
Análisis longitudinal del Registro Nacional de Alumnos sobre trayectorias educativas

Karina Videgain



Estadísticas e indicadores temáticos



Indicadores
educativos

INEE
Instituto Nacional para la
Evaluación de la Educación
México

ANÁLISIS LONGITUDINAL DEL REGISTRO NACIONAL DE ALUMNOS SOBRE TRAYECTORIAS EDUCATIVAS

Primera edición, 2015

ISBN: en trámite

Autora
Karina Videgain

D. R. © Instituto Nacional para la Evaluación de la Educación

Barranca del Muerto 341, Col. San José Insurgentes,
Deleg. Benito Juárez, C.P. 03900, México, D.F.

COORDINADORA EDITORIAL
Alejandra Delgado Santoveña

EDITORA
María Norma Orduña Chávez

CORRECCIÓN DE ESTILO
Carlos Garduño González

DISEÑO
Martha Alfaro Aguilar

FORMACIÓN
Martha Alfaro Aguilar
Heidi Puon Sánchez

Hecho en México. Prohibida su venta

Consulte el Catálogo de publicaciones en línea: **www.inee.edu.mx**

La elaboración de esta publicación estuvo a cargo de la Dirección General de Integración y Análisis de Información. El contenido, la presentación, así como la disposición en conjunto y de cada página de esta obra son propiedad del INEE. Se autoriza su reproducción parcial o total por cualquier sistema mecánico o electrónico para fines no comerciales y citando la fuente de la siguiente manera:

Videgain, Karina (2015). *Análisis longitudinal del Registro Nacional de Alumnos sobre trayectorias educativas*. México: INEE.

Análisis longitudinal del Registro Nacional de Alumnos sobre trayectorias educativas

Karina Videgain



Estadísticas e indicadores temáticos



Indicadores
educativos

INEE
Instituto Nacional para la
Evaluación de la Educación
México

ÍNDICE

7	1 Introducción
8	2 Antecedentes
11	3 Objetivos general y específicos
12	4 Justificación y alcances
14	5 Metodología
	5.1 Fuentes de datos
20	5.2 Análisis y técnicas utilizadas
23	5.3 Selección de covariantes
24	6 Marco de referencia
27	7 Análisis longitudinal de los datos
	7.1 Primaria
50	7.2 Secundaria
74	8 Conclusiones
76	9 Recomendaciones
77	Bibliografía
79	Anexo: códigos de programación en R
	1 Cohorte primaria 2008-2009
89	2 Cohorte secundaria 2008-2009

ÍNDICE DE CUADROS, GRÁFICAS Y FIGURAS

Cuadros

- 15** Cuadro 1 Número de registros de la base de datos RNA
- 15** Cuadro 2 Variables en base de datos de RNA
- 16** Cuadro 3 Estado de situación respecto a registros faltantes por ciclo
- 27** Cuadro 4 Medidas resumen de la fecha de nacimiento
- 28** Cuadro 5 Composición por sexo y edad en el momento de ingreso a primaria
- 28** Cuadro 6 Distribución de modalidades de servicio
- 29** Cuadro 7 Las 10 secuencias de modalidades más frecuentes
- 31** Cuadro 8 Distribución de trayectorias educativas, primaria temprana
- 34** Cuadro 9 Distribución de trayectorias educativas por sexo, primaria temprana
- 36** Cuadro 10 Distribución de trayectorias educativas por edad al inicio de la primaria
- 41** Cuadro 11 Distribución de grupos de trayectorias
- 43** Cuadro 12 Composición de los ocho grupos de trayectorias por covariantes
- 46** Cuadro 13 Duraciones en cada grado educativo (de 1° a 4°)
- 48** Cuadro 14 Distribución por covariantes de número de transiciones realizadas
- 49** Cuadro 15 Matriz de transiciones entre estados para población total, varones y mujeres
- 51** Cuadro 16 Medidas resumen de fecha de nacimiento
- 51** Cuadro 17 Composición por sexo y edad en el momento de su ingreso a secundaria
- 52** Cuadro 18 Distribución de modalidades de servicio al inicio de la secundaria
- 53** Cuadro 19 Las 10 secuencias de modalidades de servicio más frecuentes
- 54** Cuadro 20 Distribución de frecuencia entre modalidades de servicio en sexto primaria y primero de secundaria
- 54** Cuadro 21 Distribución de frecuencia entre modalidades de servicio en primero y segundo grado de secundaria
- 56** Cuadro 22 Distribución de trayectorias educativas
- 59** Cuadro 23 Distribución de trayectorias educativas en secundaria por sexo
- 60** Cuadro 24 Distribución de trayectorias educativas por edad al inicio de la primaria
- 66** Cuadro 25 Distribución de grupos de trayectorias
- 68** Cuadro 26 Composición de los cinco grupos de trayectorias por covariantes
- 69** Cuadro 27 Duraciones en cada grado educativo (de 7° a 9°) por sexo y edad
- 70** Cuadro 28 Duraciones en cada grado educativo (de 7° a 9°) por modalidad de servicio al inicio de secundaria
- 71** Cuadro 29 Distribución por covariantes de número de transiciones realizadas
- 72** Cuadro 30 Matriz de transiciones entre estados para población total, varones y mujeres

Gráficas

- 30** Gráfica 1 Distribución de secuencias individuales de modalidad de servicio de escuelas
- 32** Gráfica 2 Las 10 secuencias más comunes en primaria temprana
- 33** Gráfica 3 Histograma de secuencias
- 35** Gráfica 4 Las 10 secuencias más frecuentes por sexo, primaria temprana
- 37** Gráfica 5 Las 10 secuencias más frecuentes en primaria temprana, por edad al inicio de la primaria
- 38** Gráfica 6 Índice de entropía transversal en trayectorias de primaria temprana, por sexo y edad al inicio de la primaria
- 53** Gráfica 7 Distribución de secuencias individuales de modalidad de servicio de escuelas
- 57** Gráfica 8 Las 10 secuencias más frecuentes en secundaria
- 58** Gráfica 9 Histograma de secuencias en secundaria
- 61** Gráfica 10 Las 10 secuencias más frecuentes en secundaria, por sexo
- 62** Gráfica 11 Las 10 secuencias más frecuentes por edad al inicio de la secundaria
- 63** Gráfica 12 Índice de entropía transversal por sexo y edad al inicio de la secundaria

Figuras

- 40** Figura 1 Salidas gráficas del análisis *silhouette* para soluciones de clúster 6, 7 y 8
- 41** Figura 2 Estructura de árbol entre los 8 grupos de trayectorias
- 42** Figura 3 Gráficas con trayectorias individuales por grupos de trayectorias (1 a 8)
- 64** Figura 4 Salidas gráficas del análisis *silhouette* para soluciones de clúster 4, 5 y 6
- 65** Figura 5 Estructura de árbol entre los cinco grupos de trayectorias
- 67** Figura 6 Trayectorias individuales por grupo de trayectorias (1 a 5)

1

INTRODUCCIÓN

Este estudio se enmarca en las preocupaciones del Instituto Nacional para la Evaluación de la Educación (INEE) por avanzar en la construcción de indicadores longitudinales, a la luz del surgimiento de nuevas fuentes de datos que permiten, gracias al registro individualizado de los alumnos, su seguimiento en el tiempo.

De manera más precisa, se busca identificar en la totalidad de los registros de la base de datos del Registro Nacional de Alumnos (RNA) (inicio y fin de ciclos 2007-2008, 2008-2009, 2009-2010 y 2010-2011, e inicio de 2011-2012) poblaciones de interés para construir cohortes de estudio y poder seguirlas en el tiempo, con el fin de reconstruir sus trayectorias educativas.

En primer lugar, se trabajó en transformar las fuentes de datos de los distintos ciclos educativos que conforman la base de datos del RNA para construir una base única de carácter longitudinal donde la unidad de análisis la constituyeran los individuos, y cada ciclo educativo y las observaciones referentes a éste fueran columnas, lo que permitiría el seguimiento en el tiempo de distintos aspectos de la vida educativa de los mexicanos.

En segundo lugar, se seleccionaron dos cohortes de inicio según niveles educativos para seguir sus trayectorias en el periodo de tiempo del cual se tiene registro en el RNA. Esto permitió la construcción de dos bases de datos de carácter longitudinal: una que representa la cohorte de aquellos mexicanos que inician la primaria en el ciclo educativo 2008-2009, y otra de la que inicia secundaria en el ciclo educativo de 2008-2009. Se delimitaron muestras de ambas, construidas en torno a sus trayectorias educativas.

En tercer lugar, se procedió a un análisis de secuencias que incluyó la comparación y clasificación de las cohortes en tipologías, seguido de un análisis de composición sociodemográfica de dichas tipologías o grupos de trayectorias. Asimismo, se realizó la estimación de medidas estadísticas que permitieran resumir aspectos relevantes de esas trayectorias educativas, con el propósito de utilizarlas para futuros indicadores longitudinales. Estas medidas permiten una aproximación a los procesos de repetición, rezagos, promoción de grados y tasas de transiciones así como del papel que éstos juegan en la diferenciación de trayectorias educativas en cada una de las cohortes analizadas. ■■

2


ANTECEDENTES

Desde 2003, el Instituto Nacional para la Evaluación de la Educación (INEE) publica el anuario *Panorama Educativo de México, Indicadores del Sistema Educativo Nacional*, en el cual se describen la estructura y la dimensión del Sistema Educativo Nacional (población objetivo de la educación, cobertura, contexto socioeconómico); los agentes y los recursos (características de alumnos, docentes y directores; recursos informáticos; gasto público); el acceso y las trayectorias (matriculación, rezago, aprobación, deserción, abandono); los procesos educativos y la gestión (cantidad de cursos por docente), y los resultados educativos (aprendizaje, participación laboral).

Estas dimensiones, que componen el Sistema de Indicadores Educativos (SIE), han tenido leves variaciones y sus indicadores han sido construidos con base en las Estadísticas Continuas del Formato 911 (F911) de la Secretaría de la Educación Pública (SEP), en combinación con otras fuentes de información no diseñadas expresamente para la evaluación educativa (Censo Nacional de Población y Vivienda, Proyecciones de Población, Encuesta de Hogares).

El SIE reconoce entre sus antecedentes los indicadores construidos en los años sesenta en los Estados Unidos con el fin de evaluar la preparación recibida por los alumnos para enfrentarse a la competitividad académica internacional, y los construidos por la Organización para la Cooperación y el Desarrollo Económicos (OCDE) para evaluar distintas dimensiones de la educación. En la década de los noventa y en el contexto mexicano, la SEP publica información sobre el acceso y la trayectoria de los alumnos en los subsistemas educativos, y esboza las primeras ideas sobre el actual SIE, con mayor énfasis en la gestión administrativa de los procesos escolares que en la evaluación de la calidad educativa. En 2014, y en el marco de la reforma educativa, un avance en materia de información sobre el sistema educativo lo constituye el diseño y la aplicación del Censo de Escuelas, Maestros y Alumnos de Educación Básica y Especial (CEMABE). Este registro nacional de alumnos permitió complementar la información agregada captada por el Sistema de Estadísticas Continuas Formato 911 (F911).

Robles (2010) identifica tres etapas en el proceso de construcción del SIE, de las cuales se describen a continuación aquellos rasgos sobresalientes en tanto antecedentes al estudio que aquí se presenta. La primera etapa (de 2003 a 2004) consolidó una primera versión del




SIE cuyos indicadores fueron seleccionados y diseñados en función de la adaptación a indicadores internacionales y nacionales. La segunda (de 2005 a 2007) estuvo caracterizada por una mayor vinculación entre la SEP y el INEE, lo cual habilitó el consenso sobre metodologías y procedimientos de cálculo que permitió la depuración y reducción del número de indicadores presentados en anteriores años. En la tercera (de 2008 a 2010), el trabajo se orienta a precisar el modelo de evaluación de la calidad que subyace actualmente al SIE y a crear nuevos indicadores, entre los cuales sobresalen los longitudinales sintéticos. La incorporación de estos nuevos indicadores se acompaña de estudios empíricos y conceptuales que permiten sustentar marcos de referencia teóricos y normativos. En el caso de la creación de indicadores longitudinales a partir de métodos sintéticos, la discusión inaugurada por Robles (2009) es un antecedente significativo y que describiremos más adelante.

El modelo de evaluación de la calidad de la educación que subyace al SIE tiene un carácter sistémico que trata: el contexto (espacio socioeconómico que enmarca a los actores y las organizaciones escolares y que representa limitaciones o recursos para la tarea educativa); las necesidades sociales (valores, expectativas y aspiraciones compartidos con respecto a la utilidad y significados asociados a la educación); la planeación (aspectos organizacionales y curriculares), y el desempeño observado (Robles, 2010).

Entre las dimensiones consideradas en el *Panorama Educativo de México* (Robles *et al.*, 2008), interesa recuperar como antecedente la dimensión relativa a las trayectorias de los alumnos, que tiene como objetivo monitorear el avance escolar de los alumnos entre grados y entre niveles escolares para construir medidas de flujo entre grados y ciclos escolares, siempre con base en datos recogidos de manera agregada (F911). El consenso metodológico entre la SEP y el INEE con respecto a los indicadores de cobertura neta y al avance escolar en función de la edad o rango de edad normativo logrado a partir de 2006 permitió construir mediciones más pertinentes ante la ausencia de información longitudinal a nivel nacional.

Otro antecedente significativo refiere al método de cohortes reconstruidas de la UNESCO (1984). Se trata de un intento por superar los indicadores tradicionales (de tasa de cobertura, repetición, etcétera), caracterizados por la utilización de datos transversales de ciclos contiguos, para reconstruir cohortes ficticias. El método de cohortes reconstruidas simula el tránsito del alumno de un nivel a otro, estimando la repetición, deserción y egreso bajo el supuesto de que estas tasas permanecen constantes. Para esto, utiliza las matrículas de inicio de cada ciclo escolar y de cada grado de dos años consecutivos. A partir de esta información, la UNESCO calcula indicadores globales de la eficiencia del sistema educativo (coeficiente de eficiencia, los años-alumno por egresado y el abandono durante la trayectoria normativa).

Mientras que la falta de precisión de los indicadores tradicionales se debe a las reinsertaciones escolares de alumnos provenientes de otras cohortes y los traslados de alumnos entre escuelas, las dificultades asociadas a la utilización del método de cohortes reconstruidas se vinculan con el supuesto de invariabilidad de las tasas de deserción, repetición y egreso en muchos ciclos escolares, en un contexto de mejora gradual y continua del sistema educativo mexicano.



Robles *et al.* (2009) adoptan una variante que permite superar las dificultades asociadas con el supuesto de tasas fijas, pero no logran resolver el problema que suscitan los ingresos escolares y los traslados. El método de la cohorte reconstruida a partir de múltiples generaciones modela el paso de una cohorte a través de todos los grados de primaria o secundaria hasta dos ciclos después del normativo. Así, las tasas que construye varían entre ciclos y grados.

Los indicadores longitudinales presentados por estos autores dan cuenta de la eficacia y eficiencia del sistema educativo mexicano y sus avances en primaria y secundaria mediante la reconstrucción de cohortes con información proveniente de múltiples ciclos; se aproximan a un panorama más cercano al real que el que surge de indicadores tradicionales construidos con base en datos transversales. ■

3

OBJETIVOS GENERAL Y ESPECÍFICOS

El objetivo general de este trabajo es contribuir al desarrollo de indicadores longitudinales que caractericen las diferentes trayectorias educativas de los alumnos de educación básica de México, y medir el tránsito de alumnos en ciclos escolares consecutivos, así como entre niveles educativos (deserción, repetición de grados escolares). De lo anterior se desprenden los siguientes objetivos específicos:

- Depuración crítica de la base del Registro Nacional de Alumnos (RNA) de 2007 a 2011.
- Caracterización de trayectorias educativas más comunes en el Sistema Educativo Nacional.
- Análisis estadístico de las trayectorias educativas usando metodologías apropiadas para el caso, en función de variables auxiliares de los alumnos disponibles en la base de datos RNA (indicadores longitudinales).
- Análisis estadísticos complementarios de los sesgos de los indicadores educativos. ■■


4

JUSTIFICACIÓN Y ALCANCES

De acuerdo con el artículo 3º constitucional, debe garantizarse una educación de calidad para todos los niños que cursan educación básica y media superior. El análisis de trayectorias y el uso de técnicas de análisis longitudinal permiten observar algunos aspectos de dicha calidad y abordar preguntas fundamentales para la comprensión de los procesos educativos y los fenómenos asociados a ellos, por ejemplo: ¿cuáles son las trayectorias de los alumnos?; ¿cómo se distribuyen entre los miembros de una cohorte?; ¿cuán distintas son entre sí?; ¿qué patrones de trayectorias existen?; ¿hay patrones desiguales de trayectorias asociadas a desigualdades sociales y contextuales? y ¿cuáles son las tasas de transición entre grados y entre ciclos?.

