanos, la educación se debe orientar al desarrollo del sentido de la dignidad humana, debe lograr que todas las personas participen efectivamente en una sociedad libre y debe favorecer la comprensión entre todos los grupos étnicos, y entre las naciones y los grupos raciales y religiosos. De acuerdo con el CDESC, la educación en todas sus formas y niveles debe tener cuatro características interrelacionadas: *disponibilidad, accesibilidad, adaptabilidad y aceptabilidad*.

Figura 1. Características del derecho a la educación

Fuente: Icefi/Save the Children, con base en Organización de las Naciones Unidas (ONU).

1.2 Obligaciones del Estado de Guatemala respecto al derecho a la educación

El Estado de Guatemala ha asumido compromisos, tanto a nivel nacional como internacional, para garantizar el acceso universal a la educación, especialmente para personas o grupos que por diferentes circunstancias no pueden disfrutar plenamente de este derecho, tales como: mujeres, indígenas, personas con discapacidad, entre otras.

1.2.1 Obligaciones a nivel internacional sobre el derecho a la educación

Son múltiples los instrumentos internacionales por los cuales el Estado de Guatemala se ha comprometido a garantizar el derecho a la educación. En la siguiente tabla se mencionan los más importantes:

Tabla 1. Principales obligaciones del Estado de Guatemala a nivel internacional en cuanto al derecho a la educación

No.	Instrumento internacional	Artículo	Obligación
1	Declaración Universal de los Derechos Humanos	26	Establece que toda persona tiene derecho a la educación y que la educación debe ser gratuita, al menos en lo concerniente a la instrucción elemental y fundamental.
2	Pacto Internacional de Derechos Económicos, Sociales y Culturales	13	Se reconoce el derecho de toda persona a la educación. Establece que la enseñanza primaria debe ser obligatoria y asequible a todos gratuitamente. Asimismo, que la educación secundaria debe ser generalizada y hacerse accesible a todos, por la implantación progresiva de enseñanza gratuita.
3	Protocolo de San Salvador	13	Reconoce el derecho de todas las personas a la educación, establece como obligación del Estado garantizar enseñanza primaria obligatoria y asequible a todos gratuitamente. La enseñanza secundaria y superior debe ser accesible a todos, en particular, por implantación progresiva de enseñanza gratuita.
4	Convención sobre los Derechos del Niño	28	Se reconoce el derecho del niño a la educación, y para lograr en condiciones de igualdad el Estado debe: implantar la enseñanza primaria obligatoria y gratuita para todos; fomentar el desarrollo de educación secundaria mediante enseñanza gratuita y asistencia financiera; hacer la enseñanza superior accesible a todos, sobre la base de la capacidad; adoptar medidas para la asistencia regular a la escuela y reducir tasas de deserción.
5	Convención sobre eliminación de todas las formas de discriminación contra la mujer	10	Se asume la responsabilidad de adoptar todas las medidas apropiadas para garantizar la igualdad entre hombres y mujeres en la esfera de la educación, entre ellas: mismas oportunidades de acceso a programas de educación; mismos programas de estudios, exámenes, docentes, locales y equipos escolares; eliminación de conceptos estereotipados de los papeles masculino y femenino; mismas oportunidades para becas; reducción de tasas de abandono femenino y programas para mujeres que abandonan estudios prematuramente.
6	Convención Internacional sobre la Eliminación de todas las Formas de Discriminación Racial	5	Obligación de prohibir y eliminar la discriminación racial en todas sus formas y a garantizar el derecho a la educación y a la formación profesional sin distinción de raza, color y origen nacional o étnico.
7	Convención sobre los Derechos de las Personas con Discapacidad	24	Se reconoce el derecho de las personas con discapacidad a la educación, sin discriminación sobre la base de la igualdad de oportunidades, para lo cual el Estado debe asegurar que: las personas con discapacidad no queden excluidas del sistema general de educación; las personas con discapacidad puedan acceder a educación primaria y secundaria inclusiva, de calidad y gratuita en su comunidad; y, se hagan ajustes razonables en función de las necesidades individuales
8	Convención internacional sobre la protección de los derechos de todos los trabajadores migratorios y de sus familiares	30	Establece que todos los hijos de los trabajadores migratorios gozarán del derecho fundamental de acceso a la educación en condiciones de igualdad de trato con los nacionales del Estado de que se trate; asimismo, establece que acceso de los hijos de trabajadores migratorios a las instituciones de enseñanza preescolar o las escuelas públicas no podrá denegarse ni limitarse a causa de la situación irregular en lo que respecta a la permanencia o al empleo de cualquiera de los padres, ni del carácter irregular de la permanencia del hijo en el Estado de empleo.

Fuente: Icefi/Save the Children, con base en Organización de las Naciones Unidas.

Derivado de los compromisos asumidos por Guatemala en materia del derecho a la educación, el país ha sido examinado en múltiples ocasiones por los órganos encargados de verificar el pleno cumplimiento de dichos compromisos, de lo cual han derivado múltiples recomendaciones al Estado de Guatemala, entre estas se encuentran las siguientes:

Tabla 2. Principales recomendaciones al Estado de Guatemala por exámenes realizados por Órganos de Tratados, en materia de derecho a la educación

No.	Órgano	Año	Recomendación
i	Comité para la Eliminación de la Discriminación Contra la Mujer (CEDAW)	2009	Adoptar medidas proactivas para lograr que todos los niños, especialmente las niñas, tengan acceso a la educación básica
		2017	Intensificar sus esfuerzos, por conducto del Ministerio de Educación, para aumentar la inclusión y retención de las niñas en la escuela, en particular en la enseñanza secundaria, prestando especial atención a las niñas indígenas.
ii	Comité de Derechos Económicos, Sociales y Culturales (CESCR)	2014	Propiciar el pleno acceso de niñez y adolescencia a la educación secundaria. Redoblar esfuerzos para garantizar el acceso de los pueblos indígenas a la educación intercultural y en sus lenguas propias y que vele por que esa educación se adapte a las necesidades específicas de esos pueblos. Ampliar el acceso a Internet y que redoble los esfuerzos en la instauración de centros de información y educación sobre el uso de las nuevas tecnologías e Internet, particularmente a los pueblos indígenas.
iii	Comité de los Derechos del Niño (CDN)	2010	Garantizar la plena aplicación en la práctica de todas las disposiciones jurídicas que prohíben la discriminación, luche contra esta, entre otros medios, asegurando la igualdad de acceso a la educación.
			Proporcionar educación gratuita en la práctica, así como una cantidad suficiente de escuelas, materiales escolares y docentes debidamente formados. Adoptar todas las medidas necesarias para velar por que los niños completen la enseñanza primaria y la secundaria, *incluidas medidas específicas para afrontar las razones que subyacen al abandono escolar.*
		2018	Aumentar los recursos presupuestarios destinados a hacer efectivos los derechos del niño, en particular el presupuesto asignado a los sectores sociales y a los niños en situación de desventaja, y a dar aplicación a las medidas relativas a la educación

No.	Órgano	Año	Recomendación
iv	Comité para la Eliminación de la Discriminación Racial (CERD)	2019	Garantizar la disponibilidad, accesibilidad y calidad de la educación para los niños indígenas, incluso en su lengua materna, y mediante la elaboración de planes educativos interculturales que cumplan con el objetivo de promoción y preservación de la identidad cultural de los pueblos indígenas e intensificar sus esfuerzos para erradicar el analfabetismo, principalmente entre las mujeres indígenas.
v	Comité sobre Derechos de Personas con Discapacidad (CRPD)	2016	Reconocer en su legislación y políticas un sistema de educación inclusiva gratuita y de calidad en todos los niveles del sistema educativo y la garantía de ajustes razonables para los estudiantes que los precisen, con recursos presupuestarios suficientes y formación adecuada de los docentes regulares. Adoptar medidas para asegurar la escolarización de todos los niños con discapacidad. Implementar con urgencia medidas de accesibilidad en los centros educativos y de todos los materiales didácticos y asegure su uso desde el inicio del curso académico, incluyendo la disposición de textos escolares en braille e intérpretes de lengua de señas.

Fuente: Icefi/Save the Children, con base en Organización de las Naciones Unidas.

En 2009, el Relator Especial sobre el Derecho a la Educación en una visita a Guatemala realizó una serie de recomendaciones al Estado, destacando:

a. Crear un mecanismo jurídico que asegure un aumento de la inversión en educación en un 0.5% del PIB al año, hasta alcanzar progresivamente el 6% del PIB.

b. La inclusión en la política educativa de un plan transversal de género, que fomente la inclusión de educación sexual y reproductiva en el currículum educativo.

c. Desarrollar, con carácter de urgencia nacional, acciones afirmativas de estímulo y dignificación de las lenguas y las culturas indígenas, que muestren a todos los grupos sociales.

d. Apoyar los planes de expansión y fortalecimiento de la educación pública secundaria y de la universidad estatal, dotando a estos niveles de mayores recursos.

e. Mejorar la infraestructura escolar (con prioridad para la provisión de agua y electricidad).

f. Formular un sistema de atención efectiva de los niños indígenas que se han desplazado por diferentes motivos a las grandes ciudades, con el fin de mitigar la pérdida de identidad.

g. Crear programas especiales de atención a los estudiantes con discapacidades, con docentes especializados y materiales adaptados a sus necesidades.

Además de los compromisos antes detallados, en 2015 el Estado de Guatemala se comprometió a cumplir con los Objetivos de Desarrollo Sostenible (ODS), en el marco de la Agenda de Desarrollo Sostenible 2030. Dentro de esta agenda se contemplan metas relacionadas con el derecho a la educación, especialmente dentro del ODS 4**.** Comprende metas que buscan garantizar una educación inclusiva y equitativa de calidad y promover oportunidades de aprendizaje permanente para todos. En la figura siguiente se observan las metas establecidas para el ODS 4.

Figura 2. Metas ODS 4 – Educación de Calidad

Meta 4.1	• De aquí a 2030, asegurar que todas las niñas y todos los niños terminen la enseñanza primaria y secundaria, que ha de ser gratuita, equitativa y de calidad y producir resultados de aprendizaje pertinentes y efectivos
Meta 4.2	• De aquí a 2030, asegurar que todas las niñas y todos los niños tengan acceso a servicios de atención y desarrollo en la primera infancia y educación preescolar de calidad, a fin de que estén preparados para la enseñanza primaria
Meta 4.3	• De aquí a 2030, asegurar el acceso igualitario de todos los hombres y las mujeres a una formación técnica, profesional y superior de calidad, incluida la enseñanza universitaria
Meta 4.4	• De aquí a 2030, aumentar considerablemente el número de jóvenes y adultos que tienen las competencias necesarias, en particular técnicas y profesionales, para acceder al empleo, el trabajo decente y el emprendimiento
Meta 4.5	• De aquí a 2030, eliminar las disparidades de género en la educación y asegurar el acceso igualitario a todos los niveles de la enseñanza y la formación profesional para las personas vulnerables, incluidas las personas con discapacidad, los pueblos indígenas y los niños en situaciones de vulnerabilidad
Meta 4.a	• Construir y adecuar instalaciones educativas que tengan en cuenta las necesidades de los niños y las personas con discapacidad y las diferencias de género

Fuente: Icefi/Save the Children, con base en Programa de las Naciones Unidas para el Desarrollo.

1.2.2 Obligaciones a nivel nacional que garantizan el derecho a la educación

El derecho a la educación está garantizado en la Constitución Política de la República (CPR). De acuerdo con el artículo 71, es obligación del Estado proporcionar y facilitar educación a sus habitantes sin discriminación alguna, mientras que el artículo 74 establece el derecho y obligación de los habitantes a recibir educación inicial, preprimaria, primaria y básica, también establece que la educación impartida por el Estado es gratuita. Por su parte, existen normativas que reconocen el derecho a la educación.

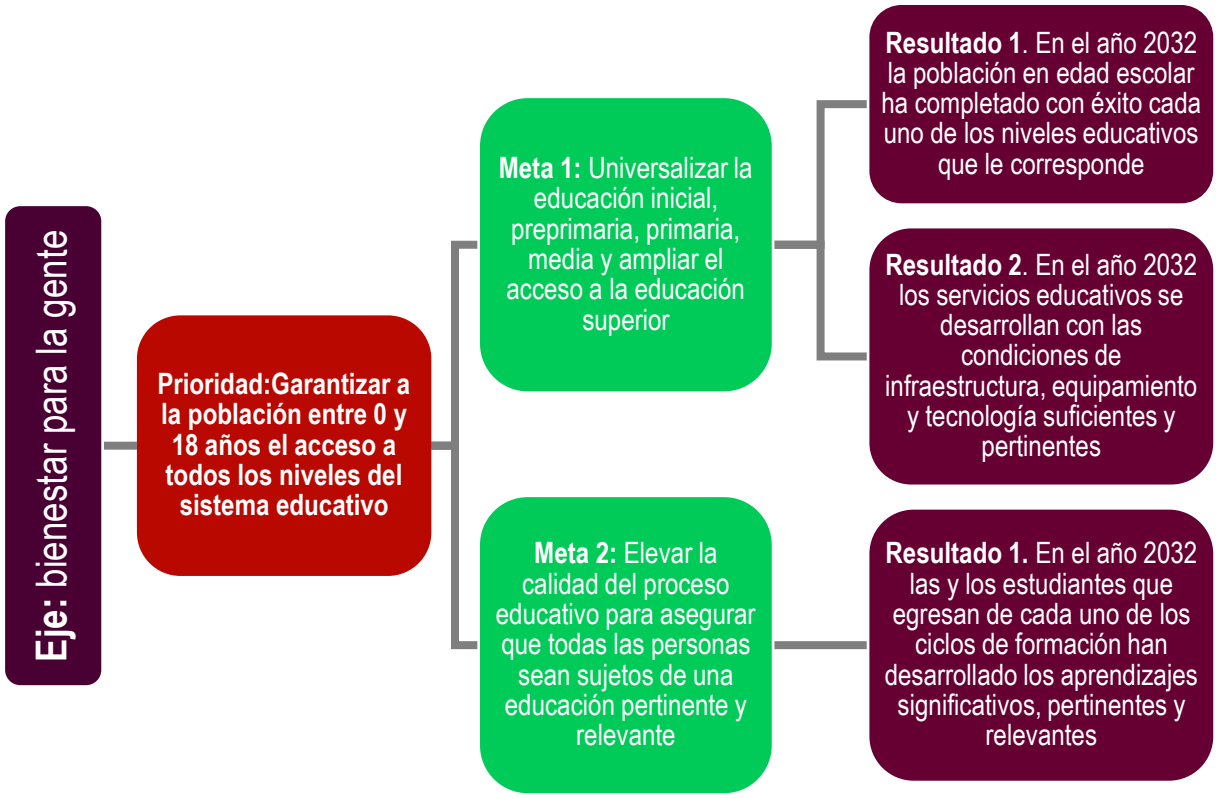
Tabla 3. Leyes que reconocen el derecho a la educación en Guatemala

No.	Legislación	Artículo	Contenido
1	Ley de Educación Nacional (Decreto 12-91)	33	Establece que es obligación del Estado propiciar una educación gratuita y obligatoria dentro de los límites de edad establecidos, así como propiciar y facilitar la educación a los habitantes sin discriminación alguna.
2	Ley de Protección Integral de la Niñez y Adolescencia (Decreto 27-2003)	36 y 37	Reconoce derecho a la educación de los niños, niñas y adolescentes, destacando el derecho a recibir una educación integral, y a la obligación del Estado de que la educación pública sea gratuita, laica y obligatoria hasta el último grado de diversificado.
3	Ley de Dignificación y Promoción Integral de la Mujer (Decreto 7-99)	9	Establece que el Estado debe garantizar a las mujeres la educación nacional en condiciones de equidad, en todas las etapas del proceso educativo, para lo cual debe extender y ampliar la cobertura escolar en todos los niveles para incrementar el acceso y permanencia de las mujeres a la educación en los distintos niveles del sistema.
4	Ley de Atención a las Personas con Discapacidad (Decreto 135-96)	25	Garantiza que las personas con discapacidad tengan derecho a la educación desde la estimulación temprana hasta la educación superior.
5	Ley de Protección para las Personas de la Tercera Edad (Decreto 80-96)	20	Establece que toda persona mayor debe tener acceso a la educación formal e informal sin que la edad sea un limitante.

Fuente: Icefi/Save the Children, con base en Programa de las Naciones Unidas para el Desarrollo.

En 2014 el Gobierno de Guatemala aprobó el *Plan Nacional de Desarrollo K'atun Nuestra Guatemala 2032 (PNDK)*, la política nacional de desarrollo de largo plazo que articula las políticas, planes, programas, proyectos e inversiones. Define prioridades, metas, resultados y lineamientos a cumplir en 2032, y contempla prioridades relacionadas directamente con el derecho a la educación, tal como se observa en la siguiente gráfica.

Figura 3. Eje, prioridades, metas y resultados del Plan Nacional de Desarrollo 2032, relacionados con el derecho a la educación



Fuente: Icefi/Save the Children, con base en Segeplan/Conadur.

Indicadores de educación en Guatemala

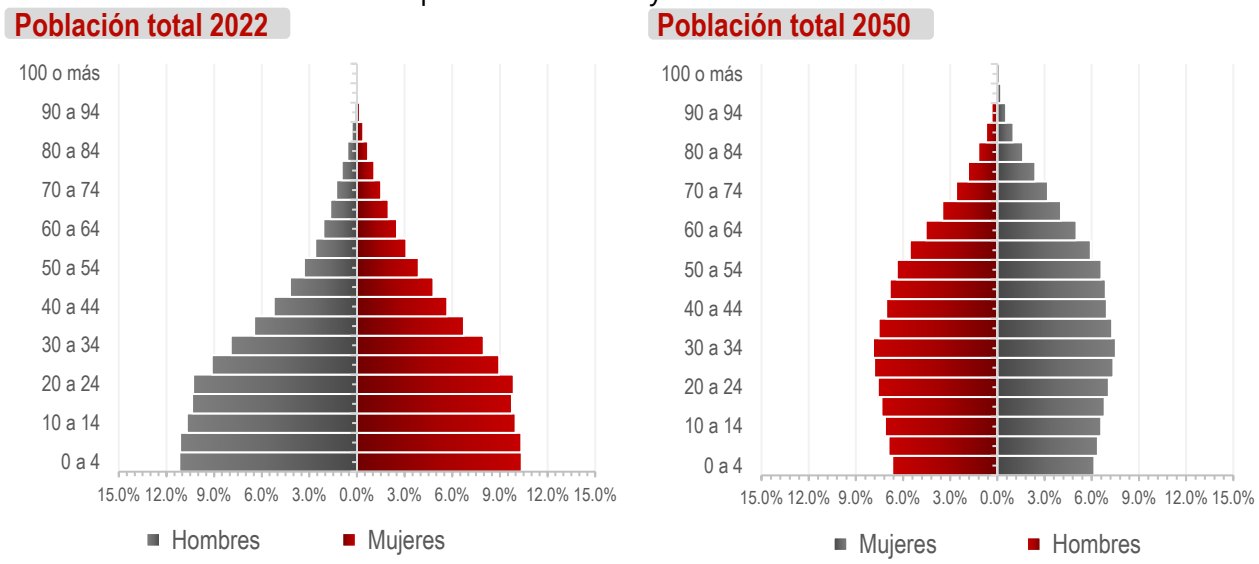
2

2.1 Contexto demográfico y algunas condiciones socioeconómicas

Desde la perspectiva demográfica, se estima que en 2022 Guatemala cuenta con un total de 17.4 millones de habitantes, de los cuales 50.8% son mujeres y 49.2% son hombres. La estructura poblacional se expande en su parte más baja, por lo que la población de Guatemala es relativamente joven, alrededor de 6.6 millones son niños, niñas y adolescentes menores de 18 años, representando el 37.9% de la población. Guatemala se ubica como el país con mayor proporción de niñez y adolescencia de la región centroamericana.