Hasta el momento, en México, ha habido algunos intentos por dar respuestas a varias de estas preguntas. Por ejemplo la UNESCO (1984) propone el método de la cohorte reconstruida mediante la simulación de las trayectorias de los alumnos desde un grado inicial hasta la conclusión de la primaria y secundaria. Este método es una aproximación parcial a la estimación de la deserción y la repetición, porque no sigue estrictamente al mismo grupo de alumnos en el tiempo y en cada ciclo educativo se agregan los repetidores de ciclos anteriores, lo cual sesga las estimaciones y se intenta resolver con supuestos de tasas de repetición y deserción. Asimismo, los fenómenos que se estudian en educación son procesos sociales y acontecen en el devenir del tiempo, entre años específicos que conforman cohortes precisas de alumnos en ciclos educativos particulares. La salida de la escuela puede ser el resultado de un acontecer abrupto o de un devenir de repeticiones, y en tal caso tener más o menos probabilidad de suceder a distintos grupos poblacionales con distintas trayectorias anteriores.

En este sentido, la posibilidad que ofrece el Registro Nacional de Alumnos (RNA) de dar seguimiento individual en el tiempo abre el camino a construir indicadores de cohorte, de carácter longitudinal, sin tener que establecer supuestos de deserciones, repeticiones y movilidad entre estados, los cuales pueden ir cambiando en el tiempo y pueden guardar diferencias importantes entre subpoblaciones. De igual manera, la representatividad nacional del RNA permite acompañar al alumno a lo largo de su trayectoria aunque cambie de escuela y de estados. De esta forma, la movilidad entre escuelas o entre estados no altera las estimaciones de la deserción.



Este estudio es un primer ejercicio exploratorio de la base de datos generada con el RNA, que por ser un registro reciente y joven tiene aún pocas rondas de observación y sólo permite el seguimiento de los alumnos por un número reducido de años y ciclos educativos. De todas formas, en el estudio se propone una estrategia de análisis longitudinal de los procesos educativos desarrollando técnicas de análisis de trayectorias e indicadores que permitan dar cuenta de la diversidad de procesos que acontecen en el dominio de la escuela. De esta manera, a diferencia de los antecedentes mencionados en el apartado anterior, este documento debe entenderse como un avance preliminar de carácter exploratorio que, sin la intención de reemplazar indicadores oficiales obtenidos de análisis agregados, propone un conjunto de medidas que den cuenta del acontecer educativo de los alumnos mexicanos y de fenómenos como el rezago, la absorción, la deserción, la interrupción o la salida a partir de estudios de cohortes educativas reales. ■

5

METODOLOGÍA

5.1 Fuentes de datos

Base de datos RNA

Las bases de datos trabajadas en este proyecto se construyen a partir de la información provista por el Registro Nacional de Alumnos (RNA), compuesto por los registros administrativos que integran el Registro Nacional de Alumnos, Maestros y Escuelas (RENAME). Se trata de información oficial, de interés nacional y de uso obligatorio para la Federación, los estados, el Distrito Federal y los municipios. Es responsabilidad de la Secretaría de Educación Pública (SEP) generar en forma regular y periódica la información del RENAME y ponerla a disposición de los usuarios por medio de los mecanismos establecidos por la Secretaría (*DOF*, 2012, 8 de agosto).

El Instituto Nacional para la Evaluación de la Educación (INEE) dispone de esta información en lo que se denomina como Base de Datos RNA, que consiste —en el momento de esta investigación— en un conjunto de bases de datos con los registros administrativos de los alumnos del país —que proveen información personal y de sus estatus educativos— para los inicios y fines de los ciclos escolares 2007-2008, 2008-2009, 2009-2010 y 2010-2011, e inicio del ciclo escolar 2011-2012. El INEE puso dichas bases a disposición de esta investigación, que las convirtió en su fuente principal de información.¹ En el cuadro 1 se detalla el total de registros a nivel nacional en cada una de esas rondas de inicio y fin de ciclo referidas.

Cada registro de inicio y final de ciclo constituye un archivo (base de datos) con las mismas 12 variables; siete de ellas informan sobre datos personales del alumno y cinco sobre la situación académica, tal y como se detalla en el cuadro 2.

¹ Se tomaron como insumo aquellas que ya habían sido trabajadas por su personal académico, con las correcciones de incongruencias en variables como la Clave Única de Registro de Población (CURP).

Cuadro 1 Número de registros de la base de datos RNA

Ciclos de registro	Número de registros totales
Inicio ciclo 2007-2008	24 147 327
Fin ciclo 2007-2008	23 897 841
Inicio ciclo 2008-2009	23 953 059
Fin ciclo 2008-2009	22 678 303
Inicio ciclo 2009-2010	24 176 662
Fin ciclo 2009-2010	22 973 782
Inicio ciclo 2010-2011	24 116 451
Fin ciclo 2010-2011	24 727 538
Inicio ciclo 2011-2012	25 095 208

Fuente: elaboración propia.

Cuadro 2 Variables en base de datos de RNA

Variables	Descripción
Identificador	Llave que identifica el registro en el período escolar correspondiente
CCT	Clave del Centro de Trabajo del alumno
Turno	Turno en el que estudia el alumno
Grado	Grado escolar en el que está inscripto el alumno en ese periodo escolar
Grupo	Grupo
CURP	Clave Única de Registro de Población
Primer apellido	Primer apellido del alumno
Segundo apellido	Segundo apellido del alumno
Nombre	Nombre del alumno
Sexo	Sexo del alumno
Fecha de nacimiento	Fecha de nacimiento del alumno
Entidad de nacimiento	Entidad de nacimiento del alumno

Fuente: elaboración propia.

El RNA es fundamental para cumplir los objetivos planteados en este estudio, porque permite la identificación y construcción de trayectorias educativas de todos los alumnos registrados en el sistema mexicano, en los periodos mencionados.

Respecto al estado de situación de los registros de inicio y fin de los ciclos, se constató falta de información para varias entidades federativas, como se observa en el cuadro 3. Debido a este problema no es posible desagregar los datos por entidad federativa, ya que en muchos casos se carece de bases completas, sobre todo en relación con los fines de ciclo de varios estados. Respecto al ciclo 2007-2008, por el peso que tiene en la población nacional el Estado de México, su ausencia en el fin de ciclo del periodo 2007-2008 torna muy débiles los registros. Por esto, se decidió en este estudio trabajar con los datos a partir del ciclo 2008-2009.

Cuadro 3 Estado de situación respecto a registros faltantes por ciclo

Ciclos de registro	Problemas de registro
Inicio ciclo 2007-2008	Coahuila presenta datos faltantes en la variable turno de todos sus registros
Fin ciclo 2007-2008	Falta la base completa del Estado de México y Morelos
Inicio ciclo 2008-2009	En Yucatán las variables Clave CCT, turno, grado y grupo presentan datos faltantes en todos los registros
Fin ciclo 2008-2009	Falta la base completa del Estado de Hidalgo, Morelos y Quintana Roo Falta la base completa del Estado de Guanajuato
Fin ciclo 2009-2010	En la base del Estado de Guerrero la variable turno presenta datos faltantes en algunos registros En el Estado de Quintana Roo la variable grupo presenta datos faltantes en todos los registros En el Estado de Sinaloa las variables de turno y grupo presentan datos faltantes en algunos registros
Inicio ciclo 2010-2011	En el Distrito Federal la variable grupo presenta datos faltantes en algunos registros En el Estado de Chiapas la variable turno presenta datos faltantes en algunos los registros
Fin ciclo 2010-2011	En el Distrito Federal las variables turno y grado presentan datos faltantes en algunos registros En el Estado de Yucatán las variables turno, grado y grupo presentan datos faltantes en algunos registros
Inicio ciclo 2011-2012	En el Estado de Baja California la variable grupo presentan datos faltantes en algunos registros

Fuente: elaboración propia con base en Valencia (2012).

Respecto a la calidad de los registros existentes, la consistencia de sus variables y la calidad de su información, se han encontrado índices aceptables en las evaluaciones realizadas (Valencia, 2012). Dos variables de la base de datos son muy importantes, la CURP y la CCT, porque resumen una gran cantidad de información del alumno y del centro donde estudia. Se exploró la estructura de la CURP; se buscaron símbolos, caracteres y dígitos, y además se contabilizaron las homonimias con la finalidad de determinar la exactitud con que esta variable es reportada en el RNA. El análisis sobre la estructura de la CURP en el RNA arrojó niveles superiores a 96.7% en los registros de inicio y fin del ciclo escolar. Respecto a la clave única de los centros educativos, se obtuvieron porcentajes de exactitud en la estructura de la CCT que oscilan entre 99.9 y 100%, lo que indica una buena calidad de las estadísticas.

Mediante un proceso de codificación, la CURP se sustituyó por un número único que identifica al individuo (llamado ID), pero que ya no refleja datos confidenciales. Esta variable se construyó asignando códigos aleatorios únicos para cada una de las claves únicas de registro de la persona y adheridos a los códigos de CURP referentes a la fecha de nacimiento. La matriz que relaciona cada CURP con el nuevo código identificador del individuo (ID) se entregó al INEE para que el organismo pueda realizar futuros seguimientos de casos.

Construcción de base única longitudinal

De esta manera, con la totalidad de las bases del RNA (para los inicios y fines de los ciclos escolares 2007-2008, 2008-2009, 2009-2010 y 2010-2011, e inicio del ciclo escolar

2011-2012) se procedió al armado de una base única longitudinal cuya unidad de análisis son los individuos identificados por el nuevo código identificador único ID (construido con base en la CURP), y todas las variables de cada uno de los periodos de observación de inicio y fin de ciclos fueron integrados como columnas. Esto permitió realizar el ejercicio de recuperación de la información que originalmente faltaba por problemas de registro en algunos periodos para algunos individuos. Siempre que se contó con información del grado en el que está inscrito el alumno en dos mediciones para cada ciclo, inicio y fin se pudo completar información faltante a la luz de ambas instancias, en concordancia y verificación lógica con la información del ciclo anterior y el siguiente. En otro tipo de casos, la ausencia de registro puede deberse a interrupciones reales que el alumno pueda haber tenido en su trayectoria educativa; estas situaciones se observan cuando desaparece el alumno del registro por un ciclo entero y luego reaparece en otro momento, inscrito en el mismo grado en que estaba cuando se fue, o en algún grado siempre menor al que el estudiante tendría de no haber estado fuera del sistema. En estos casos, el dato faltante no se corrige, y se interpreta como interrupciones en la trayectoria educativa.

Después del análisis de la calidad y cobertura de la información contenida en el RNA se decidió trabajar con los registros a nivel nacional a partir del ciclo 2008-2009 y con la información de inicio de cada ciclo usando los datos de fin de ciclo para corregir y complementar la información existente y la faltante. Trabajar con los datos de la RNA a escala nacional permite aprovechar una de las grandes ventajas de este tipo de registros, que es la posibilidad de realizar un seguimiento individual de los alumnos con independencia de su movilidad entre estados y escuelas.

No obstante, cabe señalar que la información respecto al estatus educativo registrado en cada ciclo no reporta la aprobación del grado en que el alumno está inscrito, así como tampoco existe información sobre materias reprobadas o exámenes extraordinarios en secundaria. Esta ausencia debe ser tenida en cuenta a la hora de analizar los datos, sabiendo que la aprobación se observa al verificar que en el inicio del ciclo posterior el alumno esté inscrito en el grado siguiente. Por lo tanto, los datos con que se cuenta para trabajar no son indicadores precisos de aprobación cuando hay trayectorias interrumpidas o salidas del sistema educativo, pues no permiten saber si el alumno aprobó o no el último grado en el que estaba registrado.

Asimismo, gran parte de las limitaciones en el análisis impuesto por los datos del RNA refieren a que aún es muy estrecha la ventana de observación; los periodos o ciclos educativos durante los cuales pueden seguirse los individuos constituyen un factor de truncamiento en las trayectorias educativas que hay que tener presente. Obviamente, este aspecto podrá ser superado en futuros proyectos, a medida que se sobrepongan más periodos de registro y se procesen y preparen para su uso los registros del RENAME del 2011-2012 a la fecha.

Selección de población de interés: dos cohortes educativas

A los efectos de este estudio, se decidió seleccionar de los registros de la base única longitudinal de RNA dos cohortes educativas que permitieran observar trayectorias educativas en dos ciclos educativos diferentes: a) una cohorte que inicia primaria en el ciclo 2008-2009, y

b) una cohorte que inicia secundaria en el ciclo 2008-2009. De todas maneras, esta base completa, única y longitudinal queda también como producto del trabajo realizado en este proyecto que puede servir de insumo para seleccionar otro tipo de cohortes de interés o poblaciones específicas donde se quieran observar distintos tipos de trayectorias, transiciones y fenómenos particulares en primaria, secundaria o ambos ciclos.

La cohorte de primaria se selecciona tomando como criterio que el alumno estuviera inscrito en primer grado de primaria en el inicio del ciclo 2008-2009, por lo cual no siempre se toman en cuenta alumnos de primer ingreso, es decir, pueden haber cursado ese mismo grado en periodos anteriores. La cohorte de secundaria se selecciona a partir de quienes estaban inscritos en primer grado de secundaria en el inicio del ciclo 2008-2009 y que, según la información con que se cuenta, en 2007-2008 cursaron sexto grado de primaria o que, si no cumplían con esa condición, no cursaron primero de secundaria también en 2007-2008. Así se intentó obtener una cohorte de inicio de secundaria de nuevo ingreso, o de alumnos que habían cursado primero de secundaria en etapas anteriores a 2007-2008, con alguna interrupción, y que habían ingresado nuevamente en 2008-2009. La base de datos de la cohorte de secundaria, al estar construida de esta manera, permite dar cuenta de todos aquellos alumnos que en el ciclo anterior cursaron sexto grado de primaria y estaban en posibilidad de entrar a secundaria, aunque no lo hayan hecho, lo cual permite observar la magnitud de este fenómeno. Esto también hace posible ver movilidad de modalidades de servicio escolar entre primaria y secundaria.

En el panorama que proporcionan los registros existentes hasta el momento puede observarse a los estudiantes que ingresan a secundaria en agosto del 2008, y seguirlos por los tres periodos siguientes y el inicio de uno más; la base de datos del RNA no tiene información sobre el ingreso a la preparatoria, pero el inicio del ciclo 2011-2012 —es decir, el cuarto año— permite dar cuenta de trayectorias de alumnos que repiten algún grado de secundaria o tienen interrupciones. Asimismo, para aquellos alumnos que tienen una trayectoria normativa, permite verificar la aprobación del último grado de secundaria si a principios del ciclo 2011-2012 no están repitiendo tercer grado. Respecto a la cohorte de primaria, no se pueden reconstruir trayectorias completas, y es una vía de entrada para analizar las trayectorias tempranas en la primaria, y valorar repeticiones, interrupciones y rezagos que se producen en esos primeros años.

Las cohortes de primaria y secundaria que inician en el ciclo 2008-2009 suman 2 638 735 y 2 131 087 casos, respectivamente. Ambas bases de datos ya integran las sucesivas observaciones de cada inicio y fin de ciclo de manera conjunta en una sola base, a modo de columnas. De acuerdo con las 12 variables contenidas en el registro de alumnos, se cuenta con el nuevo código identificador del individuo ID (construido con base en la CURP), así como con información de la entidad federativa por fecha de nacimiento, edad y sexo. Luego, se organizan en columnas las variables que cambian en el tiempo: nivel, grado, turno, CCT, entidad federativa donde se cursa el grado en cada periodo de inicio de 2008 a inicio de 2011, y las correspondientes observaciones de cierre.

Con las bases de las dos cohortes se procedió a construir en ellas una variable que diera cuenta de la secuencia educativa en el periodo observado. Es decir, una variable *string* que describiera el grado cursado en cada periodo de observación trabajado.

En este sentido, para la cohorte de primaria se obtuvo una variable *secuencia* que tomaba valores concatenados del grado cursado en cada inicio de ciclo 2008-2009, 2009-2010, 2010-2011 y 2011-2012. A modo de ejemplo, para un individuo 1 podría tomar valores 1234; para un individuo 2 podría tomar valores 1123, y 1*22 para un individuo 3. El primer caso indica que en cada inicio de ciclo observado el individuo cursaba un grado distinto y normativo, empezando por primero de primaria y hasta alcanzar cuarto año en el último periodo observado. El individuo 2 repite primer grado una vez, y el individuo 3 presenta una interrupción (o falta en el registro del ciclo 2009-2010 del RNA) y luego retoma segundo año de primaria, grado que a su vez repite.