Según las mismas proyecciones, se estima que este segmento poblacional ha disminuido respecto a mediciones anteriores, y continuará haciéndolo según los procesos de transición demográfica. En 2015 inició el *bono demográfico*, en el cual, por cada 2 personas en edad económicamente dependiente, hay 3 o más en edad de trabajar. Con una alta cantidad de personas en condiciones de ingresar al mercado laboral, un bono demográfico no debería ser desaprovechado. Esto también evidencia una necesidad importante de generar y actualizar las políticas, programas y proyectos enfocados en la demanda de servicios de educación, salud y seguridad social, para atender a este gran segmento de la población. En las proyecciones de población, el bono demográfico se mantiene hasta la década de 2050, cuando se esperan indicios de que dicho escenario se revertirá debido al inicio del envejecimiento de la población: la niñez y adolescencia se reduciría a un 24% del total de la población.

Gráfica 1. Guatemala: Pirámide poblacional 2022 y 2050



Fuente: Icefi/Save the Children, con base en datos del *XII Censo de población y VII de vivienda* (Guatemala, 2018) y Proyecciones de Población.

Según su ubicación geográfica, durante 2018 el 51.5% de la niñez y adolescencia residía en áreas rurales, mientras que el 48.5% en áreas urbanas. Respecto a diversidad cultural y lingüística, por pueblo de pertenencia, el 47.8% se autoidentificó con un pueblo indígena (maya, garífuna, Xinka, afrodescendiente/creole/afromestizo), mientras que 52.0% se autoidentificó como ladino. En cuanto al idioma materno, el 69.0% confirma que es español castellano; 24.2% es un idioma indígena (Q'eqchi', K'iche', Mam, Kaqchiquel, Xinka o Garífuna).

Según características lingüísticas, la niñez y adolescencia es principalmente monolingüe, 76.4% habla un solo idioma y un 23.1% es bilingüe. De estos, 65 de cada 100 habla solo español, mientras que 11 de cada 100 habla únicamente un idioma Maya, Garífuna o Xinka. La importancia de conocer el comportamiento lingüístico de este segmento de la población sirve para planificar una política de educación según sus necesidades, pues es considerado que además de fortalecer la identidad y estimular la confianza desde edades tempranas, la educación multilingüe constituye un pilar fundamental para el desarrollo de un país. Sin embargo, el que se respete esta diversidad cultural y lingüística representa actualmente uno de los grandes retos para la educación en Guatemala.

Tabla 4. Guatemala: Características lingüísticas de la niñez y adolescencia en Guatemala (2018)

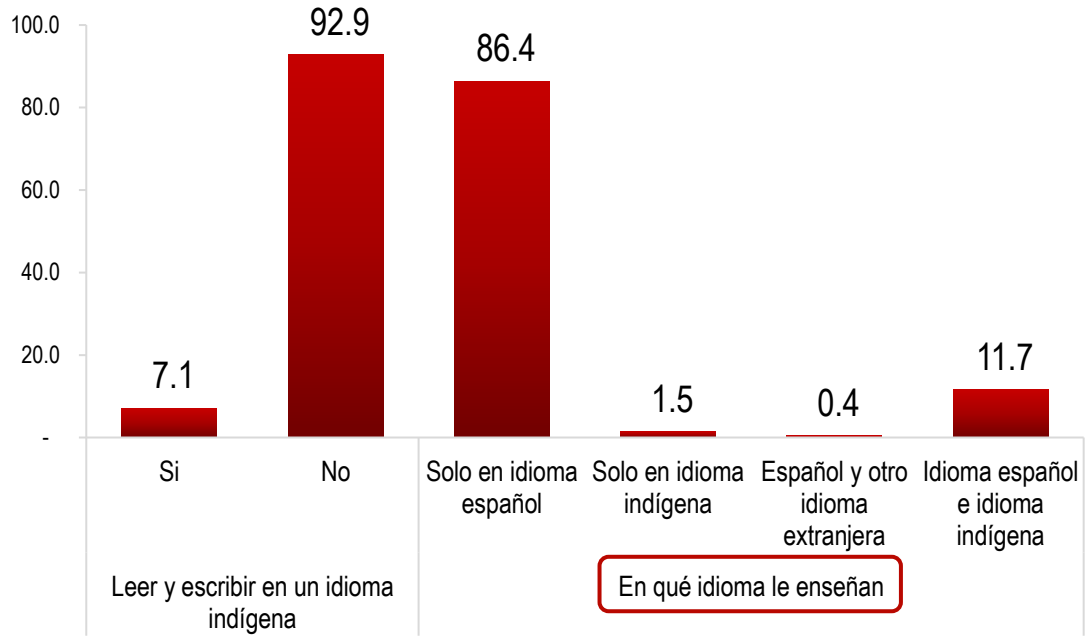
Idioma materno	Porcentaje del total	Número de idiomas	Porcentaje del total
Español	69.03	Sólo un idioma	76.4
Q'eqchi'	9.64	Bilingüe	23.1
K'iche'	7.59	Trilingüe	0.5
Mam	4.80	Políglota	0.0
Kaqchiquel	2.16	**Tipo de idioma**	**Porcentaje del total**
Xinka	0.02	Sólo español	65.0
Garífuna	0.01	Español y un idioma maya, garífuna o xinka	21.7
Otros	6.70	Solo un idioma maya, garífuna o xinka	11.2
		Otros	2.1

Fuente: Icefi/Save the Children, con base en datos del *XII Censo de población y VII de vivienda* (Guatemala, 2018)

Según la *Encuesta Nacional de Condiciones de Vida* realizada en 2014 (Encovi 2014), el 7.1% de los niños y niñas entre las edades de 7 y 17 años afirmaron saber leer y escribir en un idioma indígena, de estos más de la mitad es en idioma Q´eqchi´ (53.7% era en idioma), un 16.4% en idioma K'iche' y 12.7% en idioma Mam; conjuntamente representan el 82.8% del

total. Por el lado de la enseñanza, el 11.7% de las niñas, niños y adolescentes dijeron que aprendían en sistema mixto (idioma español e indígena) y el 1.5% sólo en idioma indígena.

Gráfica 2. Guatemala: niñas, niños y adolescentes de 7 a 17 años según lectura o escritura en un idioma indígena y el idioma en el que se le enseña en un centro educativo (2014)



Fuente: Icefi/Save the Children, con base en datos la Encuesta Nacional de Condiciones de Vida (Encovi 2014).

2.2 Principales indicadores de educación de la población guatemalteca

2.2.1 Alfabetismo y años de escolaridad total

La educación en el país pareciera que ha sido históricamente negada a buena parte de la población, ya que el 19.2% de la población de 15 años y más que no sabe leer ni escribir. Este dato se acentúa según algunas características, y crece en el caso de la mujer rural o mujer maya, los segmentos de la población con menor porcentaje de alfabetas (66.6% y 62.7% respectivamente). Los hombres urbanos y ladinos son el segmento de la población con mayores niveles de alfabetización, el 91.1% de los hombres urbanos son alfabetas y el 90.4% de los hombres ladinos saben leer y escribir.

Tabla 5. Guatemala: Población alfabeta y número de años de escolaridad. De 15 y más años de edad (2018)

Área de residencia y pueblo de pertenencia		Alfabeta			Años de escolaridad		
		Total	Hombre	Mujer	Total	Hombre	Mujer
Área	**Total**	**80.8**	**85.3**	**76.7**	**6.2**	**6.6**	**5.9**
	Urbana	87.5	91.1	84.3	7.6	7.9	7.2
	Rural	72.0	77.9	66.6	4.5	4.9	4.1
Pertenencia étnica	**Total**	**80.8**	**85.3**	**76.7**	**6.2**	**6.6**	**5.9**
	Maya	69.8	77.8	62.7	4.6	5.2	4.0
	Otro pueblo indígena	79.9	84.0	76.0	5.3	5.6	4.9
	Ladino	88.2	90.4	86.2	7.4	7.6	7.1
	Extranjero	95.5	96.7	94.3	10.8	11.4	10.4

Fuente: Icefi/Save the Children, con base en datos del *XII Censo de población y VII de vivienda* (Guatemala, 2018).

2.2.2 Alfabetismo y años de escolaridad juvenil

Aunque en menor medida que el analfabetismo total (todas las personas de 15 años y más), Guatemala también enfrenta desafíos relacionados al analfabetismo juvenil (15-24 años). En 2018 se estimó que el porcentaje de jóvenes que saben leer y escribir en el país era de 94.6% mientras que, en promedio, son las jóvenes mayas quienes tienen la menor tasa de alfabetización con un 90.3%, respecto a los jóvenes urbanos donde el 97.0% son alfabetas.

El número promedio de años de estudio de este segmento de la población (jóvenes de 15 a 24 años) es mayor que el total nacional, alcanzando 7.5 años de estudio. Al respecto, existen desigualdades significativas; mientras que en promedio los hombres y mujeres urbanas tiene 8.5 años de estudio, los hombres rurales tienen 6.7 años y las mujeres rurales 6.2 años de estudio.

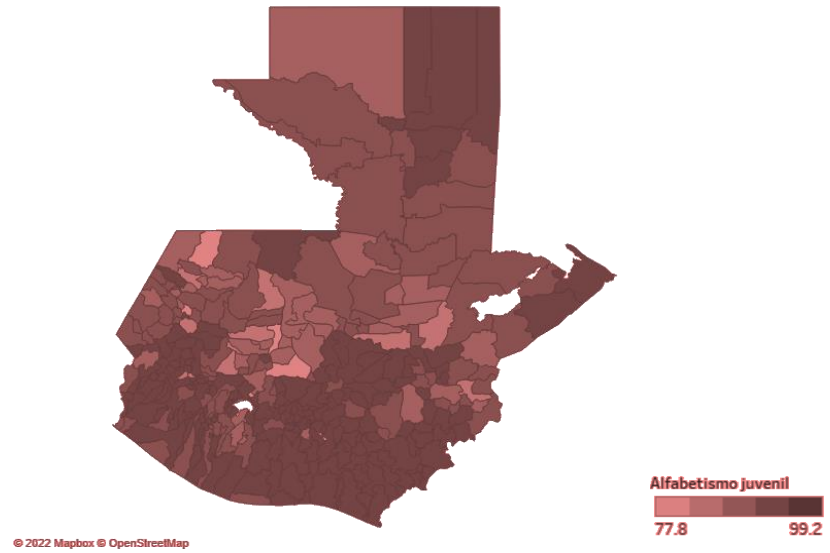
Tabla 6. Guatemala: Población alfabeta y número de años de escolaridad – de 15 a 24 años

Variable		Alfabeta			Años de escolaridad		
		Total	Hombre	Mujer	Total	Hombre	Mujer
Área	**Total**	**94.6**	**95.4**	**93.9**	**7.5**	**7.7**	**7.4**
	Urbana	96.7	97.0	96.3	8.5	8.5	8.5
	Rural	92.3	93.5	91.2	6.5	6.7	6.2
Pertenencia étnica	**Total**	**94.6**	**95.4**	**93.9**	**7.5**	**7.7**	**7.4**
	Maya	92.0	93.8	90.3	6.5	6.8	6.3
	Otro pueblo indígena	94.6	94.8	94.3	7.0	7.1	6.9
	Ladino	96.7	96.6	96.7	8.3	8.3	8.4
	Extranjero	97.7	98.1	97.4	9.1	9.1	9.1

Fuente: Icefi/Save the Children, con base en datos del *XII Censo de población y VII de vivienda* (Guatemala, 2018)

La ubicación geográfica en la que se reside también es un factor y referente de menores oportunidades básicas para las juventudes y niñez guatemalteca. Las desigualdades en oportunidades educativas coinciden en regiones con mayor pobreza y desigualdad, por ejemplo, el caso del segmento de jóvenes entre 15 y 24 que no saben leer ni escribir alcanza hasta el 22.2% en el municipio de San Mateo Ixtatán, en el departamento de Huehuetenango; el 21.9% en el municipio de Joyabaj en el departamento de Quiché; y, 17.3% en Jocotán, en el departamento de Chiquimula. Mientras que, en municipios del departamento de Guatemala, como Mixco, Santa Catarina Pínula y Guatemala, el analfabetismo alcanza al 1.0% del total de jóvenes de estas edades.

Mapa 1. Guatemala: Alfabetismo juvenil por municipio (edades de 15 a 24 años), año 2018

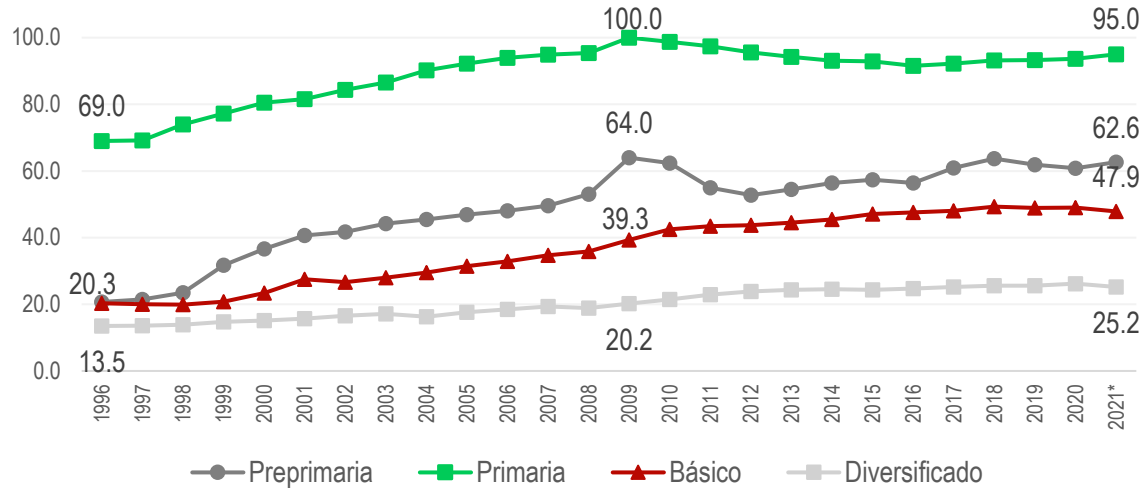


Fuente: Icefi/Save the Children, con base en datos del *XII Censo de población y VII de vivienda* (Guatemala, 2018)

2.2.3 Matricula neta estudiantil

La cobertura educativa en sus diferentes niveles, según lo demuestran las cifras, si bien ha mejorado, se ha caracterizado como un lento proceso y mejora durante las últimas décadas. En 2021 la matrícula neta que muestra la relación de niños, niñas y adolescentes matriculados en un nivel educativo respecto de grupo de edad oficial correspondiente a ese nivel, es de 95.0% matriculados en nivel primaria, siendo este el nivel educativo con mejor cobertura. Por su parte, el nivel preprimario cuenta con 62.6% de cobertura neta, mientras que solo el 47.9% de los estudiantes que terminaron la primaria alcanzan el ciclo básico y únicamente el 25.2% de las y los adolescentes en edad de estudiar alcanzaron el ciclo diversificado.

Gráfica 3. Guatemala: Tasa neta de cobertura – En porcentajes, periodo 2000 a 2021



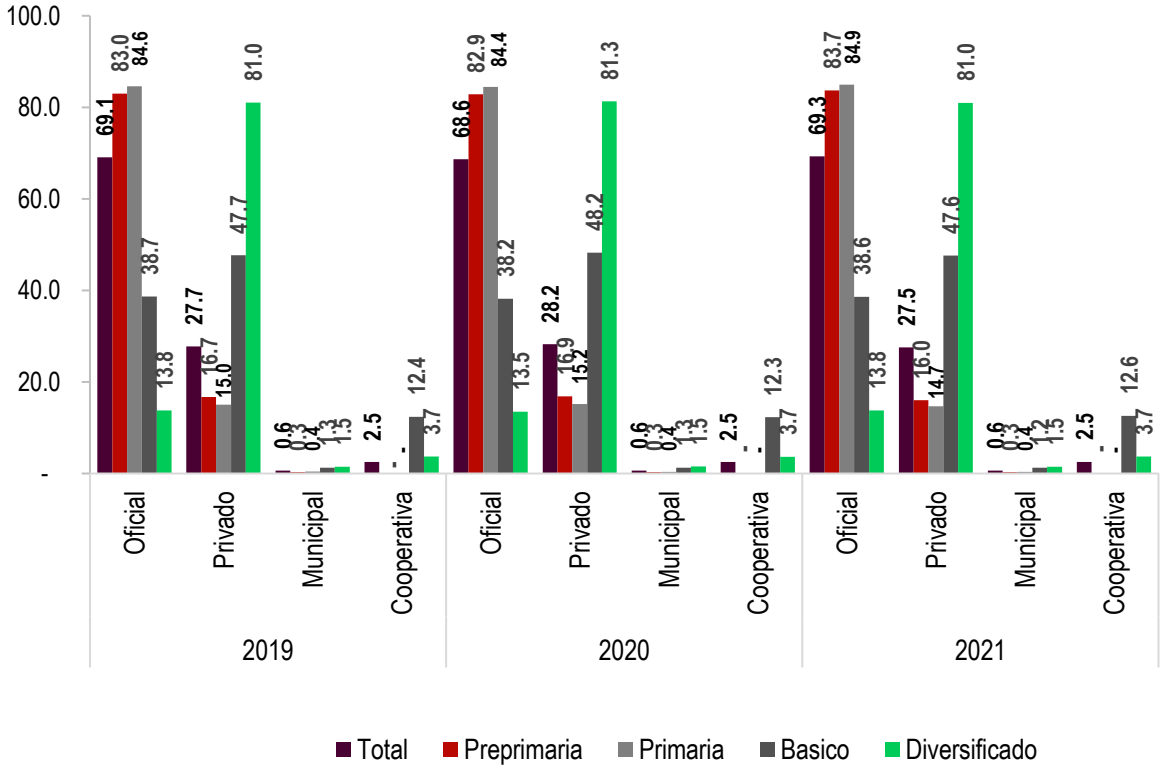
Fuente: Icefi/Save the Children, con base en anuario estadístico y estadísticas preliminares al 09 de mayo de 2022 del Mineduc; proyecciones de población oficiales estimadas con el Censo de 2018

2.2.4 Establecimientos educativos y estudiantes inscritos por nivel y sector

Según cifras del Mineduc, en 2021 el total de establecimientos educativos destinados a la educación de niñas fue 48,956, de los cuales el 69.3% es del sector oficial, el 27.5% privado, 2.5% por cooperativa y 0.6% municipal. Según nivel educativo, el 84.9% de los establecimientos disponibles son para nivel primaria y 83.7% para el nivel preprimaria, siendo los destinados al diversificado los de menor presencia del sector oficial con 13.8%

del total. De 10 establecimientos para nivel diversificado, 8 son privados. En el nivel básico, el 12.6% son establecimientos por cooperativa.

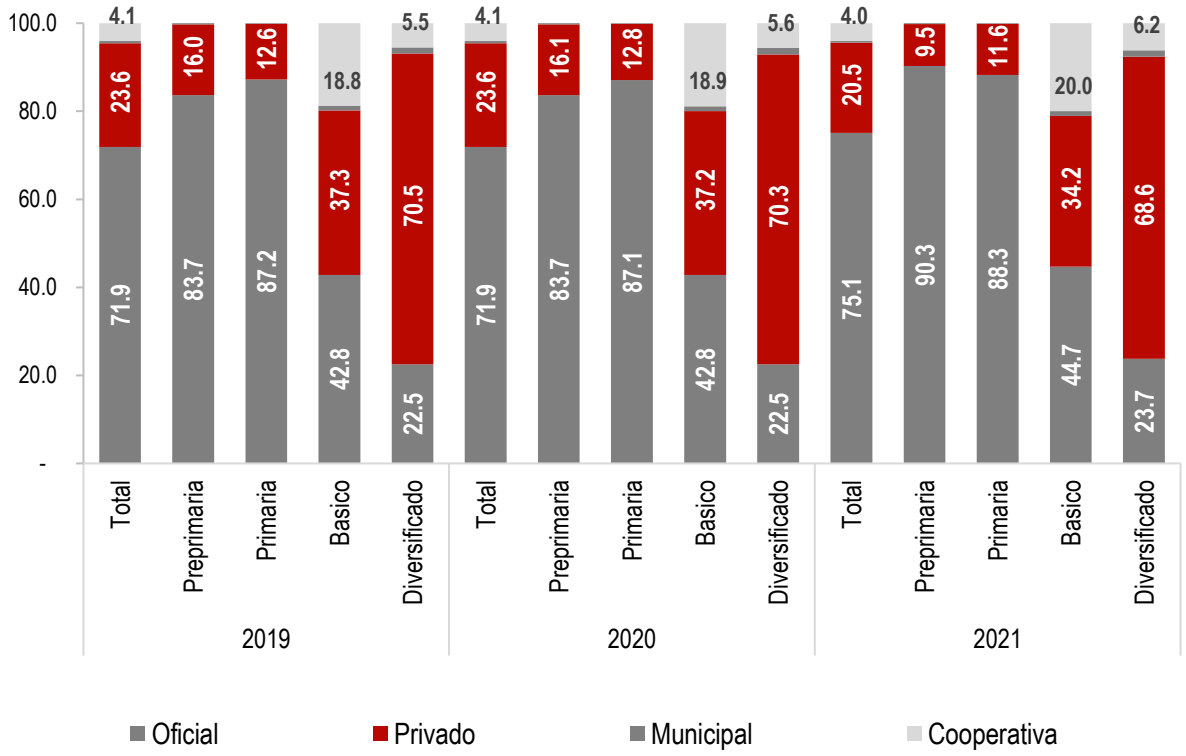
Gráfica 4. Guatemala: Porcentaje de establecimientos educativos por nivel y sector (2019-2021)



Fuente: Icefi/Save the Children, con base en anuario estadístico y estadísticas preliminares al 09 de mayo de 2022 del Mineduc.

En 2021 un total de 4,157,575 niñas, niños y adolescentes cerraron el ciclo educativo, de los cuales 75.1% asistieron a un establecimiento oficial, 20.5% a uno privado, 4.0% a uno por cooperativa y 0.4% a uno municipal. En 2021, el 90.3% de los estudiantes de preprimaria y el 88.3% de primaria asistieron a un establecimiento oficial, mientras que solo el 23.7% de los estudiantes de diversificado asistieron a un establecimiento oficial. Es decir, casi 7 de cada 10 estudiantes del nivel diversificado asisten a un establecimiento privado. En el nivel básico, los establecimientos por cooperativa han cobrado importancia durante los últimos años, dando cobertura al 20.0% de los estudiantes. En 2021, el número de estudiantes atendidos en el sector oficial aumentó respecto a los registrados en 2020 y 2019.

Gráfica 5. Guatemala: Porcentaje de estudiantes por nivel educativo y sector (2019-2021)



Fuente: Icefi/Save the Children, con base en anuario estadístico y estadísticas preliminares al 09 de mayo de 2022 del Mineduc

2.2.5 Población fuera de la escuela

Otro de los grandes retos educativos de Guatemala, es la permanencia de las y los estudiantes en las escuelas. En 2021, se estima que 1.2 millones de niñas, niños y adolescentes entre 5 a 18 años estuvieron fuera de una escuela. Entre este grupo etario, se evidencia que conforme aumenta la edad el problema se hace más común, ya que el 62.3% de las y los adolescentes entre las edades de 16 a 18 años no asistan a un instituto o colegio.

Conceptualización del abandono o deserción escolar

Según la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (Unesco), el abandono escolar como indicador se refiere al «[…] porcentaje de estudiantes que, habiendo estado matriculado en un año, deja de estudiar y no vuelve a matricularse en el siguiente año» (Unesco, 2020). Sin embargo, al abandono escolar también se le suele asociar con la deserción escolar, lo que, si bien podría considerarse un sinónimo, teóricamente representa una definición más amplia como se explica a continuación:

> «Los *desertores escolares* en la actualidad –así como también quienes se encuentran en riesgo de abandonar de forma definitiva el sistema escolar– son entendidos como aquellos niños, adolescentes y jóvenes que no se han matriculado en ninguna escuela (Lamb y Markussen, 2011; Unesco, 2012). Esta categoría incluye a quienes han decidido abandonar la escuela voluntariamente, al mismo tiempo que los estudiantes que han sido expulsados o forzados a desertar por motivos de fracaso escolar o conducta antisocial (Bradley y Renzulli, 2011; Ramírez, 2020). Entre los desertores escolares existen estudiantes con problemas de comportamiento –por ejemplo, conductas criminales–, discapacidad física, dificultades de carácter psicológico, limitaciones intelectuales, adicción a las drogas o paternidad temprana, entre otras muchas cuestiones (Fernández et al., 2016, 2011; Hawkins, Jaccard y Needle, 2013).» Citado de (Espinoza Díaz, *et al*, 2021: 117)

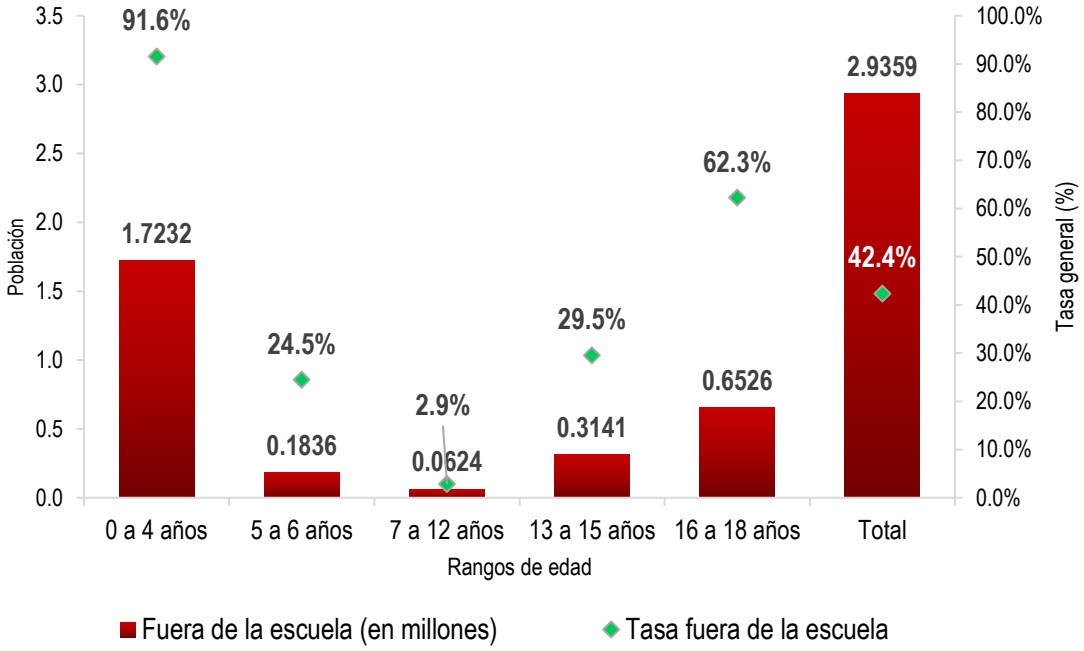
Asimismo, Espinoza Díaz, *et al,* mencionan, citando a D'Alessandre y Mttioli (2015) y Sabates, *et al* (2010), que «El abandono de los estudios puede ser de carácter temporal o definitivo, siendo la primera antesala del segundo. La deserción definitiva se ha entendido como el proceso de alejamiento y abandono, voluntario o forzoso, de la escuela, en cuanto espacio cotidiano, esto por circunstancias internas o externas al estudiante» (Espinoza Díaz, et al, 2021: 118). Por lo tanto, la deserción escolar involucra una dinámica de factores que afectan al estudiante, al sistema educativo y a la sociedad, puesto que impacta directamente en el derecho a la educación del individuo y representa serias repercusiones en su desarrollo humano.

Para el caso de Guatemala, el Ministerio de Educación incluye definiciones para estos conceptos en los glosarios de los anuarios estadísticos de educación:

a. Ausentismo escolar: es el número de personas que, de acuerdo con su edad, no se han matriculado en un año determinado. Se produce cuando el niño o la niña en edad de asistir a la escuela no lo hace. (Mineduc, 2016)
b. Deserción o abandono: acción de alumnos y alumnas al retirarse de la escuela y no finalizar el año escolar por diferentes motivos, y al no continuar en el año siguiente quedan fuera del sistema escolar. (Mineduc, 2016)

Además, tomando en cuenta a niños y niñas de la primera infancia entre las edades de 0 a 4 años, se demuestra desatención pública y privada a espacios de aprendizaje temprano, a pesar de que existen evidencias y experiencias de la importancia que tiene la educación inicial para el desarrollo integral de un ser humano en las siguientes etapas de vida. En el caso de Guatemala, a 2021 el 91.6% de la niñez en estas edades no es parte del sistema educativo.

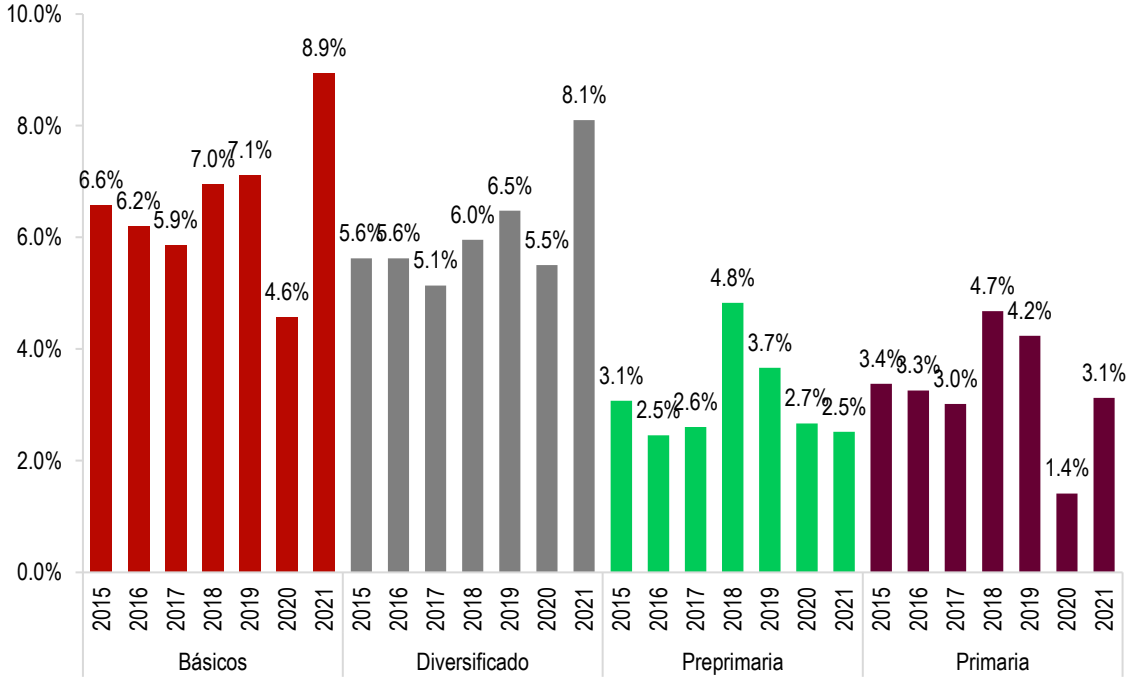
Gráfica 6. Guatemala: Estimación de la Población de 0 a 18 años fuera de la escuela en 2021



Fuente: Icefi/Save the Children, con base en el Anuario Estadístico y estadísticas preliminares al 1 de junio de 2021 del Mineduc; proyecciones de población oficiales estimadas con el Censo de 2018

El problema de la deserción escolar lo padecen todos los niveles educativos, sin embargo, las evidencias de los últimos años indican que son las y los adolescentes de los niveles básicos y diversificados quienes en mayor porcentaje abandonan la escuela. El caso más agudo se registró en 2021, en promedio el 8.9% de la matrícula del nivel básico abandonó los estudios, mientras que de los que ingresaron al nivel diversificado desertó el 8.1%. El origen o causa del abandono escolar puede explicarse por varios factores, principalmente económicos, por ejemplo, en 2021 con secuelas al alza, como consecuencia de la pandemia del COVID-19, lo que está afectando principalmente a las niñas, los niños y los adolescentes en las condiciones más desfavorables.

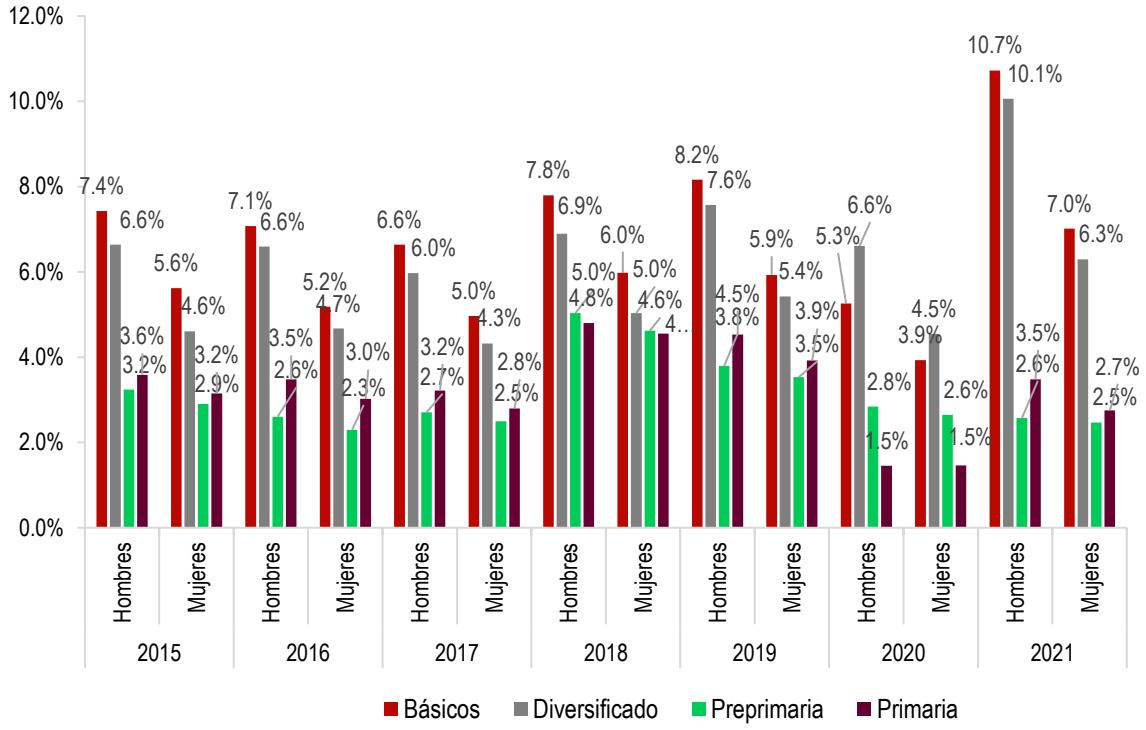
Gráfica 7. Guatemala: Deserción según nivel educativo, en porcentaje del total matriculado (2015 a 2021)



Fuente: Icefi/Save the Children, con base en anuario estadístico y estadísticas preliminares al 09 de mayo de 2022 del Mineduc.

Si bien, son las mujeres quienes presentan mayores rezagos educativos a nivel nacional, una vez dentro del sistema educativo se mantiene su matrícula respecto a los hombres, quienes en mayor porcentaje abandonan la escuela. Aunque este indicador se ha reducido levemente durante los últimos años, durante el periodo de 2015 a 2021 en promedio 7 de cada 100 adolescentes hombres abandonan el ciclo básico o el ciclo diversificado anualmente, mientras que en promedio 5 de 100 mujeres abandonan en los mismos niveles educativos. A nivel primario, 4 de cada 100 niños abandonan la escuela en promedio anualmente, y en el caso de las niñas el promedio es de 3 de cada 100 niñas que abandona el mismo nivel educativo. Aunque no existe una explicación registrada, la incorporación por parte de los hombres a una edad temprana al mercado laboral puede ser uno de los explicativos principales de este fenómeno. Para 2021, las cifras de deserción aumentaron considerablemente, principalmente en hombres en los niveles básicos y diversificado, aumentando al 10.7% y 10.1% respectivamente.

Gráfica 8. Guatemala: Deserción por nivel educativo según sexo (2015 a 2021)

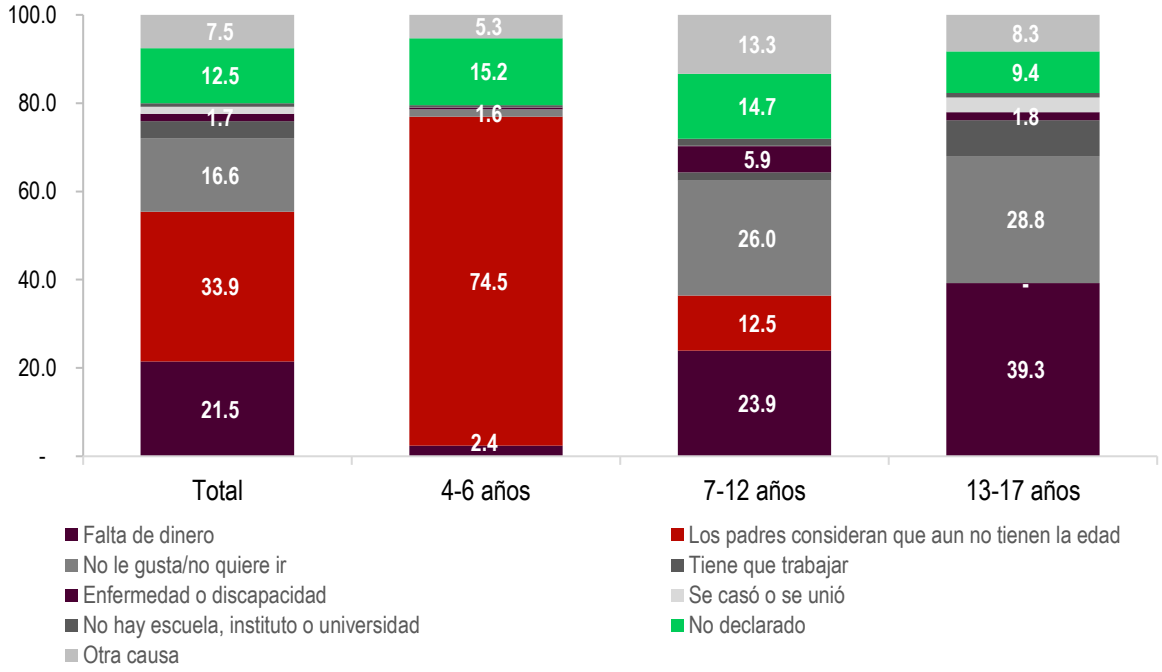


Fuente: Icefi/Save the Children, con base en anuario estadístico y estadísticas preliminares al 9 de mayo de 2022 del Mineduc.