Lo mismo se realizó con la cohorte de secundaria, pero en este caso se incluye el grado en el que se cursaba el ciclo anterior al ingreso (2007-2008), que sólo podía ser sexto de primaria o ninguno (se debe tener presente que muchos de los casos que aparecen como interrupciones en 2007-2008 pueden ser problema de calidad del registro, que por corresponder al primer año de práctica del RNAME no permiten una corrección lógica del dato faltante). En este sentido, se trabajó con una variable *string* de secuencia que tomaba valor 6 para sexto de primaria, y una sucesión de 7 a 9 entre primero y tercero de secundaria.

Selección de muestras

Una vez construida la variable *secuencia* en las bases completas de ambas cohortes de interés se estima la distribución entre distintas secuencias o trayectorias educativas. El objetivo de este procedimiento es seleccionar dos muestras, de 5000 casos cada una, que permitieran desarrollar un análisis de secuencia más exhaustivo con un menor número de casos. Estas muestras se componen de trayectorias individuales y no exclusivamente de individuos, lo cual permite utilizar un factor de ponderación que permita conocer distribuciones para el total de casos de la cohorte de interés respectiva.

Para el caso de la cohorte de primaria se encontró que ocho tipos de secuencias agrupan 94.5% de la experiencia de la cohorte, y se clasificó entonces en una novena tipología 5.5% de las secuencias. Con esa distribución se procedió a estimar una muestra aleatoria estratificada por las distribuciones de secuencias entre esos nueve tipos. La muestra obtenida es una muestra de secuencias o trayectorias, y presenta una distribución entre tipos de trayectorias educativas similar a la observada en la base completa.

Para el caso de la cohorte de secundaria se realizó el mismo procedimiento, sólo que considerando diez tipos de secuencias que concentraron 97.2% de las experiencias de la cohorte, y un tipo 11 representó 2.8%.

5.2 Análisis y técnicas utilizadas

Análisis de secuencias

A partir de la definición de los estados posibles de las dos cohortes analizadas, se construyen las secuencias en el rango de años del calendario de observación, y se realiza un análisis de secuencia propiamente dicho. El conjunto de secuencias trabajadas puede analizarse por medidas transversales o longitudinales. El uso de estas medidas está condicionado por el objeto de estudio construido en torno a la secuencia y las preguntas y objetivos planteados en la investigación.

En este estudio se utilizan la distribución de frecuencias de las distintas secuencias y recursos visuales que permiten apreciar la diversidad de situaciones —gráficas de trayectorias individuales e histogramas—. Como medida de impredecibilidad (o incertidumbre para predecir las trayectorias) se trabajó con el *índice de entropía transversal* (Billari, 2001, y Fussell, 2006), que muestra si el ciclo educativo en el que se encuentran los individuos son un buen predictor de estados de una trayectoria particular. Este índice permite obtener una medida resumen que da cuenta del grado —de primaria o secundaria para cada periodo observado— en el cual los individuos —en un mismo momento de observación (periodo o ciclo educativo)— son similares en sus estados.

La técnica propuesta consiste en calcular medidas de entropía conjuntas a partir de los estados en los que están posicionados los individuos en una edad particular o periodo de observación específico.² Se calculan entropías a partir de los distintos grados académicos en que se encuentran los alumnos (primero, segundo o tercer grado) en cada ciclo educativo (agosto de los años 2008, 2009, 2010 y 2011). El índice de entropía varía de 0 (cuando la homogeneidad es perfecta y todos los individuos están concentrados en un solo estado, por ejemplo, todos en primero grado) a un máximo, definido como entropía máxima, que depende del número total de estados posibles. Para que la medida resumen sea más fácil de interpretar, se normalizan los valores del índice en un porcentaje de la entropía máxima, y de esta forma, mientras mayor sea la proximidad a 1 en un periodo de observación, mayor será la dispersión de los individuos entre los diferentes estados.

El potencial explicativo de esta medida reside en su capacidad de medir diferencias en el calendario de rupturas de las trayectorias educativas. Si toda la cohorte que inicia en un ciclo educativo (2008-2009) y un mismo grado (primaria o secundaria, según la cohorte trabajada) sigue una trayectoria educativa exitosa y normativa, lineal y regular, en cada nuevo ciclo educativo todos los individuos estarían en el grado consecutivo al anterior hasta el fin del periodo de observación. Es decir, que el índice de entropía para cada ciclo educativo observado se mantendría siempre en 0, lo que se interpretaría como una total homogeneidad de estados

² Ecuación 1: $E = \sum p_s \ln(1/p_s)$

Donde: s = número de estatus, y p_s = la proporción de población en el estatus.

entre los individuos en cada periodo de observación. En la medida en que algunos individuos salen de la escuela o repiten grados —y lo hacen con distintos calendarios—, se deja de observar en cada ciclo escolar homogeneidad en los grados en que se encuentra la cohorte analizada. La fuerza con la que la entropía se incrementa entre periodos o ciclos educativos de observación (se aleja de 0) y las brechas que existan en las entropías entre subgrupos de población (sexo, por ejemplo) señalan que los miembros de una cohorte de primaria o secundaria comienzan a experimentar procesos diferenciales con intensidades también diferentes y en calendarios disímiles, aunque todos compartan el mismo punto de partida.

La comparación de secuencias es el aspecto medular de esta estrategia analítica; muestra la forma en que las secuencias se generan y sus características estructurales. El análisis de alineación óptima (OMA por las siglas en inglés de *optimal matching analysis*) desarrollado por Abbott permite encontrar patrones en las secuencias con base en una medida de la proximidad o semejanza entre ellas. Las secuencias de eventos se alinean por pares y se transforma una secuencia en la otra a partir de inserciones, borrados y sustituciones de estados: modificaciones que tienen "costos". El resultado del OMA son las transformaciones entre pares de secuencias con costo mínimo. Las medidas de semejanza entre las trayectorias conforman la matriz de distancias interindividuales. Esa matriz de distancias interindividuales se utiliza como insumo para el análisis de un agrupamiento (clúster).

Para validar el número de agrupaciones (es decir, la solución de clúster elegida para distinguir y agrupar a la vez el cuerpo de trayectorias educativas) se utilizó la técnica *silhouettes* (Rousseeuw, 1987), donde dado un conjunto de grupos (A , B , C) y dada una solución posible de clúster, para cada grupo de objetos distinto de A se calcula el promedio de las distancias con sus objetos.

$a(i)$ = disimilaridad promedio de i con todos los otros objetos del grupo A .

$d(i,C)$ = disimilaridad promedio de i con todos los otros objetos del grupo C (siendo C distinto de A).

Luego de hacerlo para todos los grupos $d(i,C)$, se selecciona el de menor distancia (el vecino). El vecino es la segunda mejor opción de grupo al cual pertenecer un objeto i de un grupo A .

$b(i)$ = mínimo $d(i,C)$.

Cuando $s(i)$ se aproxima a 1, significa que la disimilaridad intragrupo es mucho más pequeña que la disimilaridad entre grupos, lo cual se considera una buena clasificación; el grupo vecino no estaría tan cercano como el grupo al que efectivamente pertenece en esa solución de clúster. Que $s(i)$ se acerque a 0 indica que ese objeto estaría equidistante del grupo al que pertenece y de su vecino. La peor situación se presenta cuando $s(i)$ es próximo a -1, porque ese objeto estaría más cercano al grupo vecino que al grupo al cual efectivamente pertenece, en una solución particular de clúster, lo que se considera una mala clasificación.

En este sentido, la técnica de *silhouettes* arroja información sobre cada uno de los objetos (casos) que pertenecen a un grupo en una solución dada de clúster (grupo vecino) y una medida promedio de los $s(i)$ de ese grupo que puede variar entre -1 y 1, y finalmente una medida de la solución de clúster en su conjunto (*average silhouettes width*). Con esos elementos, el investigador tiene elementos para determinar un número específico de grupos con los cuales trabajar que depende de una partición específica de los objetos y no del algoritmo de clasificación que se utilizó para obtener esa partición. Los resultados que arroja esta técnica permiten distinguir —en una solución de k grupos de clúster— cuál o cuáles son los grupos más débiles y cuáles no, y así verificar si en una nueva solución de clúster de $k+1$ grupos esa situación mejora.

Para los fines de este trabajo se buscó una manera parsimoniosa de clasificar la diversidad de trayectorias educativas que presentan los alumnos de las distintas cohortes de estudio. En este sentido, luego de la alineación de secuencias mediante *optimal matching* y el análisis de clúster con las matrices de distancias interindividuales, la técnica de *silhouettes* proporciona una forma de verificar qué pasa con los objetos en el interior de las distintas particiones de una solución de clúster. De esta manera, es posible agrupar la diversidad de trayectorias con otras iguales o similares, y distinguir un tipo de grupos con particularidades propias para su análisis.

Estimación de medidas resumen de las trayectorias educativas

La construcción de tres medidas resumen a partir de las trayectorias educativas identificadas en ambas cohortes de estudio (estimación y distribución de duraciones por grado educativo; estimación y distribución de número de transiciones realizadas en cada cohorte; matriz de transiciones entre estados) permite una aproximación a los procesos de repetición de grado, promoción y rezago, analizados a su vez por covariantes, subgrupos de poblaciones relevantes para la investigación planteada.

La estimación de duraciones por grado proporciona una medida agregada de las veces que los individuos aparecen cursando un mismo grado educativo (definido como estado en nuestro estudio); permite además un acercamiento a los procesos de repetición de grado.

La estimación de transiciones realizadas por los individuos permite identificar en trayectorias analizadas en ambas cohortes cuántas transiciones entre estados realizan dando una magnitud a los procesos de promoción efectiva de grado así como a los procesos de rezagos.

La estimación de la de matriz de transiciones entre estados arroja una matriz con tasas de transiciones entre estados a partir de las secuencias. Las tasas de transiciones son probabilidades de transición de un estado a otro observadas en las secuencias.

5.3 Selección de covariantes

En el momento de seleccionar variables de interés analítico e indagar en su relación con las trayectorias educativas, la base de datos RNA presenta algunas limitaciones por el número reducido de elementos personales y contextuales sobre el que se indaga en el registro de alumnos. Para los fines de este proyecto, se ha decidido trabajar con dos variables individuales y una a nivel de escuela. Como variables individuales se seleccionan el sexo, que es una variable fija en el tiempo, y la edad, que si bien no es fija, se construyó como tal.

La edad obviamente varía en cada ciclo educativo; para poder fijarla al inicio de la trayectoria educativa se construyó como la relación entre la fecha de nacimiento de los individuos en el registro y el rango de edades normativas para primer grado escolar que inicia en agosto de 2008. De esta manera, la variable da cuenta de si el individuo está en el rango de edades normativas para cursar primer grado de primaria, en un rango de edades menores a la normativa o en extraedad. Cabe precisar que estas tres situaciones no sólo describen los años vividos por estudiante, sino que son también un reflejo de su trayectoria educativa anterior en el momento del inicio de la trayectoria que se está estudiando. La extraedad puede haberse adquirido por el inicio tardío de la primaria, pero también por repeticiones o interrupciones en periodos anteriores.

La variable de tipo contextual es la modalidad del servicio escolar, que también varía en el tiempo porque los individuos pueden moverse entre distintas escuelas durante su trayectoria educativa, y con ese movimiento cambiar la modalidad del servicio. Para obtener una variable fija en el tiempo se toma la modalidad de servicio al inicio de la trayectoria, primero de primaria en la cohorte primaria y primero de secundaria en la cohorte secundaria. Esta decisión descansa sobre un análisis de movilidad de servicios realizado para ambas cohortes de estudio que se presentará en el apartado correspondiente de los resultados de la investigación. ■

6

MARCO DE REFERENCIA

La idea de secuencia no es exclusiva ni originaria de las ciencias sociales, pero hay razones epistemológicas y teóricas que permiten recuperar esta idea, razones que permiten con nuevos métodos analíticos arrojar luz sobre viejas preguntas y temas teóricos respecto a los procesos sociales. Desde el punto de vista teórico, Andrew Abbott ya había destacado la importancia de considerar el tiempo en los fenómenos sociales y conceptualizarlo en sus múltiples expresiones (temporalidad, orden, procesos y etapas). Conceptualizar los fenómenos sociales como procesos, con información de su acontecer temporal, da al investigador la posibilidad de diferenciar fases dentro de éstos, así como los efectos causales de dichas fases. Para Abbott los factores causales no actúan con influencia simultánea, por lo cual, si se alcanza a descomponer un fenómeno en fases adecuadas, se tendrían alcances también a explicaciones específicas.

Para Abbott la realidad social acontece en historias (narrativas), y esas narrativas tienen propiedades: encadenamiento, orden y convergencia. El encadenamiento se define como la naturaleza que liga una etapa con la siguiente; es en la narrativa el análogo de la causalidad. Una historia presupone un particular y exacto orden de eventos que la constituye como tal. La noción de convergencia corresponde al grado en el que una secuencia social se aproxima a un estado estable. En este sentido, Abbott entiende una secuencia como una lista ordenada de elementos; ese orden tiende a ser temporal, pero es válido también pensarlo como un orden espacial.

Los eventos en una secuencia pueden ser únicos o repetidos; los estados de las secuencias tienen una dependencia entre sí, y esos niveles de dependencias pueden variar entre un total de secuencias. En este sentido, la secuencia puede ser variable dependiente o independiente en una investigación. Puede tenerse interés en los patrones de un conjunto de secuencias, o interesar lo que acontece antes de una secuencia y que hace posible su existencia particular. Asimismo, puede interesar un conjunto de secuencias que deriven en el fenómeno de interés (ejemplo: cuáles son las secuencias educativas precedentes asociadas a periodos de interrupción o salida de la escuela).

De esta forma, Andrew Abbott encontrará en el modelo analítico de las ciencias biológicas para el análisis de las cadenas de ADN el modelo analítico para analizar el acontecer de la realidad social en tanto proceso (encadenamiento de secuencias). A su vez, estos aportes de Abbott convergieron con la propuesta teórica analítica de Curso de Vida, que ya tenía


un desarrollo importante de herramientas analíticas de los procesos sociales y administración del tiempo en sus múltiples expresiones en ellos. Dentro de la perspectiva del Curso de vida, las aproximaciones “holísticas” proponen como objeto de estudio las trayectorias de vida individuales como una unidad conceptual (Billari y Piccarreta, 2005). De esta manera, la perspectiva holística (en contraposición con la que Billari denomina *pragmática*) enfatiza trabajar con trayectorias por encima de transiciones o estatus simples que permitan dar cuenta de procesos de continuidad, rupturas y quiebres en el curso de vida. El análisis de trayectorias de vida permite captar propiedades fundamentales de la perspectiva del curso de vida y del curso de vida como objeto de estudio (que aparecen desdibujadas desde el abordaje de estatus o transiciones simples), y que Abbott (2001) reconocía como propiedades principales de las narrativas: encadenamiento, orden y convergencia. Identificar el entrelazamiento entre los múltiples estatus a través de los cuales los individuos despliegan su vida, qué estatus es posible combinar (integrar) en un momento dado, cuáles se excluyen, cómo se ordenan, y cuáles no son vías de acceso a otros dominios de vida y cuáles sí no sólo nos permite conocer los caminos posibles que los individuos emplean en la construcción de sus vidas, sino también dar cuenta de procesos más generales, como los procesos de integración y diferenciación de los cursos de vida en sociedades concretas.

Entonces, si la realidad social sucede en narrativas, esas narrativas pueden ser descritas por cuatro elementos fundamentales de la perspectiva de curso de vida: trayectorias, estados, eventos y transiciones. Una *trayectoria* puede verse como un modelo de relativa estabilidad y cambio a largo plazo, como estado de una secuencia de inserciones de papeles que representa la variación a lo largo del tiempo de uno o más papeles individuales en distintas esferas de vida. Un *estado* refiere a un periodo de vida de relativa estabilidad estructural o funcional. Una transición da cuenta del cambio al final de un estado, implica entonces un cambio de estado. Un evento es lo que ocurre en un cierto tiempo y lugar, además provisto luego de sentido por los seres humanos. Estos eventos pueden ser normativos (matrimonio, graduación, retiro) o no normativos (enfermedad, guerra, muerte). Los eventos importantes están siempre asociados con transiciones.

Si se miran las secuencias por fuera del paradigma teórico analítico de curso de vida como rituales, interacciones interpersonales, textos orales y escritos, evoluciones históricas de grupos o regímenes, también pueden concebirse como procesos y descomponerlos en unidades de tiempo y escalas. Pero esas unidades de tiempo pueden ser simbólicas o artificiales en cierto sentido, dice Abbott, como los pasos de un ritual o las etapas para la manufactura de un producto. Esto permite identificar trayectorias, estados, eventos y transiciones en una multiplicidad de fenómenos y extender a ellos el análisis de secuencias propiamente dicho.