2.2.6 Niñez y adolescencia fuera de la escuela

Existen diferentes causas por las que niñas y niños abandonan la escuela o nunca asisten a una, las cuales varían dependiendo de la edad, sexo, el grado escolar, la ubicación geográfica, cultura o condición económica a la que pertenecen y, aunque se da mayormente en los niveles básicos y diversificado, este trágico hecho resulta ser común en todos los niveles educativos. En 2018, según el XII Censo de población y VII de vivienda (Guatemala, 2018), de cada 100 estudiantes entre 4 a 17 años que no asistió a la escuela: 21.5% lo hizo por falta de dinero; 33.9% porque los padres consideran que aún no tienen la edad; 16.6% porque no le gusta o no quiere ir; 3.9% por motivos de trabajo; 1.7% porque tiene una enfermedad o discapacidad; y, 1.5% porque se casó o se unió.

Gráfica 9. Guatemala: Causa principal por la que no asiste a un establecimiento educativo en 2018 niños, niñas y adolescentes entre 4 y 17 años.



Fuente: Icefi/Save the Children, con base en datos del *XII Censo de población y VII de vivienda* (Guatemala, 2018)

La edad, la ubicación geográfica y la pertenencia étnica son condicionantes para la inasistencia a un centro educativo. Para las primeras edades entre 4 y 6 años, la principal causa por la que un niño niña no asiste a un centro educativo es porque los padres consideran que es demasiado joven. Atrasar el ingreso a la escuela de un niño o niña en esta etapa de su vida ha sido una decisión común en los hogares guatemaltecos, y tiene implícita connotaciones económicas y culturales, además de que la oferta pública en educación preescolar y de cuidados para la primera infancia aun es restringida en algunas ubicaciones geográficas. Por otra parte, también es importante resaltar que entre las y los adolescentes entre 13 y 17 años, la razón principal por la que se encuentran fuera de la escuela es por falta de dinero.

Otras características específicas es que las niñas y las adolescentes son quienes presentan mayor proporción de no asistir a la escuela por realizar quehaceres del hogar o porque se casó o se unió. Mientras que son los niños y los adolescentes quienes en mayor proporción no asisten a un establecimiento educativo por motivos de trabajo.

Tabla 7. Guatemala: Causa principal por la que no asiste a un establecimiento educativo en 2018 adolescentes entre 13 y 17 años

Concepto		Total	Falta de dinero	Trabajar	No hay escuela, instituto o universidad	Los padres / pareja no quieren	Quehaceres del hogar	No le gusta / no quiere ir	Enfermedad o discapacidad	Se casó o se unió	Otra
	Total	**100.0**	**39.3**	**8.1**	**1.0**	**1.7**	**1.9**	**28.8**	**1.8**	**3.4**	**14.1**
Área	Urbana	100.0	35.8	9.8	0.4	1.4	1.8	28.3	2.3	3.1	17.2
	Rural	100.0	41.1	7.2	1.2	1.8	2.0	29.0	1.6	3.5	12.5
Pertenencia	Maya	100.0	41.0	8.0	0.9	1.9	1.8	28.8	1.4	3.3	12.9
	Otro pueblo indígena	100.0	40.2	7.2	1.0	2.4	2.4	31.3	2.7	2.8	10.0
	Ladino	100.0	36.7	8.3	1.0	1.4	2.1	28.7	2.4	3.4	16.1
	Extranjero	100.0	17.9	11.2	1.9	1.2	1.8	17.5	1.2	3.2	44.1
Sexo	Hombre	100.0	38.9	12.4	0.8	1.0	0.3	30.4	2.0	0.8	13.4
	Mujer	100.0	39.6	4.5	1.1	2.3	3.3	27.5	1.7	5.5	14.7

Fuente: Icefi/Save the Children, con base en datos del *XII Censo de población y VII de vivienda* (Guatemala, 2018)

Los datos evidencian que las razones por las que no asisten a un establecimiento educativo tienen connotaciones regidas por las normas sociales y los roles de género. El 3.3% de las adolescentes dijeron que dejaron la escuela por realizar quehaceres del hogar, mientras que únicamente el 0.3% de los adolescentes no están inscritos por esa razón. El 5.5% de las adolescentes dejaron la escuela porque se casaron o unieron, mientras que sólo 0.8% de los adolescentes reportaron esa razón. En contraste, el 12.4% de los adolescentes dejaron la escuela por un empleo remunerado, mientras que solo 4.5% de las adolescentes reportaron esa razón.

2.2.7 El acceso a herramientas tecnológicas de las y los estudiantes

El acceso a herramientas tecnológicas en la etapa escolar supone ampliar el alcance a mejores opciones de estudio y es considerada facilitadora para nuevas metodologías de aprendizaje, situación que se evidenció durante la pandemia. A partir de 2020, el acceso y uso de tecnologías se volvió indispensable, principalmente porque se incrementó significativamente la educación a distancia. Sin embargo, la última medición al respecto evidenció que una fracción reducida de la niñez y de la adolescencia guatemalteca cuenta

con los medios tecnológicos necesarios. En 2018, a nivel nacional únicamente el 31.6% de los niños, niñas y adolescentes que asistieron a un establecimiento educativo contaban con celular, el 24.1% con una computadora y el 26.1% tenían acceso a internet.

Además de que las cifras de acceso a tecnologías son reducidas, también existen desigualdades muy marcadas, principalmente según la ubicación geográfica, pues únicamente el 10.1% de la niñez y adolescencia tenía una computadora en el área rural, respecto a que el 36.3% en el área urbana. Similar razón respecto al acceso a internet, en el área rural sólo el 11.7% y 38.7% en el área urbana.

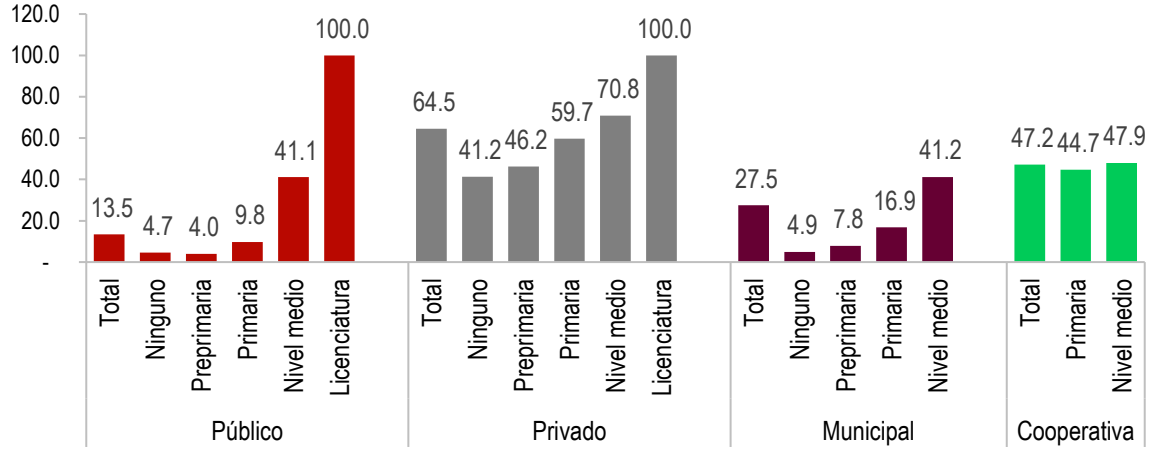
Tabla 8. Guatemala: Acceso a servicios tecnológicos de niños, niñas y adolescentes de 7 a 17 años que asistieron a un establecimiento educativo en 2018

Concepto		Celular	Computadora	Internet
Área	**Total**	**31.6**	**24.1**	**26.1**
	Urbana	42.3	36.3	38.7
	Rural	19.4	10.1	11.7
Pertenencia	Maya	20.3	14.6	14.7
	Otro pueblo indígena	21.3	13.0	15.4
	Ladino	40.5	31.5	34.9
	Extranjero	57.5	56.4	59.8
Sexo	Hombre	32.7	23.9	26.1
	Mujer	30.5	24.3	26.0

Fuente: Icefi/Save the Children, con base en datos del *XII Censo de población y VII de vivienda* (Guatemala, 2018)

Por nivel educativo y establecimiento público y privado, quienes reportaron menor acceso a una computadora fueron niños y niñas de los niveles preprimaria y primarias, inscritos en establecimientos públicos. De estos, únicamente 4 de cada 100 niños y niñas de preprimaria contaban con una computadora y 10 de cada 100 niños y niñas en el nivel primario. De manera contraria, quienes afirmaron tener mayor acceso a computadoras, fueron adolescentes de los niveles medios y universitarios de establecimientos privados. En promedio en los establecimientos privados, 71 de cada 100 adolescentes tenían acceso a una computadora, mientras que el 100% de los niveles de licenciatura en la universidad pública o privada.

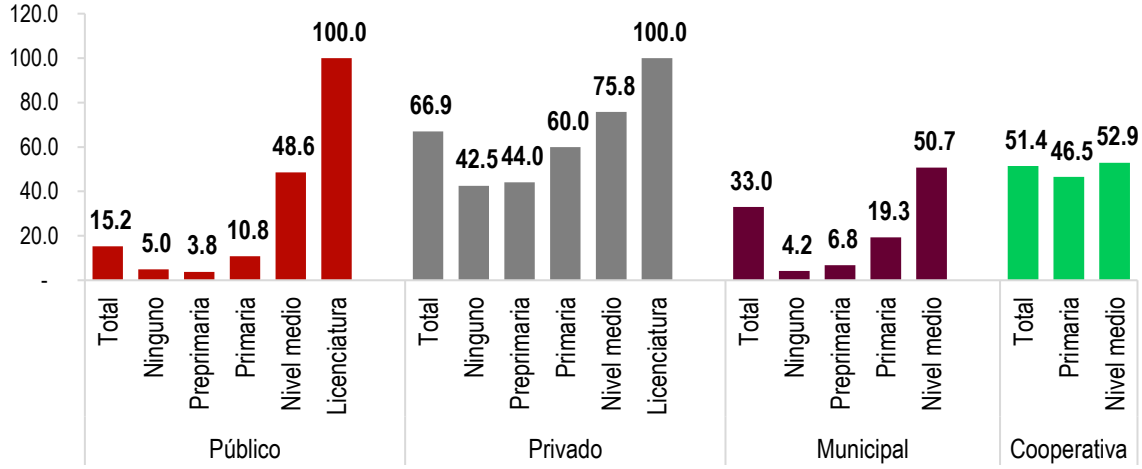
Gráfica 10. Guatemala: Acceso a computadora en niños, niñas y adolescentes de 7 a 17 años que asistieron a un establecimiento educativo en 2018



Fuente: Icefi/Save the Children, con base en datos del *XII Censo de población y VII de vivienda* (Guatemala, 2018)

El acceso desigual a internet es otro de los inconvenientes que enfrenta el sistema educativo actual, pues la capacidad de conectarse y poder desarrollar actividades en línea le es negada a la gran mayoría de niños, niñas y adolescente. En 2018, únicamente 11 de cada 100 niños y niñas de primaria en un establecimiento público contaba con acceso a internet, mientras que 60 de cada 100 estudiantes de este nivel educativo en un establecimiento privado tenían acceso.

Gráfica 11. Guatemala: Acceso a servicios de internet en niños, niñas y adolescentes de 7 a 17 años que asistieron a un establecimiento educativo en 2018



Fuente: Icefi/Save the Children, con base en datos del *XII Censo de población y VII de vivienda* (Guatemala, 2018).

2.3 Impactos en educación derivado de la pandemia del COVID-19

Existen efectos que ya han sido medidos, en cuanto a los riesgos que se estarían produciendo por el tiempo que dejaron de asistir las y los estudiantes a los establecimientos educativos, fruto de la pandemia. Aunque son escasas las cifras para evidenciar la reducción en la calidad de vida de los y las guatemaltecas debido a la pandemia por COVID-19, algunas variables indican el grado de deterioro. Según la *Encuesta Nacional de Ingresos y Egresos* (Enei 1-2021), los ingresos promedio recibidos por sueldos y salarios se redujeron en 9.1% respecto a la Enei 2-2019, pasando de Q2,429 en 2019 a Q2,207 mensual en 2021. El incremento generalizado en los precios ha hecho que el poder adquisitivo se restrinja, dada la coyuntura de choques externos sobre los precios de los energéticos y los alimentos principalmente, se estima una inflación de 7.75% al cierre de 2022, según el Banco de Guatemala,[1] la más alta en la última década. Mientras que el costo de la canasta básica alimentaria a junio 2022 fue de Q.3,311.95, muy por encima de los sueldos que en promedio se reciben como ya se mencionó en los datos de la Enei 2021.

En el caso de los indicadores de educación, en un informe publicado por el Banco Mundial (BM) en 2021, se presentan los resultados de una simulación elaborada a partir de diferentes escenarios, en los cuales se estima la pérdida de años de escolaridad (LAYS), el puntaje de las pruebas PISA y los alumnos por debajo del nivel mínimo de rendimiento (BMP) tomando en consideración 7, 10 y 13 meses de cierre de establecimientos educativos. Los resultados obtenidos son preocupantes, ante un escenario que ya superó significativamente los 13 meses de cierre de establecimientos educativos.[2] Las simulaciones realizadas por el BM muestran que, a partir de un escenario base que considera el nivel de escolaridad de Guatemala antes del 2020, de 6.3 años en promedio, pasa a 4.8 años, lo que representa una caída de 1.5 años. En este escenario se asume que los puntajes de las pruebas PISA pasan de 369 (previo el 2020) a 319, y con ello se espera que del 30% de alumnos que habrían logrado obtener un rendimiento superior al mínimo esperado, pase a sólo el 10%.

[1] Conferencia de prensa decisión de la tasa de interés líder de política monetaria. 2/9/2022. Disponible en: https://www.banguat.gob.gt/sites/default/files/banguat/Publica/conferencias/cbanguat812.pdf

[2] En marzo de 2020 se iniciaron las acciones de prevención ante la pandemia, que incluyeron el cierre de centro educativos públicos y privados. En 2022 aún se mantenía la disposición del distanciamiento social entre la población estudiantil, en vías de procurar la vacunación completa de la población menor a 18 años y con ello permitir un regreso progresivo a las clases en modalidad presencial. Sin embargo, las disposiciones aún siguen siendo parciales, con la implementación de un mecanismo de educación híbrida que incluye clases a distancia y presenciales en algunos centros educativos, según el nivel de alerta por COVID-19 en cada municipio.

Tabla 9. Guatemala: Simulación de impactos en pérdida de años de escolaridad, puntaje de pruebas pisa y nivel de rendimiento mínimo, según el tiempo de cierre de establecimientos educativos

Impacto	Pérdida de años de escolaridad (LAYS)	Puntaje PISA	Alumnos por debajo del nivel mínimo de rendimiento (BMP)
Línea base	6.3	369	70.0%
7 meses	5.5	345	81.0%
10 meses	5.1	332	86.0%
13 meses	4.8	319	90.0%

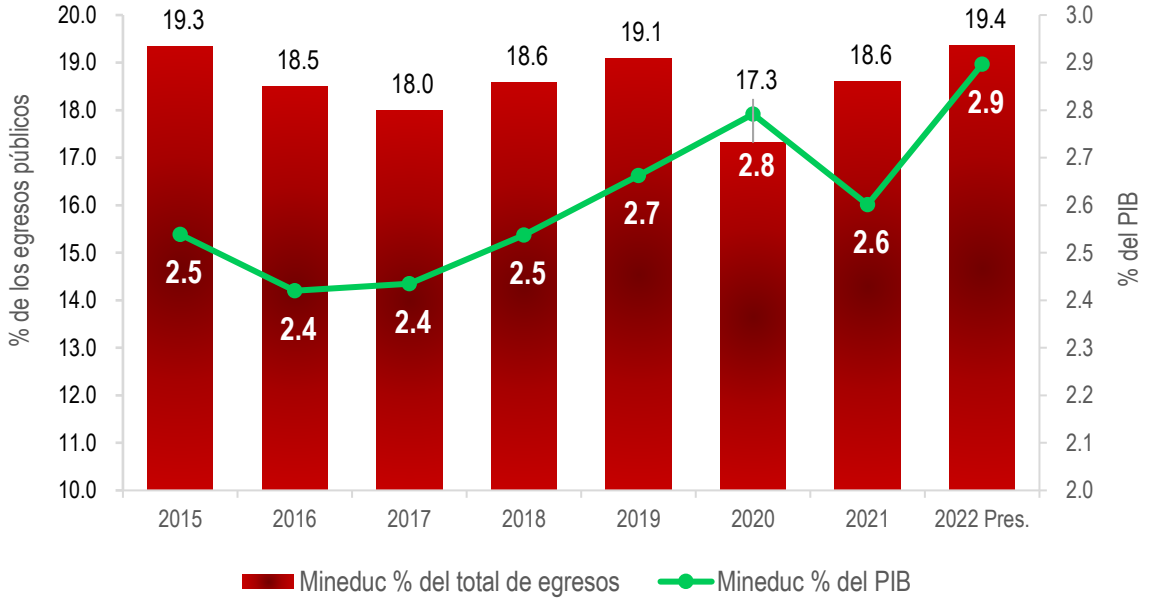
Fuente: Icefi/Save the Children, con información del informe: *Actuemos ya para proteger el capital humano de nuestros niños. Los costos y la respuesta ante el impacto de la pandemia de COVID-19 en el sector educativo de América Latina y el Caribe (2021)* Banco Mundial. Ver: https://openknowledge.worldbank.org/handle/10986/35276?locale-attribute=es

Los resultados de estas simulaciones deben tomarse en cuenta como una alerta sobre los impactos y riesgos que generan el cierre prolongado de los centros educativos, y su impacto en el contenido y calidad académica. El Gobierno enfrenta un enorme reto para aminorar estos efectos de la pandemia, lo que también involucra la necesidad por destinar recursos financieros y una planificación de acciones para atender a toda la niñez y adolescencia que ha sufrido los efectos del cierre de establecimientos educativos y la pérdida de clases. También, resulta necesario y urgente que las autoridades de educación implementen acciones directas para recuperar la matrícula estudiantil de las y los estudiantes que, dadas las circunstancias antes expuestas, han abandonado la escuela.

2.4 Inversión pública en educación por parte del Gobierno central

En Guatemala la entidad del sistema de educación público con mayor peso presupuestario es el Mineduc, en promedio durante los últimos 8 años con el 18.6% del total egresos, ubicándose solo después de las Obligaciones del Estado a Cargo del Tesoro (OEACT), como la segunda entidad con mayor presupuesto. En estos últimos ocho años, el presupuesto del Mineduc no ha superado el 3.0% del Producto Interno Bruto (PIB).

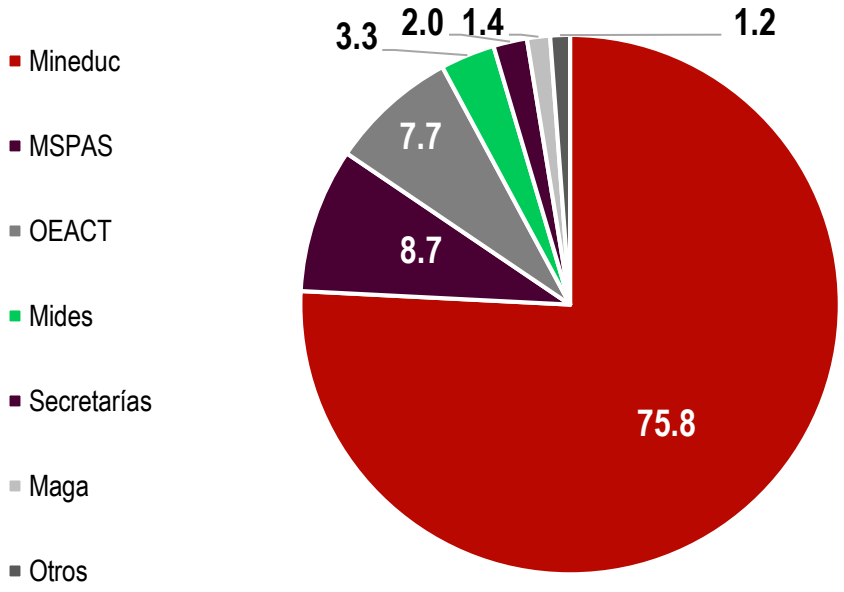
Gráfica 12. Guatemala: gasto ejecutado por el Ministerio de Educación, como proporción de los egresos totales del Gobierno central y del PIB, 2015 – 2021 ejecutado, presupuesto 2022



Fuente: Icefi/Save the Children, con base a cifras oficiales del Minfin (Sicoin).

En la actualidad, la mayor parte de la inversión pública en niñez y adolescencia (IPNA) se destina a la educación. En 2021 la IPNA representó el 23.5% del presupuesto ejecutado del Gobierno Central, mientras que en el presupuesto vigente en 2022 se estima que representa el 25.3% del total de egresos, 75.8% del cual está asignada al Mineduc. Estos datos, si bien reflejan concentración de recursos en la educación, lamentablemente no se traduce en mayor capacidad de cobertura y calidad educativa dentro del sector público. También muestra que persisten asimetrías en la capacidad del Estado para garantizar otros derechos a la niñez y a la adolescencia, que para corregirlas no se deben recortar las inversiones destinadas a la educación, sino incrementar todas las inversiones destinadas a la niñez y adolescencia.

Gráfica 13. Guatemala: Entidades[3] con mejor nivel de ejecución de la inversión pública en niñez y adolescencia (IPNA) en el presupuesto vigente del Gobierno Central en 2022.



Fuente: Icefi/Save the Children, con base a cifras oficiales.

En cuanto a la composición programática del presupuesto, durante los últimos años el Mineduc ha ejecutado 14 programas, de los cuales 12 estaban vigentes en 2021. La inversión pública en educación se encuentra principalmente concentrada en tres programas que representan el 79.2% del total: Educación escolar primaria, con más de la mitad del presupuesto total de la institución, en promedio (2016 a 2021), 56.8% del total; Educación escolar preprimaria con el 13.1% del presupuesto; y, Apoyo adecuado de alimentos, con el 9.3%.

[3] PGN = Procuraduría General de la Nación; Mineduc = Ministerio de Educación; Micivi = Ministerio de Comunicaciones, Infraestructura y Vivienda; Mingob = Ministerio de Gobernación; MSPAS = Ministerio de Salud Pública y Asistencia Social; Mides = Ministerio de Desarrollo Social; Maga = Ministerio de Agricultura, Ganadería y Alimentación; OEACT = Obligaciones del Estado a Cargo del Tesoro; y, Micude = Ministerio de Cultura y Deportes

Tabla 10. Guatemala: Egresos del Ministerio de Educación según programa, devengados en 2016 a 2021. Cifras en millones de quetzales y como porcentaje del promedio total

Programas	2016	2017	2018	2019	2020	2021	Promedio	% del total
Educación escolar de primaria	7,228.1	7,595.3	7,950.5	8,658.9	9,265.6	9,730.8	8,404.9	56.8
Educación escolar de preprimaria	1,592.0	1,703.6	1,770.4	2,012.8	2,273.4	2,308.0	1,943.3	13.1
apoyo para el consumo adecuado de alimentos	614.4	683.8	1,241.2	1,798.9	1,996.9	1,901.3	1,372.8	9.3
Educación escolar básica	882.0	936.9	930.4	1,046.8	1,047.7	1,117.6	993.5	6.7
Actividades centrales	456.6	463.0	463.4	507.4	514.5	521.2	487.7	3.3
Partidas no asignables a programas	416.7	462.8	489.3	533.7	483.1	508.4	482.3	3.3
Educación escolar diversificada	424.1	430.5	449.7	482.8	457.8	463.2	451.4	3.0
Actividades comunes a los programas de preprimaria, primaria, básico y diversificado (prg. 11, 12, 13 y 14)	185.3	206.3	264.9	297.4	288.6	300.4	257.2	1.7
Actividades comunes a los programas de preprimaria y primaria (prg. 11 y 12)	139.8	152.2	163.1	189.8	202.1	204.9	175.3	1.2
Actividades comunes a los programas de primaria, básico y diversificado (prg. 12, 13 y 14)	125.2	78.0	157.6	177.7	77.4	149.5	127.6	0.9
Educación extraescolar	84.6	91.8	96.4	101.9	85.8	86.6	91.2	0.6
Educación inicial	-	-	-	6.8	7.9	8.3	3.8	0.0
Atención por desastres naturales y calamidades públicas	-	-	5.0	-	31.8	-	6.1	0.0
Cobertura educativa escolar en preprimaria y primaria	-	14.6	8.3	-	-	-	3.8	0.0
Total	**12,148.7**	**12,818.9**	**13,990.3**	**15,814.8**	**16,732.7**	**17,300.2**	**14,800.9**	**100.0**

Fuente: Icefi/Save the Children, con base a cifras oficiales.

En contraste, existe una fuerte ausencia pública para atender los niveles básicos y diversificado. En conjunto, en los últimos años la inversión en estos niveles educativos no ha superado el 10% del presupuesto total del Mineduc, y tampoco ha logrado mejoras significativas en los montos asignados. En el caso del programa de Educación escolar básica, en promedio representa el 6.7% del total, y el programa de Educación escolar de diversificado el 3.0%. La desatención en oferta pública para estos niveles educativos ha sido

una de las razones principales que motiva a las y los adolescentes y jóvenes no continúen su educación luego de finalizar el nivel primario. Esto representa uno de los grandes desafíos para el Estado guatemalteco en cuanto a la atención del abandono escolar en los niveles de básico y diversificado, puesto que además de las altas consecuencias en el desarrollo integral de una persona, tiene repercusiones económicas y sociales para el país. Esta situación está vinculada con las bajas asignaciones presupuestarias, falta de cobertura a nivel nacional, gestión ineficiente de la calidad educativa y control académico, y finalmente, un mal seguimiento y evaluación de la gestión por resultados para medir la calidad del gasto público y cerrar brechas para mejorar la matrícula.

Metodología:
predicción de abandono escolar

3

3.1 Antecedentes sobre el modelo que predice el abandono escolar

A pesar de que existe una amplia cobertura de los servicios de educación primaria en Guatemala, aún persisten grandes brechas en el alcance de los servicios de educación secundaria para miles de adolescentes y jóvenes. El porcentaje de estudiantes que abandonan el sistema escolar tras finalizar sexto primaria y tras finalizar tercero básico –la educación primaria y la primera etapa de educación secundaria respectivamente– es significativo, y se explica por una diversidad de factores. Para afrontar esta problemática, aparte de establecer cuáles son los determinantes que están asociados al abandono de un nivel educativo, una pregunta importante que se debe investigar es *¿quiénes son los y las estudiantes con riesgo de desertar del sistema educativo?*

Identificar correctamente a los estudiantes en riesgo de deserción, permitiría optimizar las estrategias y los recursos limitados para prevenir su deserción. «Esta pregunta puede parecer relativamente sencilla de responder, particularmente en países, regiones o localidades con alta tasa de abandono escolar [...] uno pudiese suponer que los abandonos se concentran en escuelas particularmente desventajadas, o disfuncionales, o entre estudiantes con ciertas características» (Haimovich, *et al*, 2018: 3). No obstante, el abandono escolar se distribuye a lo largo de un gran porcentaje de escuelas y no se identifican únicamente con características sencillas, dada la complejidad de factores.

Para ahondar en la metodología que identifica los estudiantes con riesgo de abandono escolar, se toma como referencia el trabajo de Haimovich, *et al.,* el cual realiza predicciones de estudiantes de Guatemala y de Honduras que estarían abandonando el sistema escolar. Emplea modelos de regresión a partir de datos administrativos longitudinales de la matrícula estudiantil, y con ello predice hasta con un 80% de exactitud qué estudiantes de sexto grado de la primaria abandonaron el sistema escolar, luego de un ciclo lectivo (períodos 2013-2014, en Guatemala y 2014-2015, en Honduras).

Estas predicciones son una herramienta para decidir en dónde centrar esfuerzos para prevenir el abandono, reduciendo sustancialmente la mala asignación de recursos y su sesgo por género, etnia o ubicación geográfica. Según el estudio, los estudiantes en riesgo de desertar se distribuyen a lo largo de muchas escuelas y localidades, es decir, no se concentran solo en algunas, y por consiguiente, la priorización de recursos a nivel de

estudiantes es mucho más eficiente que a nivel de escuelas o de municipios. Para analizar las alertas a nivel individual son necesarios sistemas de información de gestión educativa que permitan seguir individualmente a estudiantes a lo largo del tiempo. En particular, es necesario que se presenten las condiciones siguientes:

1. Deben existir sistemas de información para la gestión educativa con datos a nivel individual;

2. Dichos sistemas deberían contener un código único de identificación (CUI) para rastrear el estado e información de estudiantes a lo largo del tiempo;

3. La información en estos sistemas se debe actualizar regularmente, por lo tanto, se requiere un mecanismo para que la información fluya fácilmente de las escuelas a la entidad rectora;

4. En particular, se debe captar regularmente información que identifique si un estudiante ha abandonado el sistema escolar;[4]

5. Los sistemas de información también deberían contener un CUI por escuela, para dar seguimiento y captar las particularidades en cada centro educativo; y,

6. Capturar o integrar información de predictores potenciales de deserción escolar, por ejemplo: las notas en pruebas estandarizadas, las calificaciones en los cursos, la cantidad de días de ausencia, las variables socioeconómicas del estudiante, de la escuela o de la ubicación geográfica, el historial de suspensiones o expulsiones, la repetición de años, entre otros.

Los autores presentan brevemente conceptos y precedentes de modelos de alerta de deserción escolar procedentes de los Estados Unidos de América. Explican cuáles son los datos administrativos y socioeconómicos disponibles en Guatemala y Honduras para hacer estos modelos de predicción a nivel de estudiante, y discuten el estado de esta información, puesto que estos sistemas son relativamente nuevos. También se presentan las tasas de abandono escolar en los dos países, como parte de un indicador indispensable para iniciar

[4] «El dato más importante es si ese estudiante [que tiene un identificador único] está matriculado en una fecha específica del año escolar [...] por ejemplo, dos meses después del inicio de este» (Haimovich, *et al*, 2018: 8).

el análisis y tener la referencia base sobre la medición. Los resultados de los modelos de predicción,[5] hacen una mención breve del procedimiento empleado y muestran que el desempeño de priorizar recursos a nivel de estudiantes, según el sistema de alerta temprana, es mucho más eficiente que priorizarlos a nivel de escuela o municipio,[6] y que a nivel de escuela sigue siendo más eficiente que a nivel de municipio.

Concluyen qué intervenciones para prevenir el abandono escolar pueden ser más efectivas al apuntarse específicamente a estudiantes que tienen más riesgo de desertar, además de explicar el alcance y la fiabilidad de los datos administrativos. También, sugieren dos puntos para continuar los estudios exploratorios:

i) Mejorar el análisis de alertas tempranas puede mejorar la recolección de información y ampliarla para incluir más factores asociados a la deserción escolar.

ii) Con estos análisis de alertas tempranas se pueden crear *sistemas de alertas tempranas* para afrontar esta problemática.

Sobre este último punto también hacen algunas advertencias, que incluyen que los resultados de los análisis de alerta temprana no explican las causas del abandono escolar. Esos análisis deben hacerse con cuidado, manteniendo la confidencialidad de los estudiantes, y no deben usarse para etiquetarlos negativamente o estigmatizarlos. También es necesario estudiar más a detalle qué intervenciones realmente son efectivas.

3.2 Explicación de los modelos predictivos y de aprendizaje de máquina (*machine learning*)

Un modelo predictivo es un proceso matemático que permite observar tendencias en conjuntos de información conocida y extrapolarlas a eventos desconocidos. Para fines de prevenir el abandono escolar, un modelo predictivo efectivo: toma datos de estudiantes de ciclos anteriores; determina

[5] Construidos siguiendo la metodología de: 1) Knowles, J. (2015). Of Needles and Haystacks: Building an Accurate Statewide Dropout Early Warning System in Wisconsin." Journal of Educational Data Mining 7(3): 18-67 & 2) Stuit, D., O'Cummings, M., Norbury, H., Heppen, J., Dhillon, s., Lindsay, J. & Zhu, B. (2016). Identifying Early Warning Indicators in Three Ohio School Districts. Washington, D.C.: U.S. Departamento de Educación, Instituto de Ciencias de la Educación, Centro Nacional of Education Sciences National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest

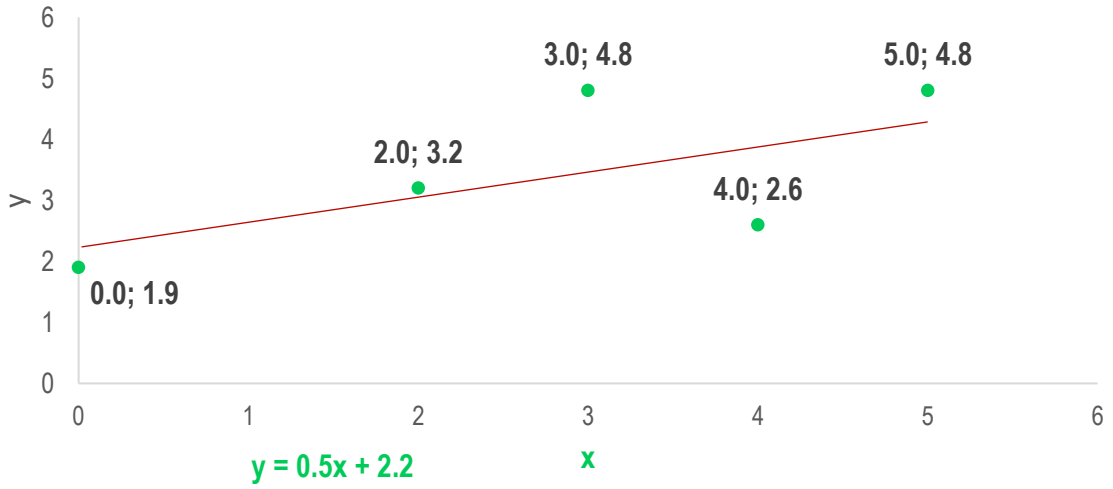
[6] Una métrica recurrente en el análisis es medir el porcentaje de falsos positivos, puesto que invertir en intervenciones para prevenir el abandono escolar en estudiantes que no iban a abandonar sería considerado un desperdicio. Por lo tanto, comparan cuánto se perdería por los falsos positivos que da el modelo con cuánto se perdería por los falsos positivos que da asignar recursos por escuela (es decir, si solo se considera una escuela entera, porque en promedio tiene tasa de deserción relativamente alta), y con cuánto se perdería por los falsos positivos por asignar recursos por municipio.

patrones estadísticos que estén asociados a la observación de si un estudiante continuó sus estudios o los abandonó; y, habiendo determinado los patrones asociados al abandono escolar, predice qué estudiantes del ciclo actual están en riesgo de desertar.

Para construir este modelo predictivo, se emplea un *algoritmo de aprendizaje de máquina supervisado*.[7] *Algoritmo de aprendizaje de máquina* se refiere a una secuencia finita de pasos o instrucciones que se dan a una computadora para que encuentre patrones dentro de un conjunto de datos.[8] El *aprendizaje* consiste en que, con esta serie de pasos, la computadora busca autónomamente el mejor patrón para describir el conjunto.

Un ejemplo de aprendizaje de máquina es una regresión lineal. Si se tiene un conjunto de datos de observaciones de una variable respuesta (o dependiente), denotada por y, y una variable predictora (o independiente), denotada por x, la regresión lineal de y en términos de x encuentra la función lineal de la forma $y = b_0 + b_1 x$ que mejor representa al conjunto de datos,[9] tal como lo muestra el ejemplo de la gráfica 14.

Gráfica 14. Ejemplo de una regresión lineal



Fuente: Icefi/Save the Children, como ejemplo ilustrado. Nota: En este caso, el conjunto de datos ilustrado en esta gráfica consiste en 5 observaciones y los parámetros óptimos para esta regresión son $b_0 = 2.2, b_1 = 0.5$.

[7] Como parte de los referentes teóricos, se sugiere consultar: https://hastie.su.domains/ISLR2/ISLRv2_website.pdf James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R* (1st ed.). Springer)

[8] Hay algoritmos de aprendizaje de máquina para clasificar variables categóricas, para hacer regresiones de variables numéricas, para crear *clústeres* (agrupaciones) de los datos o para otras aplicaciones.

[9] En el caso de un conjunto de datos que posea más variables predictoras, una regresión lineal puede generalizarse a una función de la forma $y = b_0 + \sum_i b_i x_i$.

En el ejemplo anterior, los pesos b_0 y b_1 son los que determinan cómo es la regresión lineal, es decir cuál es su pendiente y posición respecto al eje vertical (o *intercepto*). Siendo estos pesos los que representan el patrón encontrado en el conjunto de datos, se les denomina los parámetros de la regresión lineal. Teniendo estos parámetros, es posible estimar y de observaciones incompletas a partir únicamente de los valores que tengan en x. En otras palabras, sería posible hacer predicciones de la variable respuesta conociendo solo la variable predictora.