El núcleo del programa de análisis de secuencia fue estructurado entre 1990 y 2000 aproximadamente, alrededor de un número limitado de opciones metodológicas. Se estudiaban trayectorias de vida individuales en periodos contemporáneos. Las secuencias compartían la misma extensión, y la edad era el eje temporal expresado sobre todo en años.

Para la construcción de las secuencias como tales, se deben tomar al menos dos decisiones: si se trabaja con eventos o estados, y si se trabaja con referencia de tiempo interna o



externa al individuo. Cada cambio de estado es un evento y cada evento implica un cambio de estado. Sin embargo, el estado que resulta de un evento puede también depender del estado previo y de todos los eventos ya ocurridos anteriormente. En este sentido, puede ser de interés para el investigador mantener observaciones en unidades de tiempo donde no existan cambios de estado.

La otra decisión implica considerar si se tomará como unidad de tiempo para observar los estados o eventos lapsos individuales, como la edad, o lapsos externos, como años o fechas del calendario, que refieran a cohortes educativas, a cohortes con interés económico, etcétera. Como la edad individual es un elemento muy importante en el estudio de trayectorias, cuando se trabaja con periodos del calendario se pueden estudiar estados que correspondan con ciertos rasgos comunes, como: primer desempleo entre 2000 y 2010, o primera repetición escolar entre 2000 y 2010. Cabe señalar que las posibilidades y las decisiones dependerán de la disponibilidad de opciones dada la forma en la que se recabó la información original.

Como un modo de organizar la diversidad de análisis posibles es posible diferenciarlos entre los que intentan caracterizar las secuencias y los que luego intentan compararlas (agruparlas o diferenciarlas). Mediante el uso de paquetes de análisis estadístico y procesadores de datos pueden construirse estas bases de datos de secuencias para realizar posteriormente un conjunto de análisis que permita describirlas y caracterizarlas. Asimismo, pueden compararse con un análisis de similitud y distancias que puede luego asociarse a otras variables individuales y de contexto. ■ ■

7

ANÁLISIS LONGITUDINAL DE LOS DATOS

7.1 Primaria

Composición de la cohorte de ingreso a primaria

La base de datos empleada para el análisis de las trayectorias educativas de los ingresantes a primaria en agosto del 2008 es una muestra de 5 000 casos que contiene todas las variables de la base de datos original del Registro Nacional de Alumnos (RNA) para cada periodo de observación, con la cual se construyeron los estados (grado cursado en el periodo) y las variables que se utilizarán como covariantes en el análisis.

La composición por edad de la cohorte primaria puede observarse en el cuadro 4 donde se muestran algunas medidas resumen de la fecha de nacimiento. Quienes están en edad normativa para cursar primer grado de primaria en agosto de 2008 son aquellos nacidos entre el 1 de enero de 2002 y el 31 de diciembre de 2002. A partir de estas fechas de nacimiento puede saberse que 80% de la cohorte se encuentra en edad normativa y 20% en extraedad.

Cuadro 4 Medidas resumen de la fecha de nacimiento

	Fechas de nacimiento
Edad mínima	03/10/89
Primer cuartil	25/01/02
Media	27/05/02
Mediana	17/04/02
Tercer cuartil	07/09/02
Máximo	31/12/02

Fuente: elaboración propia con muestra cohorte primaria.

Respecto a la composición por sexo, de 5 000 casos 2 614 son varones (52.3%) y 2 386 son mujeres (47.7%). En el cuadro 5 puede observarse la composición conjunta de sexo y edad, donde la relación 80% a 20% entre quienes están en edad normativa y quienes están en extraedad se modifica un poco entre sexos. Los varones son los que alcanzan 21.3% en extraedad, mientras las mujeres, 18.5%.

Cuadro 5 Composición por sexo y edad en el momento de ingreso a primaria (%)

Edad	Sexo	
	Varones (%)	Mujeres (%)
Menor	0.00	0.00
En edad	78.70	81.50
En extraedad	21.30	18.50
Total	100.00	100.00

Fuente: elaboración propia con muestra cohorte primaria.

Con relación a la modalidad de servicio de la escuela en la que se inicia primaria en agosto de 2008 es necesario hacer algunos señalamientos. La modalidad de servicio puede cambiar con cada periodo de observación, como en el caso en que hay un cambio de escuela y modalidad entre un año y el otro. Con tales fines se realizó un análisis de movilidad entre modalidades de servicio en los cuatro años de observación para ver si era posible tomar como covariante en nuestro análisis la modalidad al inicio de los ciclos analizados.

Las modalidades de servicio escolar que aparecen en la base completa de la cohorte de primaria analizada, y luego en la muestra de 5 000 casos seleccionada son dos: general e indígena. En el cuadro 6 puede observarse la distribución en la base muestra.

Cuadro 6 Distribución de modalidades de servicio

	Casos	%
General	4 710	94.20
Indígena	290	5.80
Total	5 000	100.00

Fuente: elaboración propia con muestra cohorte primaria.

Con el fin de observar cambios de modalidad, se codificaron, para cada uno de los cuatro periodos observados, el ciclo educativo desde agosto de 2008 a agosto 2011 y la modalidad de la escuela en la que se cursaban los distintos grados de primaria. Cuando el individuo no

está registrado en ese periodo, no hay modalidad de servicio alguna, y se denota como NA. La modalidad General se codifica como GEN, y la modalidad Indígena como IND. De esta forma se construyó una secuencia de modalidades para ver si se modifican a lo largo de los cuatro periodos. En el cuadro 7 se pueden observar las 10 secuencias de modalidades más frecuentes para la cohorte de primaria que inicia en agosto de 2008. En la gráfica 1 se pueden observar las secuencias individuales de modalidades de servicio para la muestra en los cuatro periodos observados, los ciclos educativos que van desde agosto de 2008 hasta agosto de 2011. Cada secuencia individual tiene cuatro modalidades posibles, una por periodo de observación, donde está incluida la posibilidad de ausencia de registro (NA). Tan sólo 10 secuencias resumen 99% de las existentes en la cohorte.

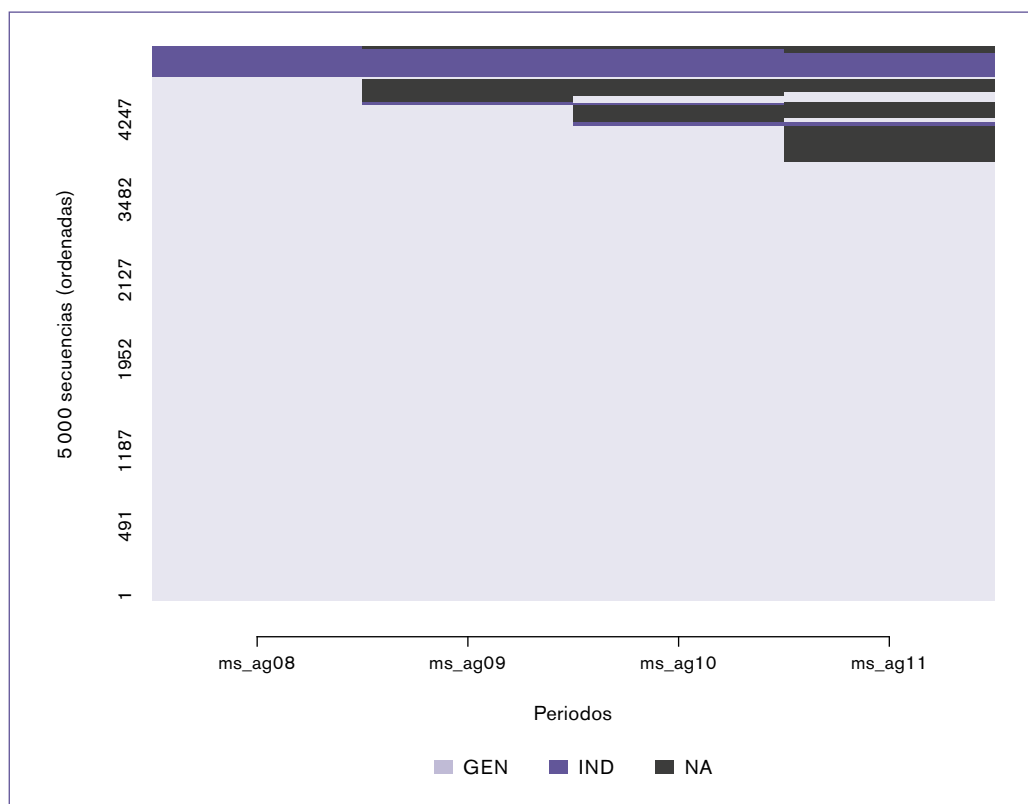
Cuadro 7 Las 10 secuencias de modalidades más frecuentes

Secuencia	Casos	%
GEN-GEN-GEN-GEN	3972	79.4
GEN-GEN-GEN-NA	320	6.4
IND-IND-IND-IND	221	4.4
GEN-NA-NA-NA	150	3.0
GEN-GEN-NA-NA	119	2.4
GEN-NA-GEN-GEN	55	1.1
GEN-GEN-NA-GEN	45	0.9
IND-IND-IND-NA	28	0.6
IND-NA-NA-NA	25	0.5
GEN-NA-GEN-NA	20	0.4
Totales	4955	99.1

Fuente: elaboración propia con muestra cohorte primaria.

Puede deducirse de dichas secuencias y sus distribuciones que la migración entre modalidades de servicio es muy baja, y por lo tanto es posible mantener como covariante de la secuencia la modalidad al inicio de la primaria, es decir, la modalidad de servicio de la escuela donde se cursó primer grado.

Gráfica 1 Distribución de secuencias individuales de modalidad de servicio de escuelas



Fuente: elaboración propia con muestra cohorte primaria.

Las trayectorias educativas

Una vez definidos los estados que los alumnos podrían tener en los diferentes periodos de observación del registro se procedió a la construcción de las secuencias individuales. Esos estados posibles eran aquellos grados que podrían estarse cursando (de primer a cuarto grado) entre los ciclos educativos iniciados en agosto de 2008 y agosto de 2011, dada la restricción impuesta de que todos los individuos debían estar cursando primer grado de primaria al inicio de la observación. A los estados definidos por los grados iniciados en cada periodo se agregó un quinto estado, que representa la ausencia del registro al inicio del ciclo, y que se interpreta como interrupciones en las trayectorias educativas —este estado fue codificado como NA y se representa en las tablas de las secuencias con un asterisco (*) en el periodo donde no hay registro—. Cabe precisar que en este trabajo se referirá el estado NA como interrupción en la trayectoria educativa sólo por una precisión metodológica, en tanto no se tienen elementos para afirmar una salida de la escuela definitiva. Asimismo, tampoco se tienen elementos para predecir un retorno a la escuela luego de alguna ausencia en el sistema

educativo, salvo en los casos en los que eso ocurre en el periodo de observación de los ciclos educativos considerados en este trabajo, y que quedan contenidos en las trayectorias identificadas. Ante la posibilidad de una ventana de años o periodos de observación mayor, con fines de investigación se podrían establecer criterios para definir el número de periodos fuera del sistema educativo a partir de los cuales podría considerarse una interrupción o ya una salida o abandono definitivo.

Cuadro 8 Las 10 secuencias de modalidades más frecuentes

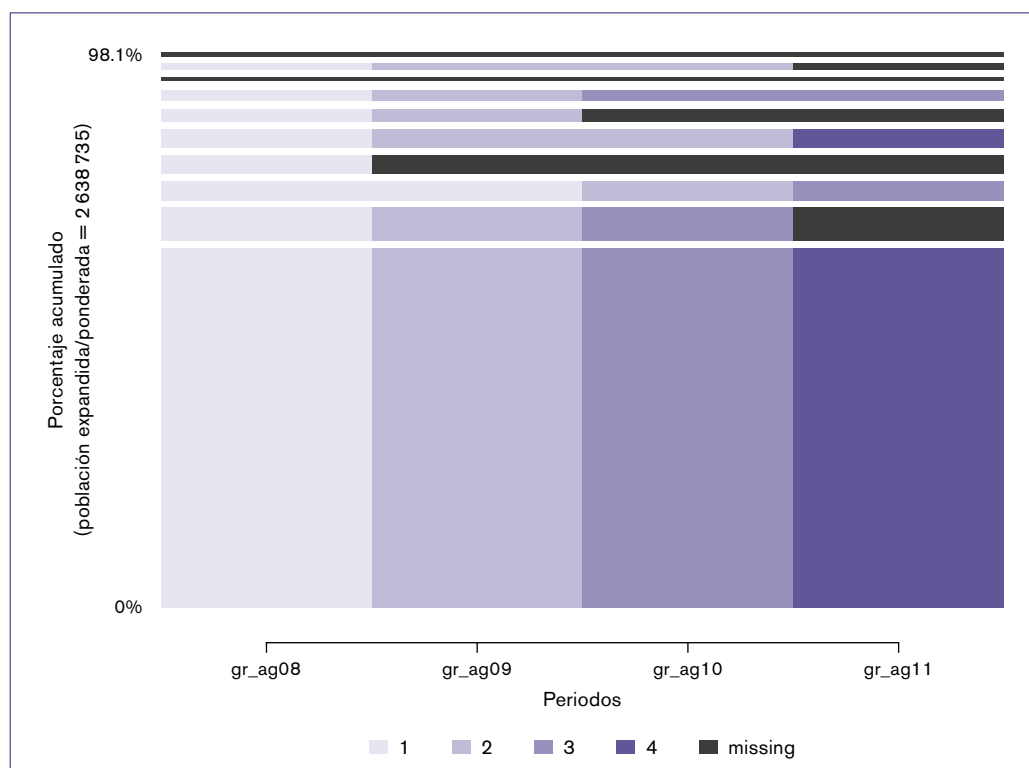
Secuencias	Base expandida		Base sin expandir	
	Frecuencia	%	Frecuencia	%
1-2-3-4	1974781	74.8	3742	74.8
1-2-3-*	169095	6.4	320	6.4
1-1-2-3	96574	3.7	183	3.7
1-**-*	92584	3.5	175	3.5
1-2-2-3	90774	3.4	172	3.4
1-2-**-*	60827	2.3	115	2.3
1-2-3-3	55176	2.1	105	2.1
1-1-2-2	17886	0.7	34	0.7
1-2-2-*	15791	0.6	30	0.6
1-1-1-2	14730	0.6	28	0.6
1-2-2-2	11574	0.4	22	0.4
1-1-2-*	7365	0.3	14	0.3
1-**-2-3	5787	0.2	11	0.2
1-2-**-3	5261	0.2	10	0.2
1-1-**-*	4738	0.2	9	0.2
1-**-*-3	2630	0.1	5	0.1
1-**-1-2	2104	0.1	4	0.1
1-1-1-1	2104	0.1	4	0.1
1-**-2-2	1578	0.1	3	0.1
1-1-1-*	1578	0.1	3	0.1
1-2-**-2	1578	0.1	3	0.1
1-**-1-*	1058	0.0	2	0.0
1-**-2-*	1058	0.0	2	0.0
1-**-*-1	526	0.0	1	0.0
1-**-*-2	526	0.0	1	0.0
1-1-**-1	526	0.0	1	0.0
1-1-**-2	526	0.0	1	0.0
Totales	2638735	100.0	5000	100.0

Fuente: elaboración propia con muestra cohorte primaria.

Como se observa en el cuadro 8, si se siguen las experiencias de la cohorte que está cursando primero de primaria en el ciclo 2008-2009 hasta inicios del ciclo 2011-2012, se obtienen 27 trayectorias posibles con la siguiente distribución de experiencias de la cohorte entre ellas. En este sentido, si se considera que se trabaja con la experiencia de una cohorte completa de inicio de primaria a nivel nacional en México, puede señalarse que no hay una gran variabilidad de experiencias, y que además la gran mayoría —casi las tres cuartas partes— comparte una misma trayectoria educativa en esta etapa de primaria temprana. De los individuos, 74.8% tiene un comportamiento normativo, pues cursa en cada ciclo educativo un nuevo y mayor grado de primaria representado en la tabla por la secuencia 1-2-3-4. De esta manera, se observa que todos los individuos que no tienen esta trayectoria normativa y lineal completan la muestra, es decir, 25% de la cohorte aproximadamente.

De igual manera, como otro indicador de concentración de experiencias y baja heterogeneidad de trayectorias educativas, se identifica que en sólo 10 secuencias (del total de las 27) se concentra la experiencia de 98.1% de la cohorte primaria, tal como se observa en la gráfica 2. Es decir, que las trayectorias de la cuarta parte de la cohorte que no tiene secuencias normativas y lineales (sin interrupciones ni repeticiones de grado) se pueden expresar y resumir a su vez en nueve secuencias diferentes.