Formalizando este principio, un modelo predictivo se obtendría al encontrar los parámetros óptimos w^* tal que se minimice una medida de error $g(\varepsilon)$ que responde a la siguiente ecuación:

$$\boldsymbol{y = f(X; w) + \varepsilon,}$$

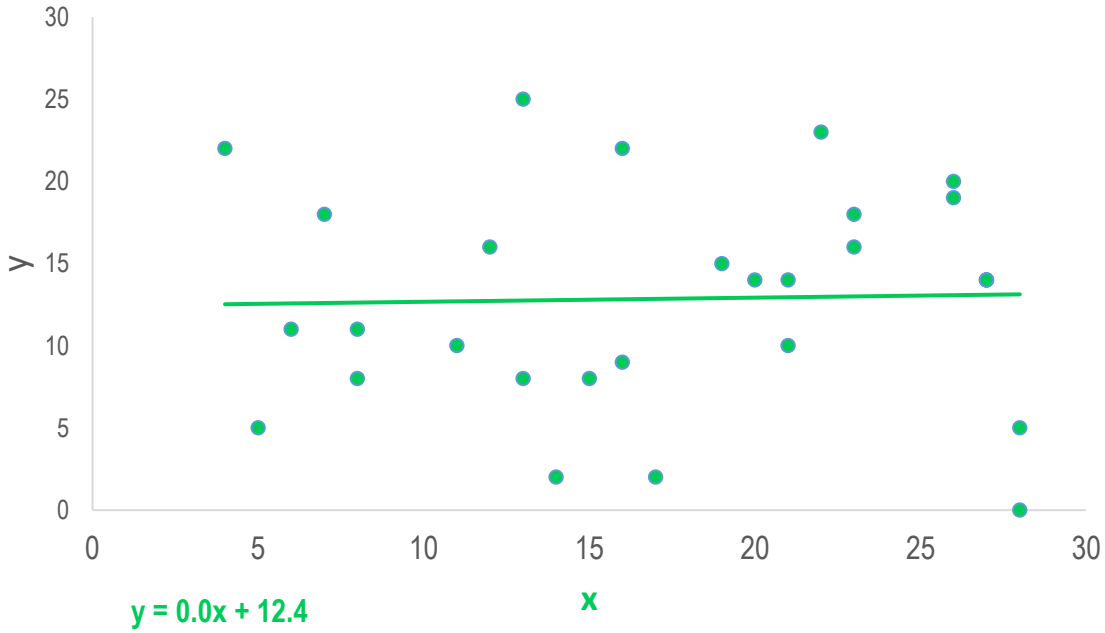
donde $\boldsymbol{y}$ es la variable respuesta, $X = \{x_1, x_2, x_3, \dots\}$ es un conjunto de variables predictoras, $\boldsymbol{f}$ es el modelo (una función que estime y), y $\varepsilon = y - f(X; w)$ es el error entre la estimación de y y y misma. Al proceso de encontrar los parámetros óptimos se le llama *ajustar* o *entrenar* un modelo predictivo. Este modelo predictivo luego puede ser empleado para hacer predicciones de y a partir de un nuevo conjunto de datos que contenga las mismas variables predictoras X siempre y cuando se suponga que el conjunto de datos sobre los que se hacen las predicciones comparte las mismas tendencias que sean estadísticamente indistinguibles al conjunto de datos original.

En el caso de la regresión lineal, la medida de error (o función de coste) que se minimiza equivale a $g(\varepsilon) = \sum_i {\varepsilon_i}^2$, es decir la sumatoria de la distancia entre cada punto y la regresión. El algoritmo con el que se *ajusta* una regresión lineal es bastante directo porque los pesos óptimos pueden calcularse analíticamente con relativa facilidad, pero otros algoritmos de aprendizaje de máquina pueden ser más complejos.

Cabe destacar que, si bien un algoritmo de aprendizaje de máquina encuentra los pesos óptimos para ajustar de la mejor forma un modelo predictivo a un conjunto de datos, no garantiza que el modelo predictivo resultante sea efectivo. Por ejemplo, en la Gráfica 15 se tiene un conjunto de datos con un alta medida de error al entrenarse una regresión lineal, lo que advierte que los datos no se ajustan a una tendencia lineal. Contrariamente, puede ser el caso que un modelo se sobreajuste a unos datos, y se encuentren tendencias que realmente se deben a *ruido estadístico*, y no necesariamente a patrones que puedan

generalizarse y usarse para predicciones (figura 4). Por lo tanto, una vez se haya entrenado un modelo, debe ser evaluado para validar si realmente es capaz de hacer predicciones, y que no esté *sub-ajustado* ni *sobre-ajustado*.

Gráfica 15. Ejemplo de una regresión lineal con poca capacidad predictiva



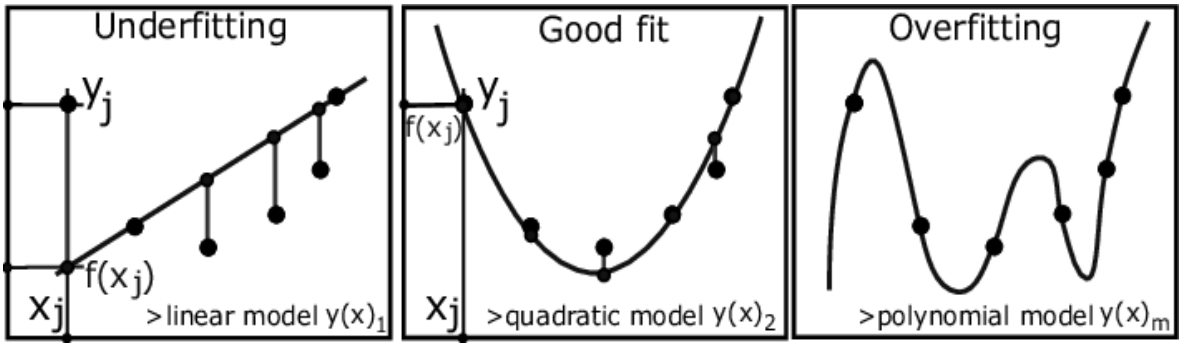
Fuente: Icefi/Save the Children, como ejemplo ilustrado. Nota: El conjunto de datos que se encuentra en esta gráfica está compuesto de 30 pares ordenados de números aleatorios, por lo que la regresión lineal no es capaz de predecir efectivamente la variable respuesta a partir de la predictora.

Evaluar un modelo predictivo para verificar su efectividad es el principio detrás de un algoritmo de aprendizaje de máquina *supervisado*. Para tal efecto, el conjunto de datos debe dividirse en dos subconjuntos estadísticamente indistinguibles:[10] un subconjunto de entrenamiento y un subconjunto de prueba. Como su nombre lo indica, con el primer subconjunto se entrena el modelo predictivo, y se verifica con la medida de error que los datos sí se ajustan al modelo que se está empleando. Luego, se suprime temporalmente la variable respuesta del subconjunto de prueba y se utiliza el modelo entrenado con el primer subconjunto para hacer estimaciones de la variable respuesta. Al comparar la variable

[10] Para cerciorarse de que sean estadísticamente indistinguibles, los subconjuntos deben generarse mediante un proceso aleatorio. Se recomienda asignar los datos a cada conjunto ya sea por lotería aleatoria simple o por asignación aleatoria estratificada para asegurar que ambos subconjuntos cuenten con características potencialmente importantes al modelo (por ejemplo, para crear el modelo predictivo de deserción escolar debe asegurarse que en ambos subconjuntos haya registros de estudiantes que continuaron y desertaron sus estudios).

respuesta real y la predicha por el modelo, se tiene una medida de qué tan extrapolable es el modelo a conjuntos de datos estadísticamente indistinguibles.

Figura 4. Sub-ajuste, ajuste correcto y sobreajuste de un conjunto de datos con tendencia cuadrática



Fuente: Tomado de Shao, Lin & Mahajan, Aishwarya & Schreck, Tobias & Lehmann, Dirk. (2017). Interactive Regression Lens for Exploring Scatter Plots. Computer Graphics Forum. 36. 10.1111/cgf.13176.

3.3 Construcción del modelo para medir el riesgo de abandono escolar

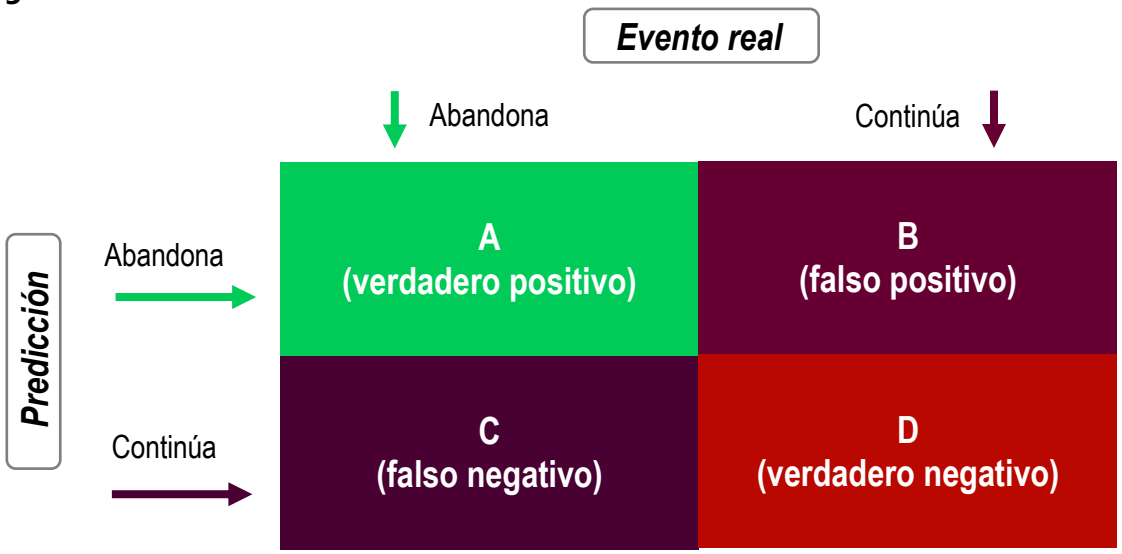
El objetivo del modelo de deserción escolar es lograr predecir si un estudiante continuará o abandonará los estudios en un determinado período, por lo que la prueba de hipótesis estadística asociada es análoga a que «el modelo es significativamente mejor para predecir si un estudiante deserta que una predicción aleatoria.» Usando la definición del abandono o deserción escolar anotada en el capítulo 1 como una variable binaria que registra un estudiante que abandona (codificado como 1) o que continúa sus estudios (codificado como 0).

El modelo funciona a partir de entrenar un algoritmo de regresión con una serie de características, también conocidas como variables predictoras, para calcular una probabilidad con valores entre 0 y 1, que indican si un estudiante presenta mayor o menor riesgo de deserción. Posteriormente, se debe definir una magnitud de corte *n*, tal que, si la probabilidad resultante de que un estudiante deserte es mayor que ese *n*, entonces se considera que desertará (codificado como 1), o que continuará sus estudios (codificado como 0).

La definición de esta magnitud *n* es la manera en la que se calibra el modelo entrenado para obtener mejores resultados. Para lograr esta calibración, es necesario comparar las

predicciones, midiendo si los estudiantes con ciertas características desertan del sistema educativo con los eventos que se espera que influyan, para luego contrastar con los datos reales. De esta comparación entre las predicciones de abandono y los eventos reales se derivan métricas para calificar al modelo, como se observa en la siguiente figura:

Figura 5. Tabla de Eventos



Estimaciones implícitas:
Exactitud = (A+D) / (A+B+C+D)
Sensibilidad = A / (A+C)
Precisión = A / (A+B)
Especificidad = D / (B+D)
Proporción de falsos positivos = 1 – Especificidad = B / (B+D)
Fuente: Icefi/Save the Children, basado en (Haimovich, *et al*, 2018)

Las mediciones a partir de la tabla de eventos resultan en diversas métricas. La sensibilidad del modelo se mide a partir de los eventos que fueron predichos como positivos y que en realidad fueron positivos, que se conocen como *verdaderos positivos* (A). Los *falsos negativos* (C) son los eventos que fueron predichos como negativos, pero en realidad fueron positivos. La sensibilidad del modelo mide la proporción de eventos realmente positivos que fueron predichos como positivos.

La precisión, se puede medir a partir de la proporción de los resultados verdaderos positivos (A), sobre la sumatoria de los mismos resultados verdaderos positivos (A) y los falsos negativos (B). Esto provee un parámetro que indica cuántos resultados asertivos se logran computar dentro del total de resultados positivos en la estimación (que incluyen resultados

falsos positivos), es decir, qué porcentaje de los resultados verdaderos positivos representa el total de resultados positivos computados.

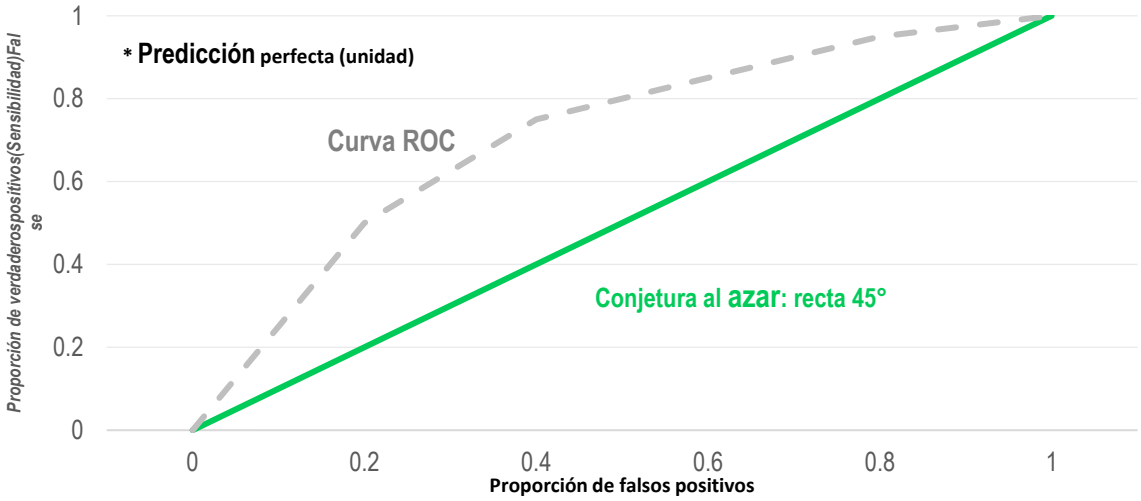
La especificidad demuestra el resultado entre los datos que representan los verdaderos negativos (D) es decir, aquellos que sí continuarían los estudios, sobre el total de la sumatoria de datos verdaderos positivos (A) y los falsos positivos (B). Esto permite determinar qué porcentaje de los resultados que indican que el estudiante va a continuar sus estudios, lo hace efectivamente sobre el total de los resultados que incluyen aquellos falsos positivos estimados.

Finalmente, la proporción de falsos positivos (B), permite observar cuánto representan estos resultados del total de datos que incluyen falsos positivos (B) y verdaderos negativos (D), lo que muestra el porcentaje de mediciones sustancialmente erróneas que predicen un abandono, cuando realmente lo que sucede es que el alumno continuó sus estudios.

Estos objetivos de maximizar la sensibilidad y minimizar los falsos positivos (también llamados *falsas alarmas*), como ya se observó, pueden estar en conflicto entre sí, ya que las predicciones que identifican a la mayoría de los desertores a menudo también pueden clasificar incorrectamente a muchos estudiantes como desertores. Los autores Haimovich, *et al,* indican que existen estudios en los que se logra una sensibilidad que alcanza más del 80%, es decir, que predicen correctamente los abandonos eventuales, pero tienen tasas de falsos positivos superiores al 50%, es decir, que su estimación resulta incorrecta en más de la mitad de los casos que eventualmente se espera que abandonen, pero en realidad continúan.

La disyuntiva entre estos dos objetivos se puede medir utilizando una *Curva de la Característica Operativa del Receptor* (ROC, por sus siglas en inglés) proveniente de la literatura de detección de señales, que simplemente traza la sensibilidad (o porcentaje de verdaderos positivos), frente al porcentaje de falsos positivos en el espacio *xy* para una medida de predicción de abandono determinada. El modelo óptimo entonces es el que minimice la proporción de falsos positivos a la vez que maximice la sensibilidad. Una forma de determinar este óptimo es mediante el área bajo una curva ROC, donde mientras más área posea, resulta mejor en cuanto a los resultados esperados. Si se obtiene una predicción sobre la deserción que clasifique correctamente a todos los estudiantes que eventualmente abandonan o continúan en la escuela, el valor alcanzado sería 1, mientras que si las conjeturas resultan siendo al azar y representan un número promedio igual en cuanto a falsos positivos y falsos negativos, se podría observar una línea recta en ángulo de 45 grados.

Gráfica 16. Ilustración de la curva ROC



Fuente: Icefi/Save the Children, basado en (Haimovich, *et al*, 2018)

Luego de explicar estos conceptos, la guía enumera cinco pasos para llevar a cabo predicciones de abandono escolar:

Paso 1 Estimar un modelo de regresión lineal de abandono escolar con las características de los estudiantes que son relevantes para predecir este resultado. La variable dependiente es de tipo binaria para representar el abandono (1) o la continuidad del estudiante (0). En lugar de regresión lineal también se pueden utilizar otros algoritmos de aprendizaje de máquinas de regresión.

Paso 2 Predecir la probabilidad de abandono para los estudiantes en la muestra de estimación y los matriculados con base a sus características observables y los parámetros estimados en el paso 1. Se realiza la estimación con un modelo de regresión lineal (o de otro algoritmo), empleando un conjunto de datos históricos. Cada registro de este conjunto de datos posee una serie de características (variables predictoras o independientes) y el resultado de si desertó o continuó en el próximo año (variable respuesta o dependientes). A este conjunto la guía lo denomina como la muestra de estimación. Por tanto, en este paso se debe calcular la probabilidad de desertar que cada registro del mismo conjunto de estimación posee. También se calcula la probabilidad de desertar que tienen los estudiantes

que actualmente están matriculados[11] utilizando el modelo entrenado con el conjunto de estimación.[12]

Paso 3 Construir la curva ROC, resultante de usar todos los posibles puntos de corte para la probabilidad estimada a partir de la cual se predice que un estudiante abandonará. Variar la magnitud de corte *n* una cierta cantidad de veces. Para cada valor de *n*, las probabilidades calculadas en el paso anterior se convierten en una predicción binaria de si desertan (1) o no (0). Por tanto, para cada valor de *n* se construye una curva ROC utilizando las predicciones de desertar contra el resultado de si realmente lo hicieron del conjunto de estimación.

Paso 4 Seleccionar el punto de corte de la probabilidad predicha que provee una combinación de sensibilidad y proporción de falsos positivos, que es la más cercana al punto de predicción perfecta. La curva ROC con mayor área debajo de ella, es decir, la que muestre más aciertos y menores falsos positivos, es generada por la magnitud de corte óptima.

Paso 5 Considerar como estudiantes en riesgo de abandono a todos los estudiantes actualmente matriculados que tengan una probabilidad predicha de abandono por encima del punto de corte seleccionado.

3.3 Procedimiento para desarrollar el modelo de predicción de abandono escolar

Tomando como base los trabajos de Haimovich, Adelman, Ham, & Vazquez (2018) descritos anteriormente, a continuación, una explicación breve sobre la metodología para predecir qué estudiantes de sexto primaria inscritos en el ciclo de 2022-2023 en el sector público están en riesgo de desertar el sistema escolar. Esta explicación se basa en el procedimiento que los autores explican en su publicación, sin embargo, también se detallan procedimientos que se ajustan a los datos que se pueden incluir dentro de un modelo con variables actualizadas, que permitiría inferir los resultados apegados a un contexto más reciente. Esto para que el modelo, además de integrar variables más recientes al contexto nacional, también tome en consideración los impactos que el distanciamiento social producido por la pandemia del COVID-19.

[11] En este caso, de ellos no se sabe si desertarán o no. No poseen variable respuesta.
[12] En estos casos, sí se conoce las características o variables predictoras o independientes, son las que se introducen en el modelo entrenado y con las cuales se estiman las probabilidades de deserción.

3.3.1 Generar base de datos

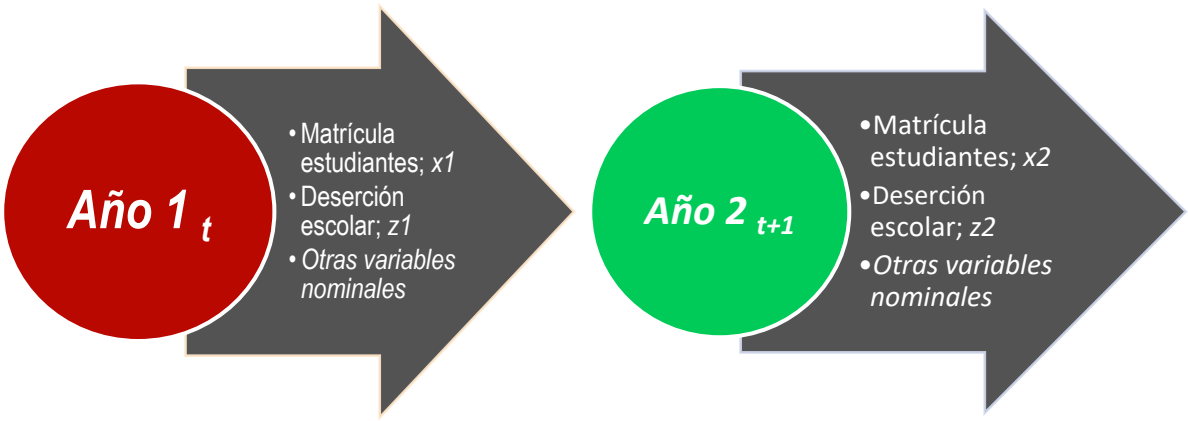
La construcción de la base de datos con la que se entrene el modelo es la piedra angular de todo el ejercicio. Las tendencias presentes en el conjunto de datos con el que se construya el modelo son las que luego se extrapolan para medir el riesgo de abandono de estudiantes en ciclos actuales. Por lo tanto, esta base de datos:

a. Mientras mejor refleje los factores que motivan el fenómeno de abandonar los estudios, el modelo será más íntegro; y,

b. Mientras más actualizados estén los datos y más pertinentes sean los patrones subyacentes a los que existen actualmente, el modelo se podrá aplicar con mayor seguridad.[13]

Para desarrollar un modelo de predicción, en términos de una regresión lineal múltiple o bien, un algoritmo que aplique el *machine learning,* es necesario contar con la materia prima, es decir los datos. Para generar una base de datos que se ajuste al modelo que predice la deserción escolar, se requieren datos administrativos de estudiantes de sexto primaria de ciclos anteriores que contenga información de si estos estudiantes en el próximo año terminaron desertando o no (conjunto de estimación).

Figura 6. Elaboración de base de datos para modelo de deserción escolar



Fuente: Icefi/Save the Children, basado en (Haimovich, *et al*, 2018)

[13] El hecho de aplicar el modelo con el conjunto de variables seleccionadas se hace bajo el supuesto de que los datos son estadísticamente indistinguibles de los que se usaron para ajustar el propio modelo.

Este conjunto será utilizado para construir el modelo de predicción de si un estudiante desertará (1) o no (0), y se deberán comprobar los resultados. Atendiendo a la condición expresada anteriormente de que una predicción solo es válida bajo el resultado positivo y verdadero (primer cuadrante de la prueba de hipótesis), el objetivo del modelo será lograr que la mayor proporción de resultados alcancen un resultado acertado. Esto se logrará al definir el conjunto de variables dependientes que mejor se ajusten a la asociación de la variable dependiente del modelo (*deserción*), la cual será una variable dicotómica (valores 1 y 0), en donde el resultado se entenderá como 1, cuando presenta riesgo de deserción y 0 cuando no presenta riesgo de deserción.

Para construir la base de datos incluyendo las variables que son importantes para el modelo, a continuación, se explica un resumen de las variables que se pueden ir integrando. En este caso, con base en el estudio de Haimovich, *et al.*, se sugiere que se pruebe el nivel de sensibilidad y precisión del modelo, tomando conjuntos de datos que se pueden ir ampliando según los resultados obtenidos. Esto permite integrar y descartar variables que presentan efectos dentro del modelo. Se podrían ir realizando pruebas de ajuste individualmente por cada variable, pero en cuanto a los conglomerados que se pueden armar, es más fácil identificar si dentro de cada nuevo conjunto de datos que se agrega, existe una mejora en cuanto a los parámetros de sensibilidad y precisión.

En la tabla 11 se describen los datos que se sugiere incluir a partir de la inforación administrativa que posee el Mineduc, así como otras variables que también se pueden incluir.

Tabla 11. Listado de variables que puede incluir el modelo de estimación de riesgo de abandono escolar

Especificación	Fuente	Variable
Modelo 1	Datos administrativos Mineduc	Sexo del alumno Edad del alumno Autoidentificación por grupo étnico, o variable que describa su grupo étnico Descripción si trabaja o no. Trabajo doméstico. Indicador de repetición de grado (actual) Indicador que determine si cursó preprimaria Indicador de disponibilidad de todo el material didáctico Indicador de la motivación del alumno Resultados de pruebas de conocimientos básicos (PISA)

Especificación	Fuente	Variable
Modelo 2	Datos administrativos y de otros instrumentos estadísticos (podría ser el Censo de Población 2018)	*Variables del Modelo 1, y adicional:* Máximo grado escolar alcanzado por los padres (dicotómica) Ayuda con la tarea (dicotómica) Calidad de la vivienda Acceso a fuente de energía eléctrica Acceso a computadora o tableta (también se puede incluir teléfono inteligente) Acceso a internet Acceso a libros Indicador de disponibilidad de bienes y servicios
Modelo 3	Datos administrativos y de otros instrumentos estadísticos (podría ser el Censo de Población 2018 y la Encuesta de Empleo e Ingresos Enei, más reciente)	*Variables del Modelo 2, y adicional:* Continuidad en la misma escuela Prima salarial de los padres

Fuente: Icefi/Save the Children, basado en (Haimovich, *et al*, 2018)

A continuación se describen algunas variables que por su relevancia y el aporte que pueden generar, se deben considerar para incluirlas en la base de datos:

I. Medidas aproximadas de pobreza como: el material de vivienda, servicios básicos, ingreso monetario;

II. Alguna medida aproximada de mercado laboral local. Para esto último podría usarse la Encovi o Enei aunque sean representativas más a nivel nacional.

III. Uso de variables a partir de la *Teoría económica*, en la cual se predice que un estudiante abandonará sus estudios cuando perciba que el costo de oportunidad de continuarlos sea mayor que los aparentes beneficios de hacerlo. Variables relacionadas con ingresos familiares, acceso a medios de transporte, costos de materiales didacticos, entre otros.

IV. Importancia de incluir variables asociadas a pandemia del COVID-19, por ejemplo: qué hizo en los ciclos lectivos durante la pandemia; qué modalidad de educación recibió; si tuvo acceso a las herramientas y materiales para atender modalidades virtuales, como internet, computadoras, tableta, o teléfono inteligente, entre otros.

La crisis sanitaria potencialmente cambió todas las tendencias y valoraciones de educación en estos últimos años, por lo que es importante incluir este tipo de variables para explroar qué tan impactantes han sido en la decisión de abandonar la escuela.

Es preciso señalar que en la tabla de variables que se pueden considerar, cada modelo que se desagrega contiene variables acumulativas. Es decir, el modelo 2 integra datos tanto administrativos del Mineduc como del Censo. Los datos son complementarios, por lo que los modelos no usan datos completamente distintos.

Esta guía se puede aplicar para generar una base de datos a partir de datos administrativos de estudiantes de sexto grado de primaria de ciclos anteriores, que provea información de estudiantes que podrían desertar del sistema escolar o no el próximo año (conjunto de estimación). Sin embargo, también se puede plantear el ejercicio para estudiantes que pasan del nivel básico al diversificado, en cuyo caso el modelo requiere datos que resultarían ser iguales o, al menos, similares al ejercicio de los estudiantes de sexto grado de primaria. En todo caso, el objetivo de la guía metodológica es lograr desarrollar el modelo de tal forma que mida el riesgo de deserción escolar entre niveles, aunque también, podría medir ese riesgo de grado a grado de cada nivel educativo. Todo depende de la disponibilidad, acceso y calidad de los datos, tanto a nivel individual como de las características de las unidades familiares a las cuales pertenece el estudiante.

Se debe tener en cuenta que la base de datos que se debe armar requiere la misma cantidad de registros para todos los campos (variables). Es decir, la estructura debe respetar el contenido de la información y el diseño de toda base de datos. En este caso, las columnas son los campos o variables del modelo, y cada fila o registro corresponde a cada dato observado, siendo en este caso, cada estudiante del registro administrativo. La figura 7 ejemplifica la estructura de la base de datos requerida, la cual puede contar con todos los registros administrativos de la matrícula pública, o podría integrar la matrícula privada y de las cooperativas, de estudiantes de sexto grado de primaria y de primero básico, de al menos 2 periodos de tiempo. Normalmente en el primer campo (columna), debería incluirse el código de estudiante, para tener presente que estos son registros individuales. No obstante, el orden de los campos y registros de la base no determina ningún resultado, por lo que también se podrían colocar otras variables al inicio, como el año, departamento, municipio, escuela y otros.

Figura 7. Ejemplo de base de datos consolidada de registros administrativos

Cod_Registro	Var_1	Var_2	Var_3	Var_4	Var_5	Var_6	Var_7	Var_8	Var …
Est_1									…
Est_2									…
Est_3									…
Est_4									…
Est_5									…
Est_6									…
Est_7									…
Est_8									…
Est_…	…	…	…	…	…	…	…	…	…

Fuente: Icefi/Save the Children, elaboración propia.

La base de datos también podría incluir algunas variables que se extrapolen de los registros administrativos. En este caso, a esas variables se les podría denominar *proxys,* ya que es una aproximación a una condición o factor que no está siendo medido dentro de los registros administrativos. Podrían incluir una condición socioeconómica o algún tipo de caracterización cualitativa como pobreza, necesidad básica insatisfecha, acceso a algún tipo de servicio, entre otros. También se pueden incluir datos cuantitativos que, si bien podrían no estar dentro de los registros administrativos, servirían para ajustar mejor la inferencia del modelo en la asociación de factores exógenos. Por ejemplo, se podrían incluir datos sobre ingresos de los padres, pago de servicios de transporte, incremento de la inflación medida por el índice de precios al consumidor, entre otros. En todo caso, algunas variables que se pueden ir adicionando podrían convertirse en variables dicotómicas, que adopten el valor de 1 o 0, en caso cumpla o no cumpla la condición, en cuyo caso se podría estimar una variable dicotómica para una variable cualitativa como el acceso a una herramienta tecnológica digital como computadora, tableta o teléfono inteligente, al que se aplicaría un criterio de 1 cuando sí posee, o 2 cuando no. Dependerá de la forma en cómo se estructure la base de datos y las variables que se integren, en cómo se incluirán ciertas variables cuyo manejo resulta ser de más simple de forma dicotómica.

Figura 8. Ejemplo de base de datos consolidada integrando variables *proxys*

	Registros admin.						*Censo_2018*			
Cod_Registro	Var_Depto	Var_Muni	Var_3	Var_4	Var_5	Var_Depto_Muni	VarCen_6	VarCen_7	VaCenr_8	Var…
Est_1										…
Est_2										…
Est_3										…
Est_4										…
Est_5										…
Est_6										…
Est_7										…
Est_8										…
Est_…	…	…	…	…	…		…	…	…	…

Fuente: Icefi/Save the Children, elaboración propia.

Si se incluyen variables *proxys,* dado que no corresponden necesariamente a la base de datos original de registros administrativos del Mineduc, es importante generar una variable de *llave* o de vinculación o asociación. En este caso, dado que, si se extrapolan datos de otras bases estadísticas como un censo de población, una encuesta de hogares, o de estadísticas socioeconómicas, se debe tomar en cuenta que los datos no estarán asociados con cada registro (estudiante) particular. En este caso, la asociación se debería lograr con una variable vinculada directamente con otra variable de la base de datos del Mineduc. En ese contexto, se podría vincular una variable por su área geográfica, llegando a nivel de comunidad, municipio o departamento. Un ejemplo de este tipo de vinculación podría ser usar una variable representativa del lugar geográfico, como los ingresos familiares promedio del municipio,[14] que asumiría que el ingreso familiar es el mismo en todo el municipio, dado que el promedio es un valor de tendencia central y que extrapola una característica general a nivel geográfico, por lo que no se puede esperar que tenga un grado de especificidad real. Es decir, el dato explicaría una condición general que caracteriza a la población que habita en un área geográfica específica, pero no está explicando con precisión una característica de cada registro (estudiante) en particular.

[14] En este caso se utiliza el valor promedio, debido a que los datos son distintos dentro de la muestra o de la base que se está utilizando de referencia. En caso de dispersión excesiva en los datos, se recomienda utilizar la mediana. En cualquiera de los casos, hay que recordar que esta variable sería de aproximación, no un dato que explique exactamente o con precisión la situación de cada individuo. Una variable *proxy* se debe de usar únicamente si el nivel de asociación es significativo para explicar el modelo.

3.3.2. Aplicación del modelo de medición de riesgo de abandono escolar

Una vez se ha consttruido una base de datos que incluya todas las variables que requiere el modelo, se procede aplicarlo para medir el riesgo de abandono escolar. Para ello es necesario utilizar un programa informático que trabaje modelos estadísticos de los que están disponibles en el mercado, como *SPSS, Stata, Phyton* o *RStudio*. Lo importante es aplicar las funciones del modelo y lograr generar los resultados que midan si existe o no riesgo de abandono escolar en las y los estudiantes. En estas condiciones, se siguen los siete pasos siguientes:

➔ **Paso 1. Dividir base de datos en un subconjunto de entrenamiento y subconjunto de prueba.**

Se debe separar una parte de la base de datos para desarrollar un subconjunto de entrenamiento, el cual servirá para correr el modelo y medir la sensibilidad de las variables. El otro subconjunto ayudará a realizar las pruebas midiendo los resultados reales del abandono.

➔ **Paso 2. Estimar un modelo de regresión lineal de abandono escolar en las características de los estudiantes que son relevantes para predecir este resultado.**

En este caso, solo hay dos predictores de abandono que se pueden utilizar como variables explicativas en el modelo de deserción escolar: el género (capturado por la variable *man*) y la escuela a la que asiste el estudiante (para la cual se puede incluir un *efecto fijo / variable dummy* por escuela). La variable dependiente es *dropout* (abandono), que solo se observa en la muestra de estimación (estudiantes que se matricularon en sexto grado en 2019 y no se matricularon en 2020). Este modelo se puede estimar en Stata® usando, por ejemplo, el siguiente código:

```
regress dropout man i.id_school if est_sample==1
```

Paso 3. Predecir la probabilidad de abandono para los estudiantes en la muestra de estimación y aquellos actualmente matriculados con base en sus características observables y los parámetros estimados en el Paso 1.

Esto se puede lograr utilizando el programa informático *Stata*® usando el comando de *post-estimación predict* seguido del nombre de la nueva variable que capturará la *probabilidad* predicha:

predict prob_dropout

Esto generará una nueva variable con la predicción lineal del modelo del Paso 1 para las observaciones que están tanto dentro como fuera de la muestra de estimación. Por lo tanto, los estudiantes matriculados en 2020 ahora tienen una *probabilidad* predicha de abandono $\hat{pi}$ basada en su género y la escuela a la que asisten.[15]

➔ **Paso 4. Construir la curva ROC resultante de usar todos los posibles puntos de corte para la probabilidad predicha a partir de la cual se predice que un estudiante abandonará.**

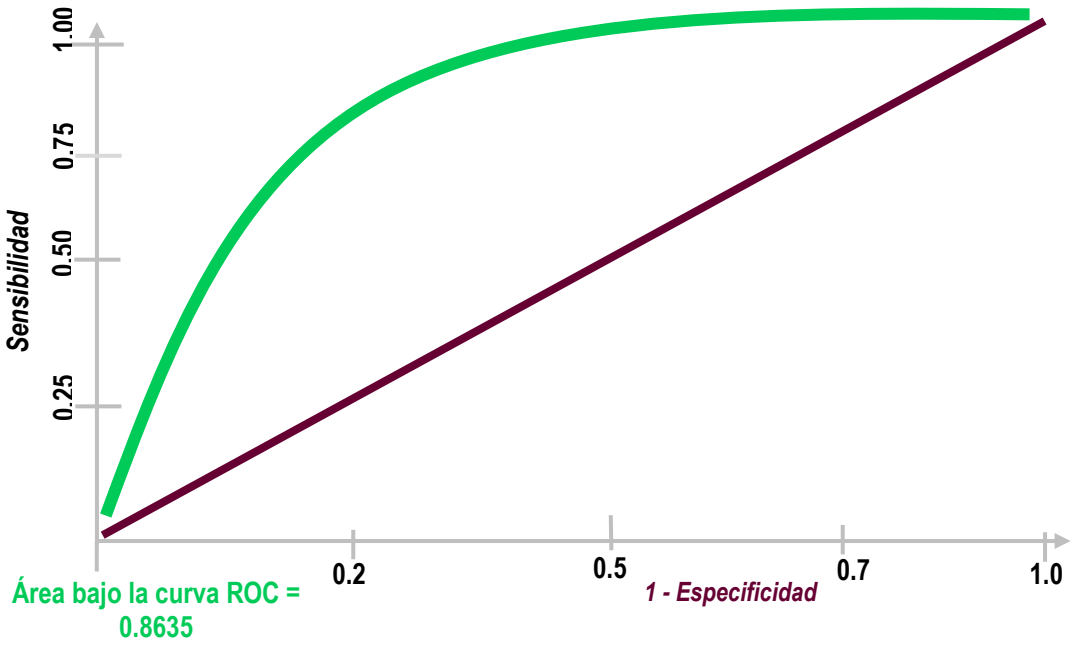
Si bien este paso es más conceptual y no es estrictamente necesario, se puede implementar utilizando la opción *graph* del comando *roctab*, que debe ir seguido de la variable de referencia que indica el estado real de la observación (en este caso, abandono o no abandono), codificado como 0 o 1, y la probabilidad de abandono predicha, con valores más altos indicando un mayor riesgo:

roctab dropout prob_dropout if est_sample==1, graph

[15] Un método alternativo para estimar el modelo de efectos fijos y las *probabilidades* predichas que es menos intensivo computacionalmente, consiste en predecir la probabilidad de abandono sin los efectos fijos y el efecto fijo por separado, y luego sumarlos. En este caso, el código para los pasos 1 y 2 sería:
areg dropout man if est_sample==1, absorb(id_school)
predict predicted_probability_without_fe
predict school_fe_sample if e(sample), d
egen school_fe=max(school_fe_sample), by(id_school)
gen prob_dropout = predicted_probability_without_fe + school_fe

que aplicado a este ejemplo da la siguiente curva ROC:

Figura 9. Curva ROC



Fuente: Icefi/Save the Children, elaboración a partir de (Haimovich, Adelman, Ham, & Vazquez, 2018)

Paso 5. Seleccionar el punto de corte de la probabilidad predicha que provee una combinación de (sensibilidad, proporción de falsos positivos) que es la más cercana al punto de predicción perfecta.