Gráfica 2 Las 10 secuencias más comunes en primaria temprana

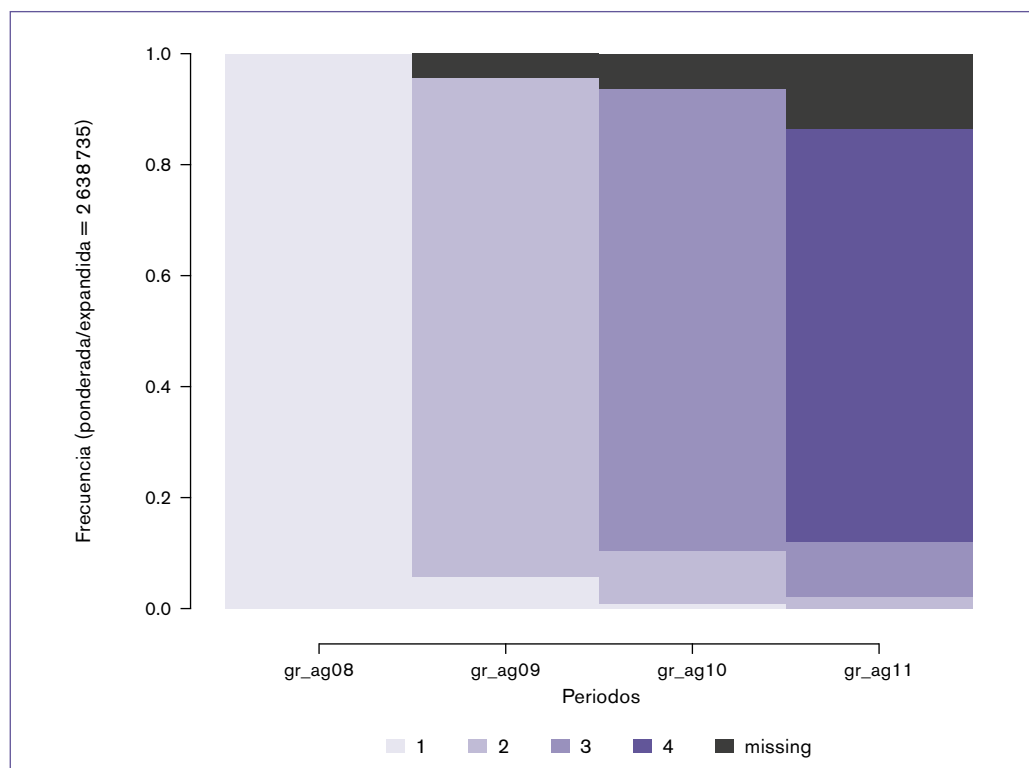


Fuente: elaboración propia con muestra ponderada cohorte primaria.

Tanto en la gráfica 2 como en el cuadro 8 puede identificarse que algunas de estas trayectorias están signadas por repeticiones de grado, otras por interrupciones y otras por ambas situaciones (alguna repetición de grado con posterior interrupción). Incluso hay una trayectoria en particular, con 3.5% de las experiencias de la cohorte, que representa a los individuos inscritos en agosto de 2008 en primer grado de primaria pero ausentes luego del sistema en todos los grados. Por las características del registro no puede decirse si estos individuos aprobaron o no primer grado de primaria, y tampoco si esta interrupción supone una salida definitiva de la escuela, pero indudablemente esa proporción representa en este trabajo las experiencias con mayor tiempo fuera de la escuela: en cuatro ciclos educativos de observación, tres cuartas partes del periodo no aparecen inscritos en grado alguno.

En la gráfica 3 se pueden observar la distribución de estados por periodos de observación, y el aporte de las repeticiones de grado y las interrupciones a la heterogeneidad de estados a medida que se avanza en los ciclos educativos. Cuando se sigue una cohorte en el tiempo, como en este ejercicio, puede señalarse que a medida que se avanza en el tiempo disminuye la proporción de individuos de esa cohorte que ocupa los estados esperados para los años que lleva estudiando. De la misma manera, en el tiempo va aumentando la proporción de individuos que ya no están inscritos (identificada en la gráfica por el color negro de los NA).

Gráfica 3 Histograma de secuencias



Fuente: elaboración propia con muestra ponderada cohorte primaria.

Cuando se observan la distribución de las trayectorias por sexo, los resultados presentan cambios importantes (cuadro 9). Son las mujeres quienes presentan mayor proporción de trayectorias normativas (1-2-3-4), pues sobresalen entre el promedio general de la cohorte hasta con 78%, mientras los varones presentan menor concentración que las mujeres, y el promedio general de la cohorte 71.9%. Asimismo, mujeres y varones presentan diferencias en la distribución y el ordenamiento de los otros tipos de trayectorias educativas.

Cuadro 9 Distribución de trayectorias educativas por sexo, primaria temprana

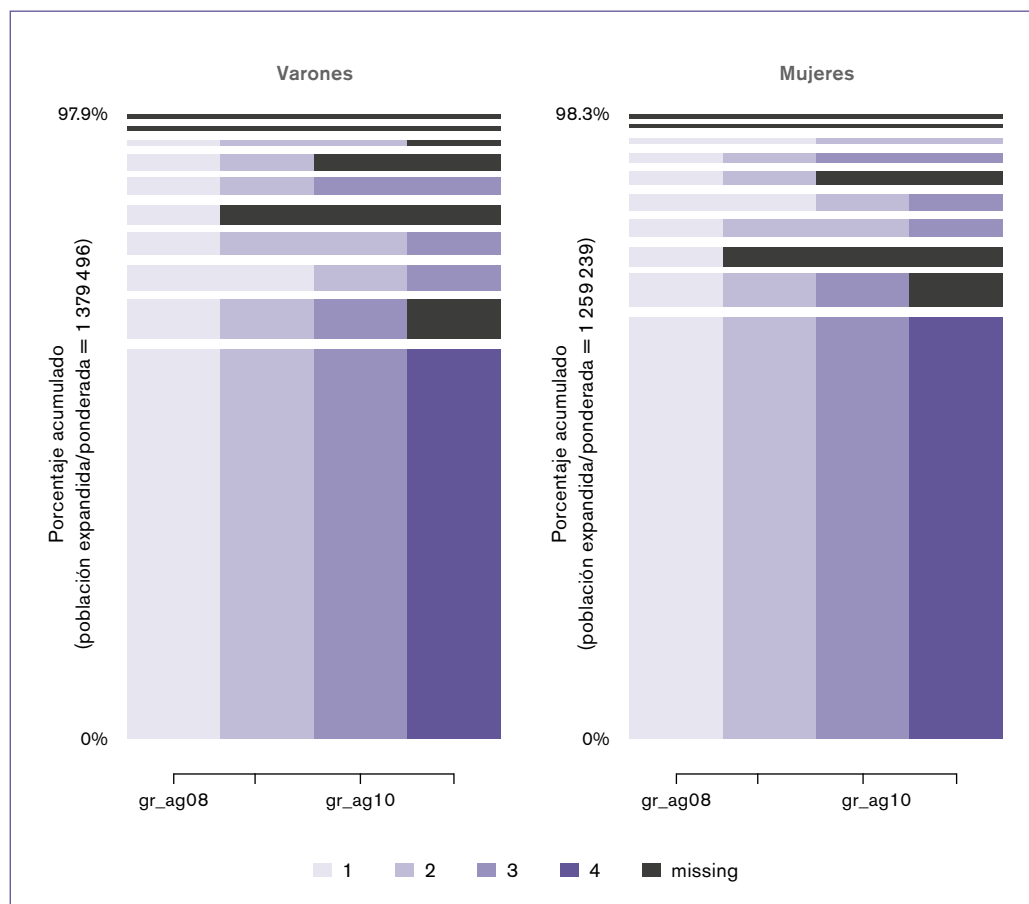
Mujeres			Varones		
Secuencias	Casos	%	Secuencias	Casos	%
1-2-3-4	1862	78.0	1-2-3-4	1880	71.9
1-2-3-*	138	5.8	1-2-3-*	182	7.0
1-**-*	80	3.4	1-1-2-3	121	4.6
1-2-2-3	76	3.2	1-2-2-3	96	3.7
1-1-2-3	62	2.6	1-**-*	95	3.6
1-2-**-*	56	2.3	1-2-3-3	72	2.8
1-2-3-3	33	1.4	1-2-**-*	59	2.3
1-1-2-2	18	0.8	1-2-2-*	19	0.7
1-2-2-*	11	0.5	1-1-1-2	18	0.7
1-1-1-2	10	0.4	1-1-2-2	16	0.6
1-2-2-2	10	0.4	1-2-2-2	12	0.5
1-2-**-3	7	0.3	1-1-2-*	9	0.3
1-1-2-*	5	0.2	1-*-2-3	8	0.3
1-1-**-*	4	0.2	1-1-**-*	5	0.2
1-**-1-2	3	0.1	1-**-3	4	0.2
1-**-2-3	3	0.1	1-1-1-1	3	0.1
1-1-1-*	2	0.1	1-2-**-2	3	0.1
1-**-*-1	1	0.0	1-2-**-3	3	0.1
1-**-*-3	1	0.0	1-**-2-2	2	0.1
1-**-1-*	1	0.0	1-**-*-2	1	0.0
1-**-2-*	1	0.0	1-**-1-*	1	0.0
1-**-2-2	1	0.0	1-**-1-2	1	0.0
1-1-1-1	1	0.0	1-**-2-*	1	0.0
Total	2386	100.0	1-1-**-1	1	0.0
			1-1-**-2	1	0.0
			1-1-1-*	1	0.0
			Total	2614	100.0

Fuente: elaboración propia con muestra cohorte primaria.

A pesar de esta brecha importante entre la concentración de experiencias de mujeres y varones en la trayectoria normativa, puede observarse en la gráfica 4 que los niveles de heterogeneidad de experiencias y trayectorias educativas entre mujeres y varones siguen siendo relativamente bajos. De esta forma, 98.3% y 97.9% de las experiencias de mujeres y varones respectivamente se concentran en sólo 10 tipos de trayectorias.

Mujeres y varones no sólo difieren en las posibilidades de tener trayectorias normativas y relativamente exitosas en la primaria temprana, sino que aquellos que no presentan estos caminos en uno y otro grupo se distribuyen en opciones de trayectos también diferentes entre sí. Existe mayor presencia de varones en trayectorias con repeticiones de grado, pero se profundizará en estos aspectos en análisis posteriores de este informe, mediante indicadores y medidas resumen que permiten hacer visible este fenómeno de manera más adecuada.

Gráfica 4 Las 10 secuencias más frecuentes primaria temprana, por sexo



Fuente: elaboración propia con muestra ponderada cohorte primaria.

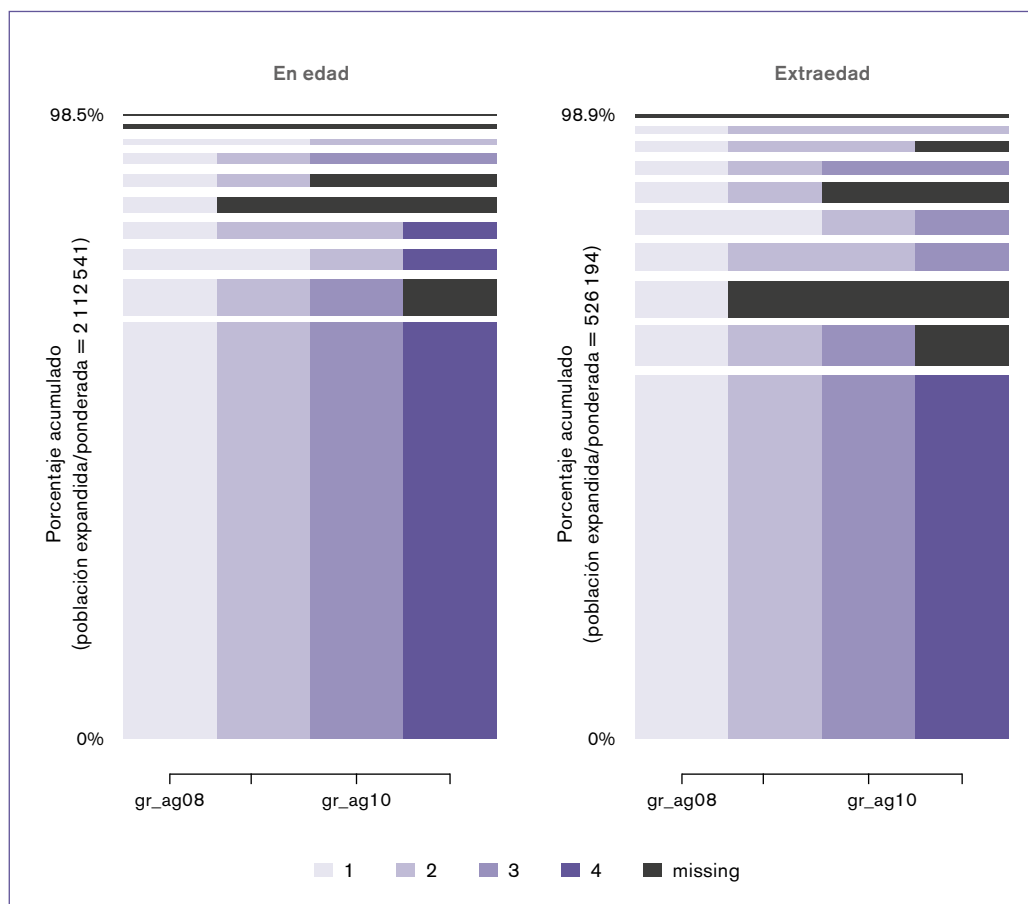
En el cuadro 10 y la gráfica 5 se presenta la distribución de trayectorias por edad al inicio de la primaria. Las trayectorias por edad al inicio de la primaria en agosto de 2008, tal como está construida la variable (en relación con que se esté o no dentro de rango de edad normativa), arroja resultados muy interesantes. En este sentido, aparecen los individuos que tienen extraedad con trayectorias más irregulares, sea por repetición, interrupciones o ambas situaciones.

Cuadro 10 Distribución de trayectorias educativas por edad al inicio de la primaria

En edad normativa al ingreso			Extraedad al ingreso		
Secuencias	Casos	%	Secuencias	Casos	%
1-2-3-4	3081	77.0	1-2-3-4	661	66.3
1-2-3-*	251	6.3	1-2-3-*	69	6.9
1-1-2-3	145	3.6	1-*-*	65	6.5
1-2-2-3	124	3.1	1-2-2-3	48	4.8
1-*-*	110	2.7	1-1-2-3	38	3.8
1-2-*	82	2.0	1-2-*	33	3.3
1-2-3-3	82	2.0	1-2-3-3	23	2.3
1-1-2-2	28	0.7	1-2-2-*	15	1.5
1-1-1-2	23	0.6	1-2-2-2	8	0.8
1-2-2-*	15	0.4	1-1-2-2	6	0.6
1-2-2-2	14	0.4	1-1-1-2	5	0.5
1-1-2-*	10	0.3	1-1-2-*	4	0.4
1-*2-3	8	0.2	1-*2-3	3	0.3
1-2-*3	8	0.2	1-1-1-1	3	0.3
1-1-*	7	0.2	1-*3	2	0.2
1-*3	3	0.1	1*-1-2	2	0.2
1*-1-2	2	0.1	1-1-*	2	0.2
1*-2-2	2	0.1	1-2*-2	2	0.2
1-1-1-*	2	0.1	1-2*-3	2	0.2
1-*3	1	0.0	1-*3	1	0.1
1*-1-*	1	0.0	1*-1-*	1	0.1
1*-2-*	1	0.0	1*-2-*	1	0.1
1-1*-2	1	0.0	1*-2-2	1	0.1
1-1-1-1	1	0.0	1-1*-1	1	0.1
1-2*-2	1	0.0	1-1-1-*	1	0.1
Total	4 003	100.0	Total	997	100.0

Fuente: elaboración propia con muestra cohorte primaria.

Gráfica 5 Las 10 secuencias más frecuentes en primaria temprana, por edad al inicio de la primaria

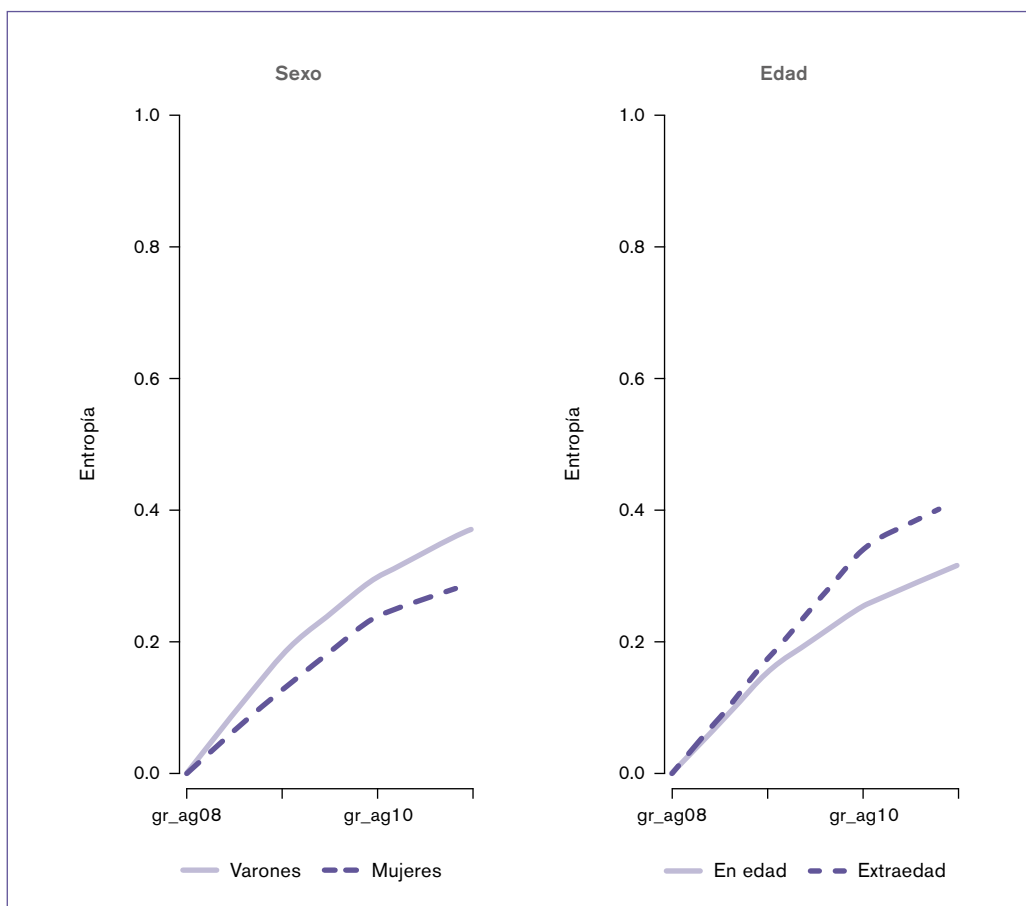


Fuente: elaboración propia con muestra ponderada cohorte primaria.

Cabe señalar que dicha extraedad puede ser el resultado de interrupciones o repeticiones en ciclos anteriores, entre otras posibles causas. De todas maneras, en el momento en que se realizó este estudio los individuos con extraedad en la cohorte de primaria quedan posicionados con aquellos en edad normativa al menos de manera hipotética en un juego de iguales chances al inicio. Sin embargo, se observa que las posibilidades de alcanzar trayectorias regulares son muy diferentes al grupo de edad normativa, solo un 66.3% ante el 77% que presentan en la trayectoria “1-2-3-4” lo de edad normativa al ingreso. Asimismo, los de extraedad presentan una mayor concentración de experiencias en la trayectoria “1-*-*-*”, con un 6.5% de ellos que sólo inician primer grado y salen al menos por tres períodos consecutivos más, que es cuando se tiene posibilidades de observarlos.

Para resumir los comportamientos desiguales entre sexos y rangos de edades al inicio de la primaria la gráfica 6 permite ver los índices de entropía por periodos o ciclos educativos según esas dos covariantes. A medida que se avanza en el tiempo, los varones presentan mayor heterogeneidad de estados en cada ciclo en relación con las mujeres. Algo similar acontece con quienes están en extraedad respecto a los que inician primaria en su edad normativa. Esto refleja la mayor proporción en estas subpoblaciones de repeticiones e interrupciones. Si todos los individuos en cada uno de los ciclos ocuparan los mismos estados, la entropía sería 0, y 1 en una igual distribución entre los 5 estados posibles.

Gráfica 6 Índice de entropía transversal en trayectorias de primaria temprana, por sexo y edad al inicio de la primaria



Fuente: elaboración propia con muestra cohorte primaria.

Es importante destacar que el patrón de la entropía transversal para todos los grupos de población reflejados en la gráfica 6 es ascendente, lo que respalda lo ya identificado en las trayectorias y el histograma propiamente dichos: a medida que se avanza en el tiempo y pasan los ciclos educativos, la heterogeneidad entre estados es mayor, porque conviven los que avanzan de grado conforme pasan los años con aquellos que cursan aún grados anteriores para su cohorte, los rezagados, que son resultado de procesos de repetición de grado e interrupción de sus trayectorias educativas. Esto, aunque en mayor medida en varones y en individuos que inician primaria con extraedad, también es una tendencia general en el patrón de las entropías transversales de mujeres y de quienes inician sus estudios en edad normativa.

Análisis estadístico de trayectorias

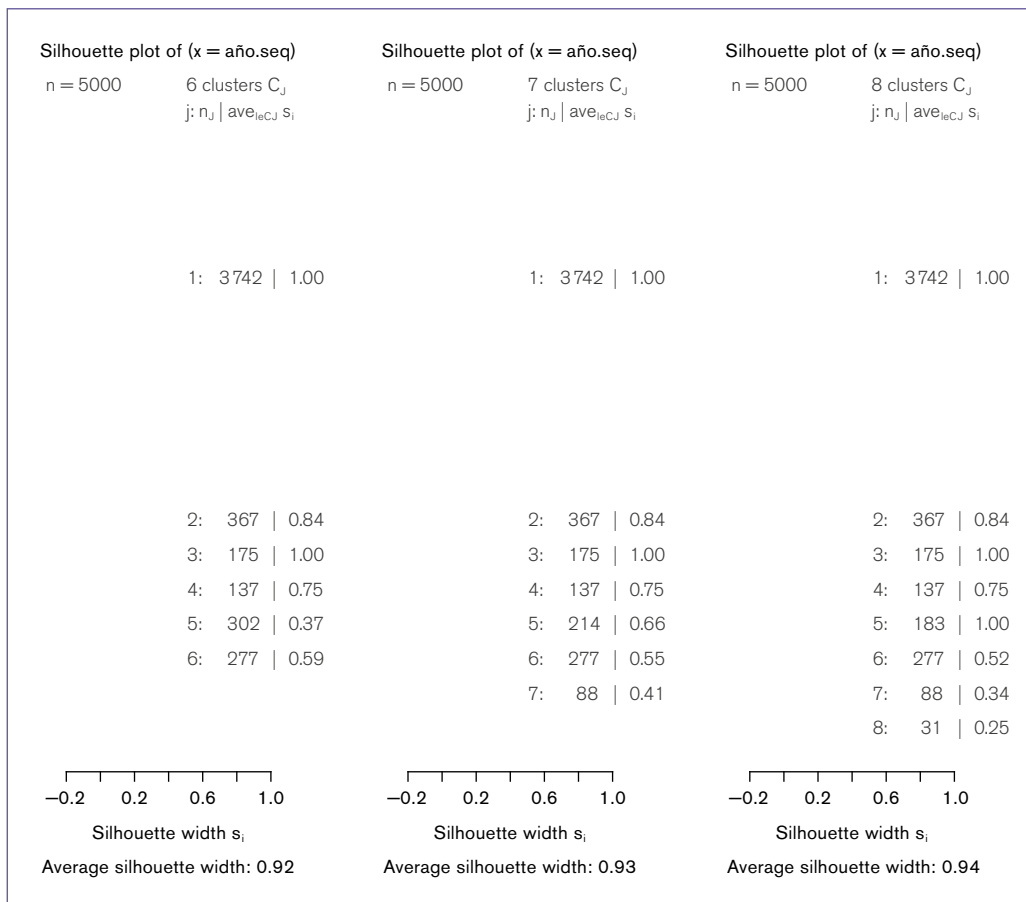
En esta fase del análisis de las secuencias educativas se busca mediante el análisis de agrupamiento de trayectorias comparar y clasificar el conjunto total de las trayectorias para encontrar diferentes patrones generales en las trayectorias educativas de la cohorte y analizar algunas covariantes asociadas a ellas.

Con el objetivo de identificar tipos de trayectorias educativas a partir de las secuencias anteriormente constatadas, se estimó una matriz de distancia por medio de *optimal matching* a partir de una matriz de costos y sustituciones entre estados basada en sus tasas de transición. Esa matriz de distancias se usó como base de un análisis jerárquico de clúster con el criterio *ward*, que permitió identificar ocho grupos de trayectorias. Con el fin de validar el número de agrupaciones, la solución de clúster elegida para distinguir y agrupar a la vez el cuerpo de trayectorias educativas de la cohorte primaria 2008, se utiliza la técnica *silhouette*. En la figura 1 se presentan los resultados del análisis de *silhouette* para las soluciones de clúster de 6, 7 y 8 grupos de trayectorias.

A partir de una solución de clúster de 6 grupos comenzaron a obtenerse valores de *average silhouette width* superiores a 0.9 que se incrementaban en 0.01 a medida que se pasaba de 6 a 7 y de 7 a 8 grupos. Se decidió trabajar con 8 grupos porque esto proporcionaba mayor nivel de homogeneidad interna al grupo de trayectorias número 5 al constituirse un grupo nuevo e independiente del 5, que sería el grupo de trayectorias número 8. La estructura de árbol de los grupos de la solución 8 del clúster puede verse de manera clara en la figura 2.

Este octavo grupo que se obtiene en la solución 8 de clúster refleja la experiencia de muy pocos individuos y no es muy homogéneo en su interior, pero tiene una particularidad que lo diferencia de los otros de manera importante: es el único que guarda las experiencias de trayectorias con interrupciones y regresos. Justamente porque esas interrupciones y regresos no coinciden para todos los casos en el mismo periodo o ciclo educativo, ni se regresa o sale exactamente de idénticos grados, es que presenta niveles de *sil* relativamente cercanos a cero (0.25). Pero es indiscutiblemente un grupo de trayectorias muy particular que interesaba poder clasificar de manera disgregada del resto de las otras experiencias educativas.

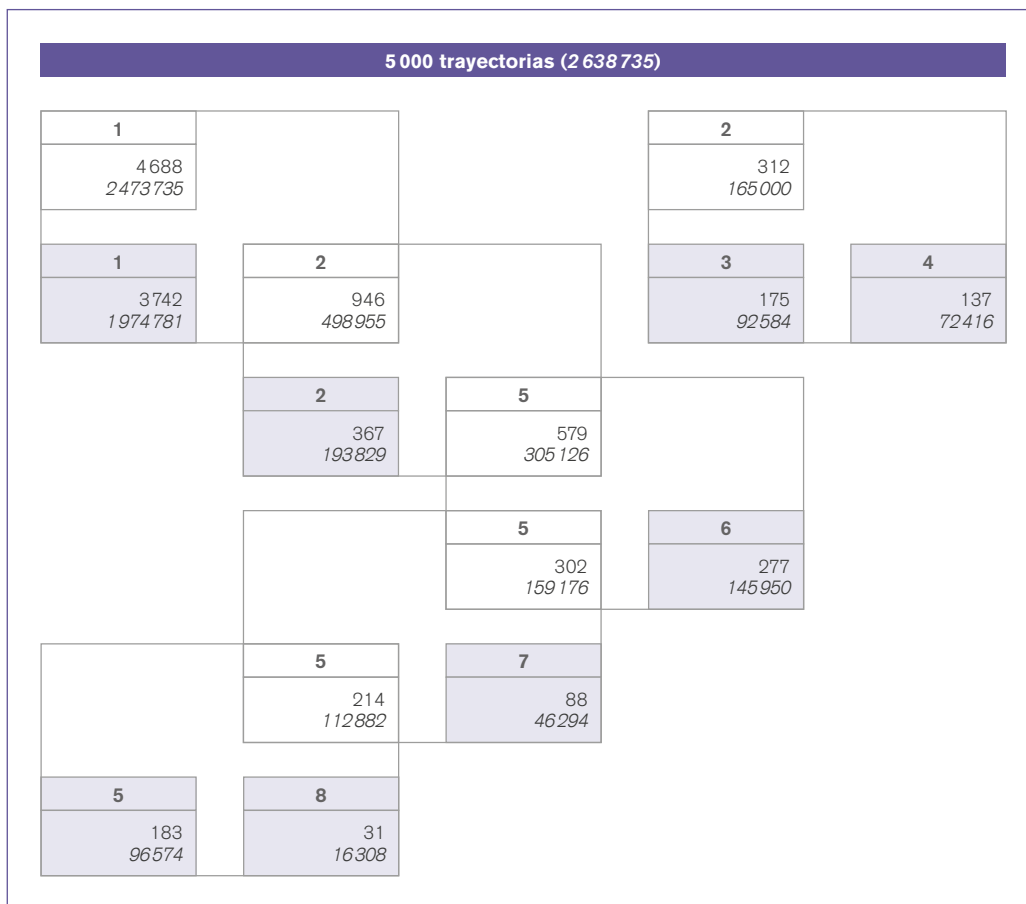
Figura 1 Salidas gráficas del análisis *silhouette* para soluciones de clúster 6, 7 y 8



Fuente: elaboración propia con muestra cohorte primaria.

En el cuadro 11 puede observarse la distribución de las trayectorias educativas ente los ocho grupos identificados mediante el análisis de clúster. Existe un primer grupo de trayectorias, el de mayor frecuencia, que representa casi 75% de las experiencias educativas de la cohorte; ya se distinguía así en las distribuciones de frecuencias el grupo que da cuenta de las trayectorias regulares, lineales y normativas, donde los individuos en cada periodo de observación y ciclo educativo avanza en los grado conforme pasan los años, sin repeticiones ni interrupciones. Los restantes siete grupos dan cuenta de la diversidad de situaciones por las que transita 25% de la cohorte primaria analizada.

Figura 2 Estructura de árbol entre los ocho grupos de trayectorias



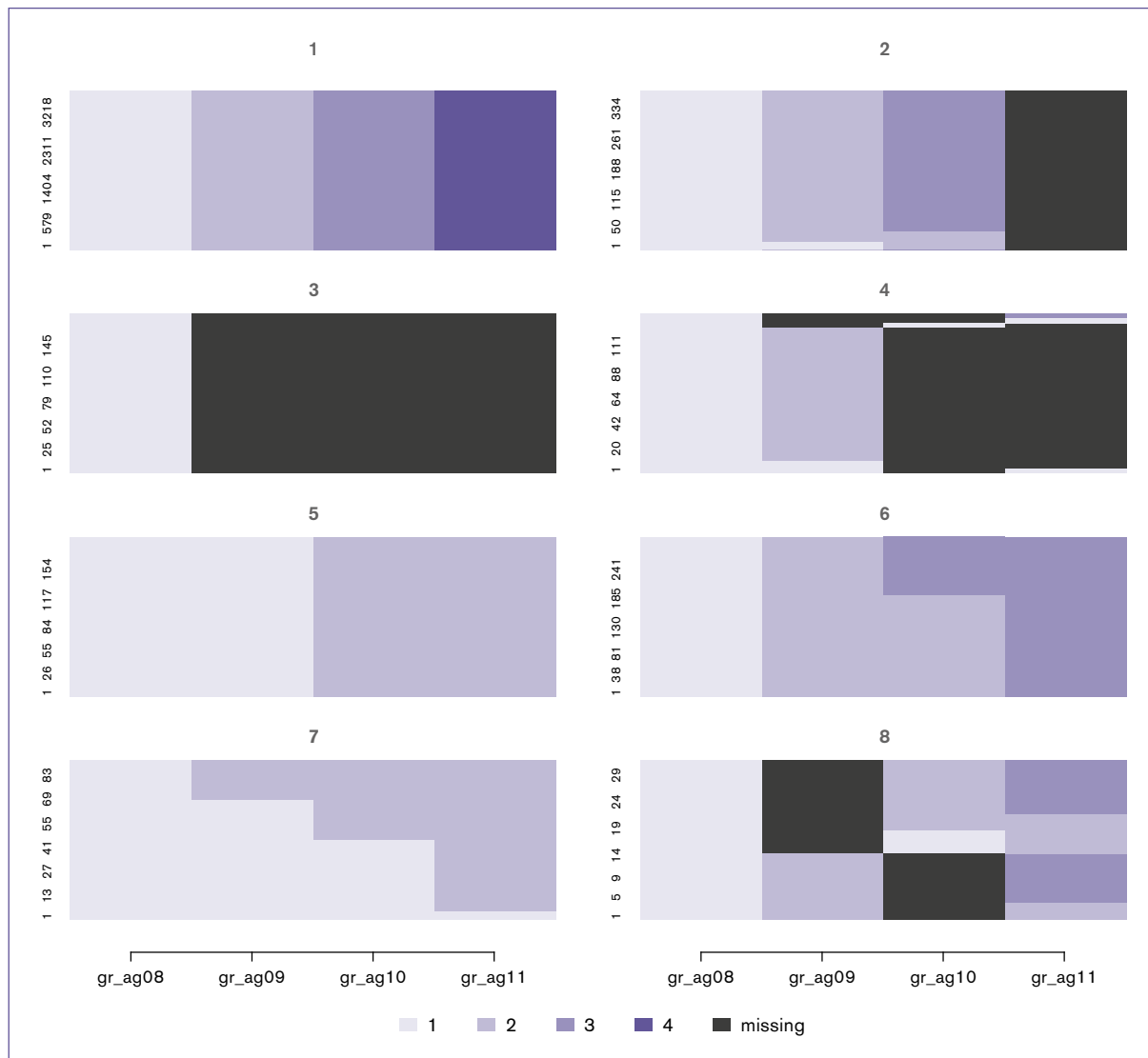
Fuente: elaboración propia con muestra cohorte primaria.