Este paso puede implementarse con el comando *cutpt*, desarrollado por Clayton (2013), utilizando la opción *nearest* para seleccionar el punto más cercano al punto de predicción perfecta, y la misma sintaxis empleada en el paso 3:

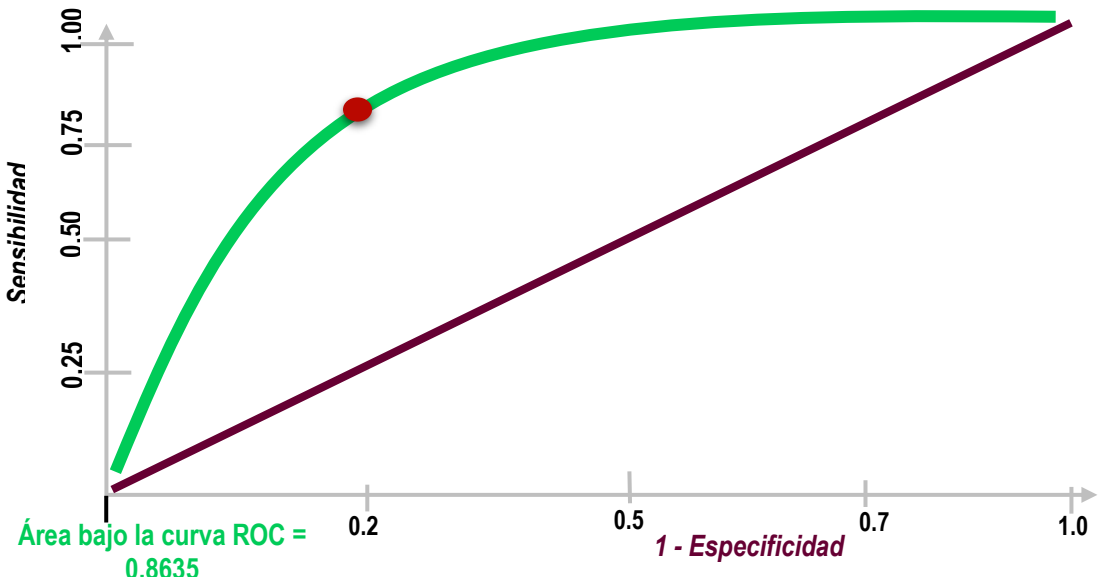
cutpt dropout prob_dropout if est_sample==1, nearest

```
. cutput dropout prob_dropout if est_sample==1, nearest

Empirical cutpoint estimation
Method:                                 Nearest to (0, 1)
Reference variable:                     drpout (0=neg, 1=pos)
Classification variable:                prob_dropout
Empirical optimal cutpoint:             .28462084
Sensivity at cutpoint:                  0.78
Specifity at cutpoint:                  0.77
Area under ROC Curve at cutpoint:       0.78
```

En este ejemplo, el valor de corte óptimo $X*$ es 0.28 (redondeado a dos decimales). Así, predecir que los estudiantes con probabilidad predicha $\hat{pi}$ mayor a 0.28 abandonarán la escuela resulta en una sensibilidad de 78% y una proporción de falsos positivos de 1-especificidad = 23%. Este punto es destacado en la curva ROC que se muestra abajo.[16]

Figura 10. Curva ROC con el punto de sensibilidad



Fuente: Icefi/Save the Children, elaboración a partir de (Haimovich, Adelman, Ham, & Vazquez, 2018).

[16] Este gráfico se puede obtener usando la opción addplot en el paso anterior: roctab dropout prob_dropout if est_sample==1, graph addplot(scatteri `e(sens)' `=1 - e(spec)') legend(off)

➔ Paso 6. Considerar como estudiantes en riesgo de abandono a todos los estudiantes actualmente matriculados que tengan una probabilidad predicha de abandono por encima del punto de corte seleccionado.

En esta aplicación, los estudiantes actualmente matriculados son aquellos cuya variable *dummy est_sample* toma un valor de 0, ya que están matriculados en sexto grado de primaria en 2020. Por lo tanto, se pueden marcar con un 1 utilizando el siguiente código y sus identificadores se pueden utilizar para obtener una lista de la población que está en riesgo de abandono mientras aún está inscrita en el sistema educativo:

```
gen at_risk = .
replace at_risk = 1 if est_sample==0 & prob_dropout> e(cutpoint) & prob_dropout < .
replace at_risk = 0 if est_sample==0 & prob_dropout<=e(cutpoint)
list id_student if at_risk==1
```

Si los parámetros estimados en el modelo en el Paso 1 son relativamente estables en el tiempo, se espera que el uso de este algoritmo con los estudiantes actualmente matriculados resulte en una sensibilidad y una proporción de falsos positivos de alrededor del 78% y 23%, respectivamente, de acuerdo con la evaluación en el conjunto de entrenamiento del modelo. Este procedimiento analítico ha demostrado tener un buen rendimiento fuera de la muestra (Adelman *et al*. 2018), pero con 3 en lugar de 2 años de recopilación de datos (por ejemplo, 2018, 2019 y 2020). El desempeño específico fuera de la muestra de cualquier modelo se puede examinar comparando las predicciones de la muestra de estimación (por ejemplo, 2018), con los resultados reales de abandono (de 2019), y luego aplicar ese modelo para predecir el abandono en 2020.

➔ Paso 7. Probar el modelo predictivo con el conjunto de prueba

Se comprueba el modelo predictivo con el conjunto de prueba. Esto significa medir el *out-of-sample performance* del modelo (como lo describen muy brevemente Haimovich, *et al*). Básicamente, se toma el subconjunto de prueba, se suprime temporalmente la variable de si abandonaron o no la escuela; se calcula la probabilidad de abandonar según el modelo entrenado y los parámetros que tengan incluidos; se aplica la magnitud de corte *n* para ver quienes se predice que abandonaron o no; y, finalmente, se calculan las métricas de la matriz de eventos para ver el porcentaje de falsos positivos, falsos negativos, prueba estadística y demás métricas descritas.

Conclusiones

4

1. La educación es un derecho fundamental de todas las personas, su garantía facilita o permite el goce pleno de otros derechos humanos. El Estado de Guatemala debe velar por garantizar este derecho, prestando los servicios de educación en todos los niveles escolares, y para toda la población que habita el territorio nacional. Sin embargo, existen brechas en la cobertura educativa, y para lograr que miles de niñas, niños y adolescentes logren acceder al sistema educativo y no deserten.

2. El Estado de Guatemala ha ratificado compromisos internacionales sobre el derecho a la educación, y posee un marco normativo nacional que refuerza las entidades públicas que deben garantizar los servicios de educación. Actualmente el Gobierno está adherido a la Agenda 2030 de los Objetivos de Desarrollo Sostenible, que en su Objetivo 4, garantizar una educación inclusiva, equitativa y de calidad y promover oportunidades de aprendizaje durante toda la vida para todos, tiene asociadas metas sobre la universalización y cobertura total de la educación. Este compromiso es importante para el desarrollo humano de la población y requiere el esfuerzo del Estado para cumplirlo, lo que a su vez requiere una mayor cantidad de recursos financieros, en conjunto con eficiencia y calidad del gasto público, midiendo sus resultados.

3. Previo a la crisis derivada por la pandemia del COVID-19, los indicadores mostraban que las coberturas de la matrícula estudiantil se habían reducido en algunos de los niveles educativos. También, evidenciaban que para los niveles de educación básicos y diversificado, las matrículas de estudiantes son muy bajas y requieren esfuerzos económicos y políticos para mejorar las condiciones de acceso a la educación de miles de adolescentes y jóvenes que lo demandan. Existe a su vez un componente significativo de deserción escolar, que es más alto en estos niveles de educación y que requieren atenciones directas del gobierno, especialmente para alumnos que pasan de un nivel de educación a otro. Datos preliminares sugieren que estos indicadores se deterioraron como resultado del impacto de la pandemia.

4. El uso de modelos de alerta temprana ante el riesgo de abandono escolar ha resultado eficiente, incluso en Guatemala y Honduras, donde el Banco Mundial ya estimó indicadores de riesgo de deserción. Sin embargo, este modelo puede ser actualizado ante las situaciones que actualmente están sucediendo en Guatemala, dentro de un contexto en el cual la crisis por la pandemia, además de afectar desde lo económico, también trasciende al plano social. Enfrentar estos desafíos requiere estrategias y planificación de intervenciones mejor adaptadas al nuevo contexto, especialmente para proteger a poblaciones vulnerables como la niñez y adolescencia.

5. Un modelo de alerta temprana ante el riesgo de abandono escolar permite contar con datos que facilitan la implementación de acciones, acordes a las necesidades de los centros educativos, para además de evitar el riesgo que un alumno abandone sus estudios, se creen condiciones para que el resto de los estudiantes permanezcan en la escuela y cuenten con todos los elementos necesarios para una mejor educación.

Recomendaciones

5

1. El Icefi recomienda al Gobierno de Guatemala, especialmente a las autoridades del Mineduc, planificar y programar su presupuesto público en función de las nuevas necesidades de la población estudiantil, asociadas al impacto de la pandemia del COVID-19. Debe destinar recursos para que la niñez y la adolescencia que se encuentra fuera de la escuela retome sus estudios inmediatamente, y reconocer que para ellos hace falta incrementar la cobertura de los servicios, los materiales didácticos, el personal del magisterio que imparte las clases, recursos tecnológicos para facilitar las clases a distancia, entre otros aspectos de una realidad socioeconómica compleja y desafiante.

2. Es importante que las autoridades del Mineduc actualicen los datos estadísticos de la matrícula estudiantil y sean considerados información pública. Así se tendrá una mejor visión del impacto de la pandemia en el sistema educativo nacional durante más de 2 años. Con la información actualizada también se podrán seguir los indicadores educativos, y, en este caso, implementar un modelo de predicción de abandono escolar mucho más ajustado a la realidad que vive Guatemala.

3. Se recomienda al personal técnico del Mineduc replicar el modelo de indicadores de abandono escolar, usando la información de los registros administrativos detallados por estudiante. Esto proveerá información que permitirá trazar la evolución de cada estudiante y su continuidad en el sistema educativo. La base de registros administrativos del Mineduc debe contar con información amplia de las características socioeconómicas de cada estudiante, lo que enriquecerá el modelo de indicadores de abandono escolar.

4. Se pueden utilizar otros instrumentos que aporten mediciones de variables *proxys* para que el modelo de predicción de abandono escolar se adapte de mejor forma a ciertas condiciones de cada estudiante. Sin embargo, se debe tener el cuidado necesario que estos datos sean contemporáneos a la información base, y que al extrapolar datos se puedan asociar a los registros individuales. Utilizar datos que representen situaciones a nivel geográfico muy agregado, puede que no generen resultados de asociación significativos.

5. Los análisis de alerta temprana no explican las causas del abandono escolar, pero son una aproximación, que, a partir de la asociación de ciertas variables, pueden predecir que un estudiante probablemente desertará del sistema escolar. Estos análisis deben hacerse con cuidado de mantener la confidencialidad de los estudiantes, y no deben usarse para etiquetarlos negativamente o estigmatizarlos. Se debe estudiar a profundidad qué intervenciones realmente son efectivas para evitar la deserción escolar.

6. Es importante que los resultados que se obtengan con un buen nivel de sensibilidad en el modelo se tomen en consideración para implementar acciones de prevención ante el riesgo de abandono escolar. Para ello, el Mineduc debe realizar intervenciones financieramente viables. Estas intervenciones deben coadyuvar a garantizar el derecho a la educación de los estudiantes que por situaciones que ya tenían o se agravaron con la pandemia, abandonaron sus estudios. Esta es una catástrofe social que requiere medidas urgentes para que estas niñas, niños y adolescentes permanezcan en su escuela o instituto.

Referencias bibliográficas

6

Banco Mundial. (2021). *Actuemos ya para proteger el Capital Humano de Nuestros Niños. Los costos y la respuesta ante el impacto de la pandemia de COVID-19 en el sector educativo de América Latina y el Caribe.* 1818 H Street NW, Washington, DC 20433: Banco Mundial.

Comisión Económica para América Latina y el Caribe (CEPAL). (2013). Consenso de Montevideo sobre población y desarrollo. .

Comité contra la Tortura (CAT). (24 de junio de 2013). *CAT/C/GTM/CO/5-6, Observaciones finales sobre los informes periódicos quinto y sexto combinados.*

Comité de Derechos Económicos, Sociales y Culturales (CDESC). (2000). *E/C.12/2000/13, Principios de Limburgo relativos a la aplicación del Pacto Internacional de Derechos Económicos, Sociales y Culturales.*

Comité de Derechos Económicos, Sociales y Culturales (CDESC). (2003). *E/C.12/1/Add.93, Observaciones finales del Comité de Derechos Económicos, Sociales y Culturales : Guatemala.*

Comité de Derechos Económicos, Sociales y Culturales (CDESC). (2014). *E/C.12/GTM/CO/3, Observaciones finales sobre el tercer informe periódico de Guatemala.*

Comité de Derechos Humanos (CCPR). (2012). *CCPR/C/GTM/CO/3, Examen de los informes presentados por los Estados. Observaciones finales del Comité de Derechos Humanos a Guatemala.* New York.

Comité de los Derechos del Niño (CDN). (2010). *CRC/C/GTM/CO/3-4, Examen de los informes presentados por los Estados Partes en virtud del artículo 44 de la convención.*

Comité de los Derechos del Niño (CDN). (2018). *CRC/C/GTM/CO/5-6, Observaciones finales sobre los informes periódicos quinto y sexto combinados de Guatemala.*

Comité para la Eliminación de la Discriminación contra la Mujer (CEDAW). (2009). *CEDAW/C/GUA/CO/7, Observaciones finales del séptimo informe periódico de Guatemala.*

Congreso de la República de Guatemala. (2014). *Dictamen Iniciativa 4826* .

Congreso de la República de Guatemala. (22 de agosto de 2021). *Noticias del Congreso de la República de Guatemala*. Obtenido de https://www.congreso.gob.gt/noticias_congreso/6804/2021/3#gsc.tab=0

Consejo Nacional de Desarrollo Urbano y Rural (Conadur) y Secretaría General de Planificación y Programación de la Presidencia (Segeplán). (2014). *Plan Nacional de Desarrollo K'atun: nuestra Guatemala 2032.*

Consejo Nacional de la Juventud (Conjuve). (2012). *Política Nacional de la Juventud 2012-2020.* Guatemala.

Defensoría de la Juventud. (2021). *Informe de supervisión al proceso de implementación de la Política Nacional de Juventud 2021-2032 y su Plan de Acción Interinstitucional.* Procurador de los Derechos Humanos.

Espinoza Díaz, Ó., González Fiegehen, L., & Loyola Campos, J. (2021). *Factores determinantes de la deserción escolar y expectativas de estudiantes que asisten a esculeas alternativas.* Educación y Educadores. doi:10.5294/educ.2021.24.1.6

Grupo de trabajo sobre juventud de la Plataforma de Colaboración Regional para América Latina y el Caribe (GTJ). (2021). *Las juventudes latinoamericanas y caribeñas y la Agenda 2030 para el Desarrollo Sostenible: una mirada desde el sistema de las Naciones Unidas.* Organización de las Naciones Unidas, Santiago.

Haimovich, F., Adelman, M., Ham, A., & Vazquez, E. (2018). *Predicting school dropout with administrative data: new evidence from Guatemala and Honduras.* Education Economics. doi:10.1080/09645292.2018.1433127

Mineduc. (2016). *http://estadistica.mineduc.gob.gt*. Recuperado el 8 de abril de 2022, de http://estadistica.mineduc.gob.gt/anuario/data/anexos/glosario.html#Ausentismo_escolar

Mineduc. (2016). *http://estadistica.mineduc.gob.gt*. Recuperado el 22 de abril de 2022, de http://estadistica.mineduc.gob.gt/anuario/data/anuario/2016/data/Anexos/Glosario.htm#Ausentismo_escolar

Organización de Estados Americanos (OEA). (1979). *Resolución N° 448, Estatuto de la Corte Interamericana de Derechos Humanos.* .

Organización de las Naciones Unidas (ONU). (2015). Resolución de Asamblea General 70/1, Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible.

Organización Iberoamericana de Juventud (OIJ). (2017). Convención Iberoamericana de Derechos de los Jóvenes + protocolo adicional. Badajoz.

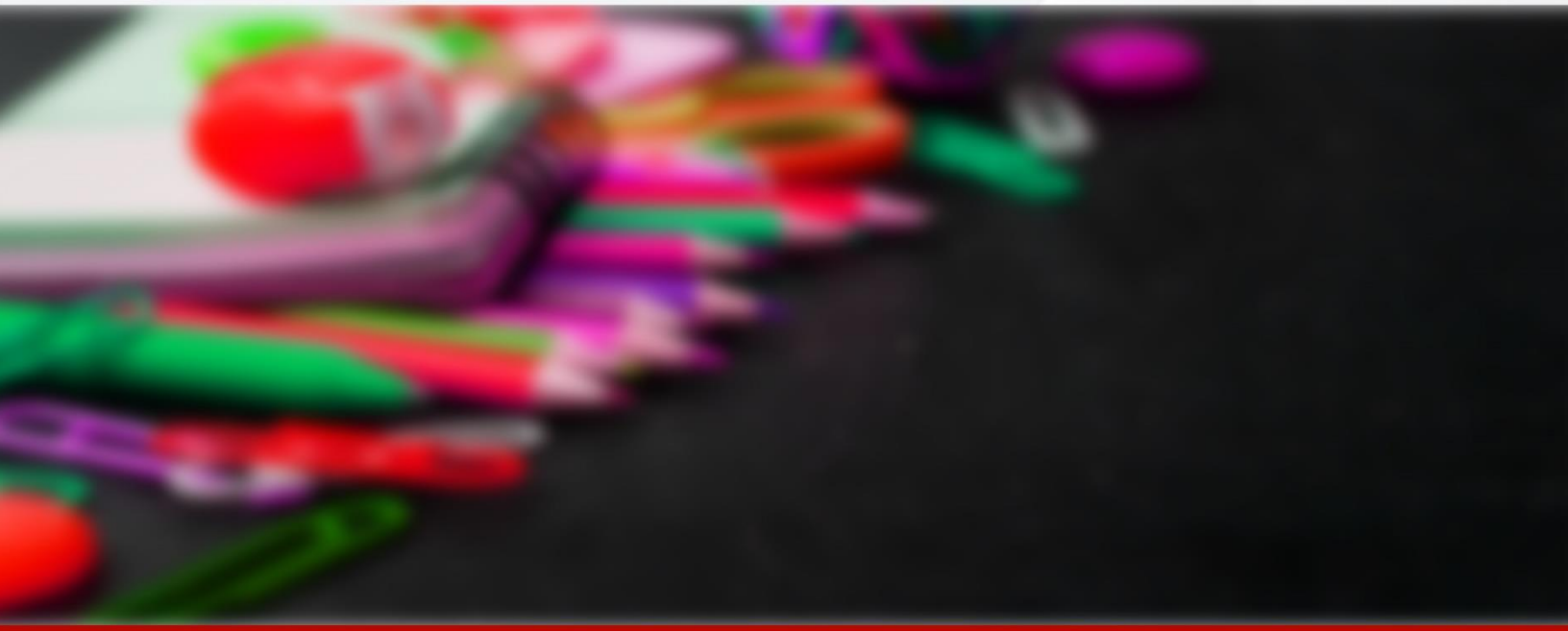
Organización Mundial de la Salud (OMS). (1986). *La salud de los jóvenes: un desafío para la sociendad.* Ginebra.

Secretaría de Planificación y Programación de la Presidencia (Segeplán). (2017). *Guía para la formulación de políticas públicas.* Guatemala.

Secretaría de Planificación y Programación de la Presidencia (Segeplán). (23 de agosto de 2021). Obtenido de https://www.segeplan.gob.gt/nportal/index.php/sala-de-prensa/2016-01-26-18-14-30/noticias/2093-politica-nacional-de-la-juventud-2021-2032-sera-actualizada

Unesco. (22 de septiembre de 2020). *https://es.unesco.org.* Recuperado el 4 de abril de 2022, de https://es.unesco.org: https://es.unesco.org/news/formacion-tecnica-como-respuesta-bisagra-reducir-potencial-desercion-causada-coronavirus

Conoce más
acerca del Icefi en:



Save the Children

ICEFI
INSTITUTO CENTROAMERICANO DE ESTUDIOS FISCALES

Guatemala, Centroamérica, septiembre del 2022