Cuadro 11 Distribución de grupos de trayectorias

Grupos	Casos	%
1	3 742	74.8
2	367	7.3
3	175	3.5
4	137	2.7
5	183	3.7
6	277	5.5
7	88	1.8
8	31	0.6
Total	5 000	100.0

Fuente: elaboración propia con muestra cohorte primaria.

Figura 3 Gráficos con trayectorias individuales por grupos de trayectorias (1 al 8)



Fuente: elaboración propia con muestra cohorte primaria.

La figura 3, que agrupa las gráficas de las trayectorias individuales de cada uno de los 8 grupos o tipologías (que arrojó el análisis de clúster a partir de la matriz de distancia interindividual estimada por *optimal matching*), muestra las particularidades de cada grupo para intentar nombrarlos. Como se observa a continuación:

- Tipología 1: Trayectorias educativas normativas (regulares y lineales).
- Tipología 2: Trayectorias de interrupción a partir de tercer grado.
- Tipología 3: Trayectorias de interrupción a partir de primer grado.
- Tipología 4: Trayectorias de interrupción a partir de segundo.
- Tipología 5: Trayectorias con una repetición de primer grado.
- Tipología 6: Trayectorias con una repetición de segundo o tercer grado.
- Tipología 7: Trayectorias con doble repetición.
- Tipología 8: Trayectorias con interrupciones y retornos.

Se analizará a estos grupos de trayectorias o tipologías a la luz de las trayectorias individuales que se observan en las gráficas de la figura 3 y el cuadro 12, donde se describe la composición de los grupos por medio de tres covariantes: sexo, edad al inicio y modalidad de servicio de la escuela al inicio de la primaria.

Cuadro 12 Composición de los ocho grupos de trayectorias por covariantes (%)

	1	2	3	4	5	6	7	8
Sexo								
Varones	50.2	57.5	54.3	53.3	66.1	60.6	55.7	54.8
Mujeres	49.8	42.5	45.7	46.7	33.9	39.4	44.3	45.2
Edad al inicio de secundaria								
En edad	82.3	75.7	62.9	70.1	79.2	74.4	75.0	67.7
Extraedad	17.7	24.3	37.1	29.9	20.8	25.6	25.0	32.3
Modalidad de servicio al inicio de la secundaria								
General	95.3	92.1	85.7	96.4	92.3	89.5	87.5	100.0
Indígena	4.7	7.9	14.3	3.6	7.7	10.5	12.5	0.0

Fuente: elaboración propia con muestra cohorte primaria.

Tipología 1: trayectorias educativas normativas (74.8%)

Este grupo de trayectorias representa las experiencias regulares y lineales que ya han sido mencionadas como las de mayor presencia en la cohorte analizada de primaria. Asimismo, es el grupo con distribución más igualitaria entre sexos y mayor presencia de individuos en edad normativa al inicio de la primaria en el ciclo 2008-2009.

Tipología 2: trayectorias de interrupción a partir de tercer grado (7.3%)

Este grupo de trayectorias se caracteriza por su interrupción a partir de tercer grado, pues ninguno de los alumnos que lo conforman inicia cuarto grado en el ciclo educativo 2011-2012. Este calendario de interrupción parece ser el elemento clasificador de estas trayectorias, ya que forman parte de este grupo aquellas que, sin ser lineales ni regulares hasta tercer grado,

presentan algunas repeticiones de primer o segundo grado, y todas se interrumpen igualmente en el ciclo que inicia en agosto de 2011. Respecto a su composición, tiene mayor presencia de varones, y casi una cuarta parte de sus miembros tenían extraedad al inicio de la primaria.

Tipología 3: trayectorias de interrupción a partir de primer grado (3.5%)

El grupo 3 de trayectorias concentra y distingue de todos los otros grupos o tipologías a aquellas secuencias que, si bien inician en primer grado de primaria con la cohorte analizada, ya no tienen más participación educativa durante el resto del periodo observado. Este grupo es el de mayor participación de individuos con extraedad al inicio de la primaria, y alcanza 37.1% de los casos. De la misma manera, tiene más varones (54.3%) que mujeres. Este grupo de trayectorias es el que tiene la mayor presencia de escuelas de modalidad indígena; 14.3% de estas experiencias son individuos que estudiaron primer grado de primaria en dicha modalidad.

Tipología 4: trayectorias de interrupción a partir de segundo grado (2.7%)

Este grupo de trayectorias da cuenta de aquellas experiencias educativas de los miembros de la cohorte analizada que pasan la mitad del periodo de observación sin registro, es decir con interrupciones. En la otra mitad del tiempo observado hay un avance de primero a segundo grado y algunos casos de repetición de primer grado y posteriormente interrupción. Asimismo, existe un número pequeño de secuencias en este grupo que muestra ese patrón de interrupción por dos años, pero entre dos participaciones, es decir, cursan un año, interrumpen dos, y vuelven a participar en el último ciclo educativo analizado.

Tipología 5: trayectorias con una repetición de primer grado (3.7%)

Este grupo de trayectorias presenta una homogeneidad total y da cuenta de aquellas secuencias educativas donde se repite una vez primer grado y hay un avance posterior a segundo. En este sentido, dado nuestro periodo de observación, dejan de observarse todos en tercer grado de primaria. En la conformación de grupos mediante el análisis de clúster este grupo integraba a los otros tres que se analizarán a continuación, y que guardan en común episodios de repetición de grados. Este grupo presenta mayor proporción de varones (66.1%).

Tipología 6: trayectorias con una repetición de segundo o tercer grado (5.5%)

En este grupo de trayectorias todos los individuos superan el primer grado y, más adelante, repiten una vez segundo o tercer grado. De esas trayectorias, 60.6% son varones, y 10% pertenecen a escuelas de modalidad indígena.

Tipología 7: trayectorias con doble repetición (1.8%)

En este grupo de trayectorias se concentran las experiencias educativas de dobles procesos de repetición en las etapas tempranas de la escuela primaria. La doble repetición se produce tanto por repetir dos veces un mismo grado (cursarlo tres periodos) o repetir una vez un

grado (dos periodos) y otra vez el siguiente (por dos periodos). Este patrón en las trayectorias conlleva a que ninguno de sus miembros curse tercer grado en el momento de inicio del ciclo en 2011, cuando se detiene la observación en este trabajo. Dentro de la poca participación de escuelas con modalidad indígena, este grupo es uno de los que concentra mayor participación relativa de trayectorias que acontecen en escuelas con esta modalidad (12,5%).

Tipología 8: Trayectorias con interrupciones y retornos (0.6%)

Este último grupo de trayectorias es muy pequeño, pero no por eso residual, sino que tiene una característica única respecto a todas las otras trayectorias de los demás grupos y tipologías: presenta una interrupción educativa durante un solo ciclo. Puede observarse que en muchos casos esta interrupción no estuvo ligada a procesos de repetición, ya que los alumnos se incorporan en un grado mayor al que tenían en dos periodos anteriores. Más de la tercera parte de estas experiencias (32.3%) tienen con extraedad al inicio de la primaria.

Medidas resumen de las trayectorias educativas

Cuando se construyeron algunas medidas resumen a partir de las trayectorias educativas identificadas en la cohorte de primaria 2008-2009, se buscó alcanzar una medida proxy que diera cuenta de los procesos de repetición de grado. Un número importante de trayectorias describían este fenómeno, y además se diferenciaban entre sí por los grados repetidos y el número de veces de las repeticiones que podía existir en una trayectoria individual. Así, se procuró estimar la distribución de duraciones por grado educativo para identificar estos casos de manera más directa y analizarlos por covariantes.

Es importante señalar que el escaso rango de periodos observados impuso una restricción metodológica al análisis de los resultados, sobre todo a los últimos dos grados (tercero y cuarto), porque les otorga menos años del calendario para observar repeticiones respecto a los dos anteriores. En este sentido, no deben compararse distribuciones de duraciones entre los distintos grados, lo cual llevaría a la conclusión de que es más probable repetir primero que cuarto, en tanto no se otorgue igual tiempo de observación a la cohorte para ver, por ejemplo, si hay más o menos repeticiones en cuarto que primer grado.

De todas maneras, en estos grados (tercer y cuarto) es interesante ver las duraciones 0, porque dan cuenta de aquellos individuos que, por repeticiones o interrupciones, nunca llegaron a cursarlos en el periodo de observación de este trabajo (cuatro ciclos educativos).

En el cuadro 13 puede observarse la distribución de duraciones por grado para las subpoblaciones por sexo, edad al inicio de la primaria y modalidad al inicio de la primaria. Debido a que se observa un conjunto de fenómenos desde distintos ángulos e intentos de operacionalización, se ven las coincidencias de algunas duraciones con las frecuencias de grupos de trayectorias que clasifican y dan cuenta de las distintas particularidades. En resumen, estas distribuciones de duraciones por grados educativos constituyen otra vía para el estudio

Cuadro 13 Duraciones en cada grado educativo (de 1° a 4°)

Duraciones	Grado de primaria							
	Primero		Segundo		Tercero		Cuarto	
Sexo	Varones	Mujeres	Varones	Mujeres	Varones	Mujeres	Varones	Mujeres
0	0.0	0.0	4.2	3.8	9.5	8.5	28.1	22.0
1	93.2	95.5	90.1	91.4	87.8	90.1	71.9	78.0
2	5.9	3.9	5.2	4.4	2.8	1.4	0.0	0.0
3	0.8	0.5	0.5	0.4	0.0	0	0.0	0.0
Totales	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Edad al inicio	En edad	Extraedad	En edad	Extraedad	En edad	Extraedad	En edad	Extraedad
0	0.0	0.0	3.1	7.5	7.5	15.1	23.0	33.7
1	94.5	93.7	92.3	84.5	90.4	82.5	77.0	66.3
2	4.9	5.3	4.2	7.2	2.0	2.3	0.0	0.0
3	0.6	0.7	0.3	0.8	0.0	0.0	0.0	0.0
Totales	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Modalidad al inicio	General	Indígena	General	Indígena	General	Indígena	General	Indígena
0	0.0	0.0	3.7	9.7	8.5	17.6	24.3	39.0
1	94.5	91.4	91.4	79.0	89.5	78.3	75.7	61.0
2	4.8	7.6	4.5	10.0	2.0	4.1	0.0	0.0
3	0.6	0.7	0.4	1.4	0.0	0.0	0.0	0.0
Totales	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Fuente: elaboración propia con muestra cohorte primaria.

de fenómenos como la repetición de grados, los procesos de rezago y la disminución de la cohorte a medida que se avanza en el tiempo, y para el proceso de análisis de dichos fenómenos por sexo y otras covariantes.

Respecto a las duraciones en cada grado por sexo, interesa destacar que siempre son mayores los niveles de repeticiones (duraciones 2 y 3) en varones que en mujeres. Asimismo, como ya se veía en la distribución por sexo de trayectoria de tipo 1, normativa, regular y lineal (1-2-3-4), hay más mujeres que varones, y siempre la duración 1 para los cuatro grados analizados es menor en varones.

Algo similar sucede con la edad, sólo que quienes tienen extraedad al inicio son aquellos que presentan mayores niveles de duraciones 2 y 3, y menos de duraciones 1 en los distintos grados.

Respecto a la distribución de duraciones por modalidad de la escuela en la que se inicia primaria, la modalidad indígena es la que presenta mayores niveles en la duración 2 y 3, e incluso duplica las proporciones de la modalidad general en segundo y tercer grado. Asimismo, los individuos de modalidad indígena son quienes presentan proporciones menores en la duración 1 en todos los grados, lo que evidencia mayor nivel de rezago, interrupciones o ambos procesos combinados. Mientras las tres cuartas partes de los individuos de la cohorte analizada que inician en escuelas generales alcanzan cuarto grado en agosto de 2011, sólo lo hace 61% de quienes asisten a escuelas indígenas. Además, 20% de quienes estudian en esa modalidad en la cohorte analizada sólo estudia hasta segundo grado de primaria.

Otra manera de agrupar información de las trayectorias identificadas en la cohorte de primaria es mediante el número de transiciones efectivamente realizadas por la cohorte analizada en ese periodo de tiempo en particular. Con *transiciones* se alude a los cambios de estado — de primero a segundo, de segundo a tercero y de tercero a cuarto—, y el término se usa con el fin de definir una medida aproximada para verificar niveles de promoción de grados en la población estudiada en el periodo planteado.

En el análisis anteriormente realizado a partir de las tablas de distribución de la totalidad de trayectorias identificadas para el total de la población y ordenada por subgrupos de interés (covariantes), en los cuadros 1, 2 y 3 pudo observarse que las posibilidades de alcanzar en el tiempo analizado la realización de todas las transiciones entre estados posibles eran muy desiguales entre sexos y edades al inicio de la primaria. En el cuadro 14 se presenta la distribución por sexo, edad y modalidad de servicio escolar al inicio de primaria de las transiciones realizadas; como puede observarse, coinciden la proporción final de cada subpoblación que alcanza a realizar las tres transiciones posibles en el tiempo trabajado y la trayectoria normativa y regular que daba cuenta de las tres transiciones (codificada como 1-2-3-4).

Sin embargo, existe otro tipo de situaciones donde hay sólo una o dos transiciones que se distribuyen entre diferentes tipos de trayectorias: cuando se repiten grados, cuando se interrumpen o cuando se experimentan ambos fenómenos. De esta forma, un análisis centrado en transiciones y sus distribuciones proporciona información más apropiada para dar cuenta de procesos de avance entre grados en una cohorte.

Respecto a las transiciones realizadas por la cohorte en el periodo analizado, es importante señalar que en todos los grupos, incluso en aquellos con menores niveles de avance, siempre es relativamente baja la proporción de los individuos que no realizan ninguna transición o sólo realizan una. En este sentido, el segundo lugar en prevalencias lo ocupa la transición 2 para todos los subgrupos poblacionales analizados. Sin embargo, las brechas por covariantes son más grandes cuando se considera la edad de inicio y la modalidad de servicio escolar al inicio de la primaria, respecto al sexo.

Cuadro 14 Distribución por covariantes de número de transiciones realizadas (%)

Sexo	Varones	Mujeres
0	4.1	3.7
1	5.6	4.9
2	18.4	13.4
3	71.9	78.0
Total	100.0	100.0

Edad al inicio	En edad	Extraedad
0	3.0	7.3
1	4.5	8.0
2	15.4	18.4
3	77.0	66.3
Total	100.0	100.0

Modalidad servicio al inicio	General	Indígena
0	3.5	9.7
1	5.1	7.9
2	15.7	21.4
3	75.7	61.0
Total	100.0	100.0

Fuente: elaboración propia con muestra cohorte primaria.

El análisis de duraciones y transiciones puede ser complementado con la estimación de tasas de transición entre estados. Dichas tasas se presentan como matriz de transiciones entre estados en el cuadro 15; para la población total y por sexo. Igual que lo que sucedía con el análisis de duraciones, el truncamiento en esta investigación debido a un periodo de observación estrecho, que no alcanza a seguir a la cohorte primaria en todo su ciclo, impone restricciones analíticas. Si bien se presentan las tasas de transición entre todos los estados, se debe tener cuidado a la hora de interpretarlas. En este sentido, cuarto grado de primaria en esta matriz es un punto de llegada pero no de partida, porque la observación no continúa. Cabe señalar que se incorporan como un estado posible los casos faltantes, es decir, aquellos que no estaban registrados en algunos periodos de observación, y se indican en la matriz con asterisco (*). Su análisis es interesante como estado de destino y como estado de partida. Permiten identificar con cierta precisión el peso de la tendencia a abandonar o interrumpir los estudios a partir de ciertos grados anteriores, así como a retornar a otros grados tras alguna interrupción.

De manera general, la matriz de transición presenta algunas celdas o casillas de gran relevancia. Una de ellas es la diagonal señalada con color amarillo en el cuadro 15. Permite ver la proporción de individuos que partiendo de un grado escolar alcanzan el siguiente (una medida de aproximación a la promoción entre grados de la cohorte). Esta diagonal también permite observar el proceso de pérdida y reducción que suponen para la cohorte analizada las repeticiones de grados y las salidas e interrupciones en cada ciclo educativo consecutivo. En el caso de la población total, se observa una reducción aproximada de 10% en los alumnos de cada ciclo educativo. Esa reducción se distribuye entre dos situaciones: la repetición y la interrupción o el abandono (que forman parte del estado señalado con el asterisco).

Así, las celdas resaltadas con color rosa permiten ver las tasas de retención por grado educativo. En este sentido, primer y segundo grado presentan tasas mayores que tercer grado; son próximas a 6%, frente a 2.5% de retención en tercer grado de primaria.

Cuadro 15 Matriz de transiciones entre estados para población total, varones y mujeres

Total	[1->]	[2->]	[3->]	[4->]	[*->]
[1->]	0.059	0.899	0.000	0.000	0.041
[2->]	0.000	0.057	0.908	0.000	0.035
[3->]	0.000	0.000	0.025	0.898	0.077
[4->]	0.000	0.000	0.000	0.000	0.000
[*->]	0.015	0.040	0.028	0.000	0.917
Varones	[1->]	[2->]	[3->]	[4->]	[*->]
[1->]	0.071	0.886			0.043
[2->]		0.060	0.904		0.036
[3->]			0.034	0.881	0.085
[4->]					
[*->]	0.010	0.056	0.024		0.909
Mujeres	[1->]	[2->]	[3->]	[4->]	[*->]
[1->]	0.046	0.915			0.039
[2->]		0.053	0.913		0.034
[3->]			0.016	0.916	0.068
[4->]					
[*->]	0.021	0.021	0.033		0.925

Fuente: elaboración propia con muestra ponderada de cohorte primaria.

Las celdas resaltadas con morado indican los distintos niveles de atracción que tiene la interrupción desde los distintos grados de la primaria temprana. Si bien nunca alcanza 10%, mientras desde primer y segundo grado está próximo a 4%, a partir de tercer grado se aproxima a 8%. Esto da lugar a conformar un tipo de trayectorias clasificadas como el grupo 2.

Si se observan estos comportamientos en las matrices de transición por sexo, se pueden constatar los patrones diferenciales entre varones y mujeres que se han resaltado a lo largo de todo el análisis ya realizado para la cohorte de primaria 2008-2009. La proporción de los varones es siempre mayor a las de mujeres en las tasas de retención en el mismo grado (celdas rosas); pero la brecha es mucho mayor en el caso de tercer grado (la tasa se duplica). Asimismo, siempre son menores a 9% los niveles de pérdida de mujeres a medida que se avanza entre los ciclos educativos, mientras las tasas son superiores a 10% en el caso de los varones (celdas amarillas).

7.2 Secundaria

Composición de la cohorte de ingreso a secundaria

La base de datos que se usó para el análisis de las trayectorias educativas de quienes ingresaron a secundaria en agosto de 2008 es una muestra de 5 000 casos que contiene todas las variables de la base de datos original del RNA para cada periodo de observación, con la cuales se construyeron los estados (grado cursado en el periodo) y las variables que se usan como covariantes en el análisis. Se eliminó un caso duplicado de la base muestra, lo cual se refleja en que en el total de todos los cuadros y gráficos se presentan 4 999 casos.

La composición por edad de la cohorte de secundaria puede observarse en el cuadro 16, donde se muestran algunas medidas resumen de la fecha de nacimiento de la cohorte. Aquellos que están en edad normativa para cursar primer grado de secundaria en agosto de 2008 son los nacidos entre el 1 de enero de 1996 y el 31 de diciembre de 1996. A partir de estas fechas de nacimiento se observa que 0.3% de los individuos de la cohorte son menores a la edad normativa al inicio de la secundaria, 58.6% de ellos está en edad normativa y 41.1% presenta extraedad al inicio de la secundaria. Los que presentan extraedad pueden estar ubicados hasta con 15 años aproximadamente por encima de la edad normativa, mientras que los que presentan edades menores son un número reducido y están por debajo del rango normativo por apenas unos meses.

Respecto a la composición por sexo de 4 999 casos, 2 541 son varones (50.8%) y 2 458 son mujeres (49.2%). En el cuadro 17 puede observarse la composición conjunta de sexo y edad, donde la relación entre los que están en edad normativa y los que están en extraedad se modifica entre sexos. Los varones alcanzan 43.8% en extraedad, mientras las mujeres, 38.4%, lo que da cuenta de trayectorias educativas anteriores a la analizada menos regulares y lineales en los varones que en el caso de las mujeres. De todas maneras, la extraedad aparece como un fenómeno con presencia en ambos sexos.

Cuadro 16 Medidas resumen de fecha de nacimiento

	Fechas de nacimiento
Edad mínima	27/06/81
Primer cuartil	22/09/95
Media	11/02/96
Mediana	08/12/95
Tercer cuartil	24/06/96
Máximo	02/10/97

Fuente: elaboración propia con muestra cohorte secundaria.

Cuadro 17 Composición por sexo y edad en el momento de su ingreso a secundaria

Edad	Sexo		Total
	Varones	Mujeres	
Menor	0.4	0.3	0.3
En edad	55.9	61.4	58.6
Extraedad	43.8	38.4	41.1
Total	100.0	100.0	100.0

Fuente: elaboración propia con muestra cohorte secundaria.

Respecto a la modalidad de servicio de la escuela en la que inician secundaria en agosto de 2008, interesa evaluar la movilidad entre modalidades de servicio en las distintas escuelas que cursan los distintos grados y, a su vez, el pasaje de nivel de escuela primaria a modalidad de secundaria.

Las modalidades de servicio escolar que aparecen en la base completa de la cohorte secundaria analizada y en la base muestra son cuatro: general, técnica, telesecundaria y trabajo. En el cuadro 18 se presenta la distribución en la muestra de cohorte secundaria. La categoría NA da cuenta de la ausencia de registro en ese periodo.

En el cuadro 18 puede observarse que la modalidad general es la más frecuente en el inicio de la secundaria: casi la mitad de los individuos de la cohorte asiste a este tipo de escuelas. Una cuarta parte está inscrita en escuelas técnicas; casi 16% en telesecundarias, y una pequeña proporción (0.4%) en la modalidad de trabajo.

En el caso de la cohorte secundaria, que inicia en agosto de 2008, se incluyó la información del ciclo anterior (2007-2008) para ver la modalidad de servicio de la escuela primaria de la que los individuos proceden. En este sentido, puede hacerse un análisis de la movilidad de

Cuadro 18 Distribución de modalidades de servicio al inicio de la secundaria

Modalidad	Casos	%
General	2289	45.8
Técnica	1234	24.7
Telesecundaria	788	15.8
Trabajo	19	0.4
NA	669	13.4
Totales	4999	100.0

Fuente: elaboración propia con muestra cohorte secundaria.

modalidades de servicio escolar entre primaria y secundaria, y luego ver la movilidad entre modalidades de escuelas secundarias. Este último análisis es determinante para validar la modalidad de servicio escolar al inicio de la secundaria como una variable fija en el tiempo y covariante en el análisis de las trayectorias educativas realizado en este proyecto.

Para analizar y visibilizar los cambios de modalidad de servicio, se codificaron para cada uno de los cuatro periodos observados —ciclos educativos que iniciaron en agosto de 2008 y agosto de 2011— la modalidad de la escuela en la que se cursaba el último grado de primaria y la de los tres ciclos en secundaria. Obviamente, cuando el individuo no estaba registrado en algún periodo de observación, no hay modalidad de servicio alguna, y se emplea el código NA. La modalidad general se codifica como GEN, y la modalidad indígena como IND; además, la modalidad técnica como TEC, la modalidad telesecundaria como TEL y la modalidad trabajadores como TRA. De esta forma, se construyó una secuencia de modalidades para ver el pasaje de primaria a secundaria entre un primer y un segundo periodo de observación, y luego entre los tres periodos finales en secundaria.

En el cuadro 19 se pueden observar las 10 secuencias de modalidades más frecuentes para la cohorte de secundaria que inicia en agosto de 2008; el primer registro señala la modalidad en la que se cursó sexto grado de primaria.

Las 10 secuencias de modalidades más frecuentes resumen la experiencia de 87.7% de la cohorte secundaria analizada. En la gráfica 7 puede observarse la totalidad de las secuencias individuales de modalidades de servicio para los cuatros periodos observados (desde agosto de 2007 hasta agosto de 2011).

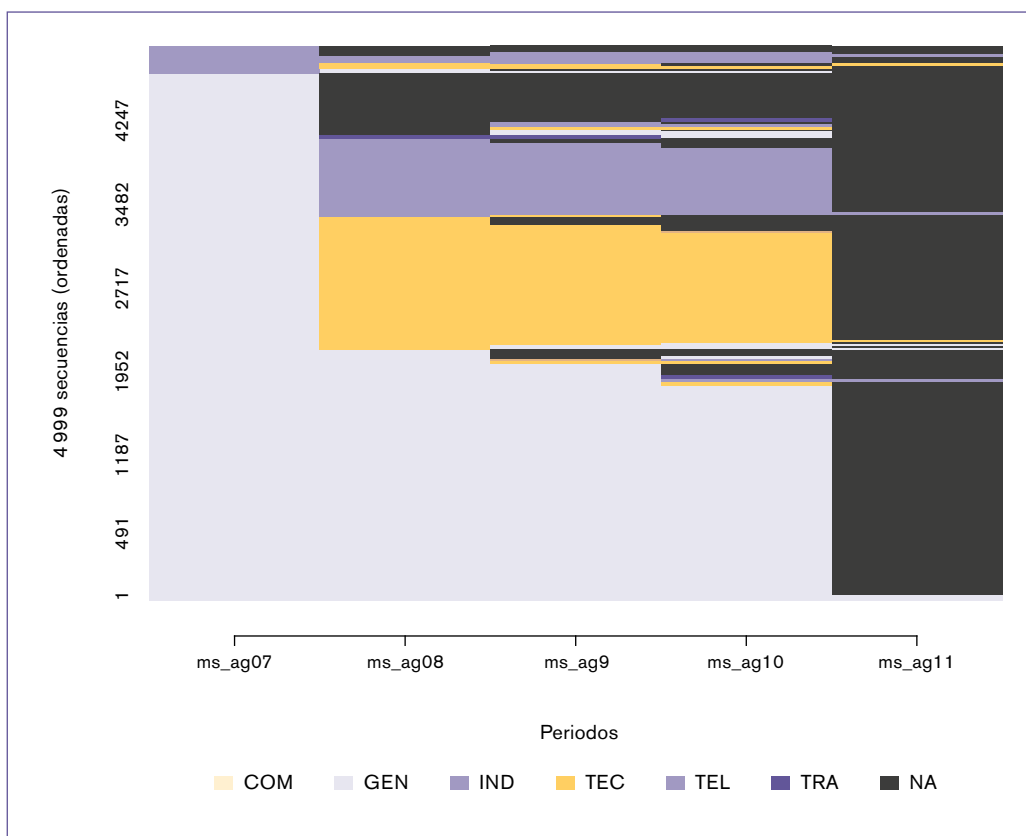
En las secuencias de modalidades de servicio puede observarse que la gran mayoría de aquellos que cursan sexto de primaria en escuela de modalidad general continúan en secundarias con la misma modalidad durante todos los periodos, lo cual ocurre en casi 38% de los casos. Existe un grupo de individuos que cursan sexto de primaria en escuelas de modalidad general y pasan a otro tipo de modalidades de servicio en la secundaria, como técnica y telesecundaria.

Cuadro 19 Las 10 secuencias de modalidades de servicio más frecuentes

Secuencia	Casos	%
GEN-GEN-GEN-GEN	1893	37.9
GEN-TEC-TEC-TEC	962	19.2
GEN-TEL-TEL-TEL	607	12.1
GEN-NA-NA-NA	419	8.4
GEN-GEN-GEN-NA	168	3.4
GEN-TEC-TEC-NA	86	1.7
GEN-GEN-NA-NA	73	1.5
IND-TEL-TEL-TEL	72	1.4
IND-NA-NA-NA	56	1.1
GEN-NA-TEL-TEL	49	1.0
Totales	4 385	87.7

Fuente: elaboración propia con muestra cohorte secundaria.

Gráfica 7 Distribución de secuencias individuales de modalidad de servicio de escuelas



Fuente: elaboración propia con muestra cohorte secundaria.

En el cuadro 20 se presenta la distribución cruzada entre modalidades de servicio de la escuela primaria donde se cursó sexto grado y la secundaria donde se cursó primer grado. Como puede verse, existe una pequeña proporción de individuos que proceden de escuelas primarias de modalidad de servicio indígena, y que pasan en su gran mayoría a telesecundarias, y en menor medida a técnicas y generales.

Cuadro 20 Distribución de frecuencia entre modalidades de servicio en sexto de primaria y primero de secundaria (%)

Modalidad servicio 6° primaria	Modalidad servicio inicio de secundaria					Total
	General	Técnica	Telesecundaria	Trabajo	NA	
Primaria comunitaria	0.0	0.0	100.0	0.0	0.0	100.0
Primaria general	47.7	25.0	14.8	0.4	12.1	100.0
Primaria indígena	12.5	18.3	31.9	0.0	37.3	100.0

Fuente: elaboración propia con muestra cohorte secundaria.

En el cuadro 21 se presenta la distribución cruzada de modalidades de servicio en primer grado de secundaria y segundo grado de secundaria. Puede concluirse que la migración entre modalidades de servicio en los grados de secundaria es muy baja, por lo cual se torna viable mantener como covariante para las secuencias la modalidad al inicio de la secundaria, es decir, la modalidad de servicio de la escuela donde se cursó primer grado de secundaria.

Cuadro 21 Distribución de frecuencia entre modalidades de servicio en primero y segundo grado de secundaria (%)

Modalidad servicio 1° secundaria	Modalidad servicio 2° secundaria					Total
	General	Técnica	Telesecundaria	Trabajo	NA	
General	94.10	1.00	0.80	0.00	4.10	100.00
Técnica	3.40	91.20	0.70	0.10	4.60	100.00
Telesecundaria	0.90	0.90	93.50	0.00	4.70	100.00
Trabajo	5.30	0.00	0.00	78.90	15.80	100.00

Fuente: elaboración propia con muestra cohorte secundaria.

Las trayectorias educativas

Se definieron los estados posibles que los miembros de la cohorte secundaria 2008-2009 podrían adquirir en el periodo de observación, y se construyeron las secuencias educativas para el periodo de observación. Como se inicia la observación en el estado del ciclo anterior al inicio de secundaria en agosto de 2008, se parte del estado posible en agosto de 2007. Es decir, sexto grado de primaria (codificado como 6), y luego, primer grado de secundaria (codificado como 7), segundo grado de secundaria (codificado como 8) y tercer grado de secundaria (codificado como 9). A estos cuatro estados posibles (6, 7, 8 y 9) se agregó uno quinto que indica la ausencia en los registros en alguno de los periodos de observación; se codifica como estado NA y en las tablas se representa con un asterisco (*).

En el cuadro 22 se presenta la totalidad de trayectorias educativas de la cohorte que inicia secundaria en agosto de 2008 —39—, así como la distribución en la base expandida y sin expandir.

Como se observa en el cuadro, existe una secuencia que refleja la trayectoria normativa y regular de la cohorte analizada; en ella se muestra en cada ciclo educativo el curso del grado esperado (6-7-8-9-*), y concentra 76% de las experiencias. Igual que lo identificado en la cohorte de primaria, en secundaria también existe una cuarta parte de la cohorte, aproximadamente, con trayectorias que pueden entenderse como no normativas o irregulares.

El número total de trayectorias identificadas en la cohorte de secundaria es mayor que en la de primaria, pero de todas formas, un pequeño grupo concentra la experiencia de 97.7% de la cohorte, tal como se ilustra en la gráfica 8.

Las nueve trayectorias que siguen a la primera, la cual representa a la normativa y regular, dan cuenta de 21.7% de la cohorte. Son secuencias que describen mayoritariamente interrupciones, repeticiones de grado y también la combinación de ambas situaciones. Entre estas 10 trayectorias se destaca como la segunda con mayor proporción aquella que indica la no transición de la primaria a la secundaria; representa 9.5% de la cohorte de inicio de secundaria, y está compuesta por los individuos que deberían haber iniciado secundaria, pues habían cursado sexto grado de primaria el ciclo anterior, y sin embargo no lo hicieron.

La tercera y cuarta trayectoria representan secuencias donde los alumnos sólo cursan uno o dos años de secundaria e interrumpen después sus estudios, al menos durante el resto del tiempo de observación (representan de manera conjunta 7.2% de la cohorte).

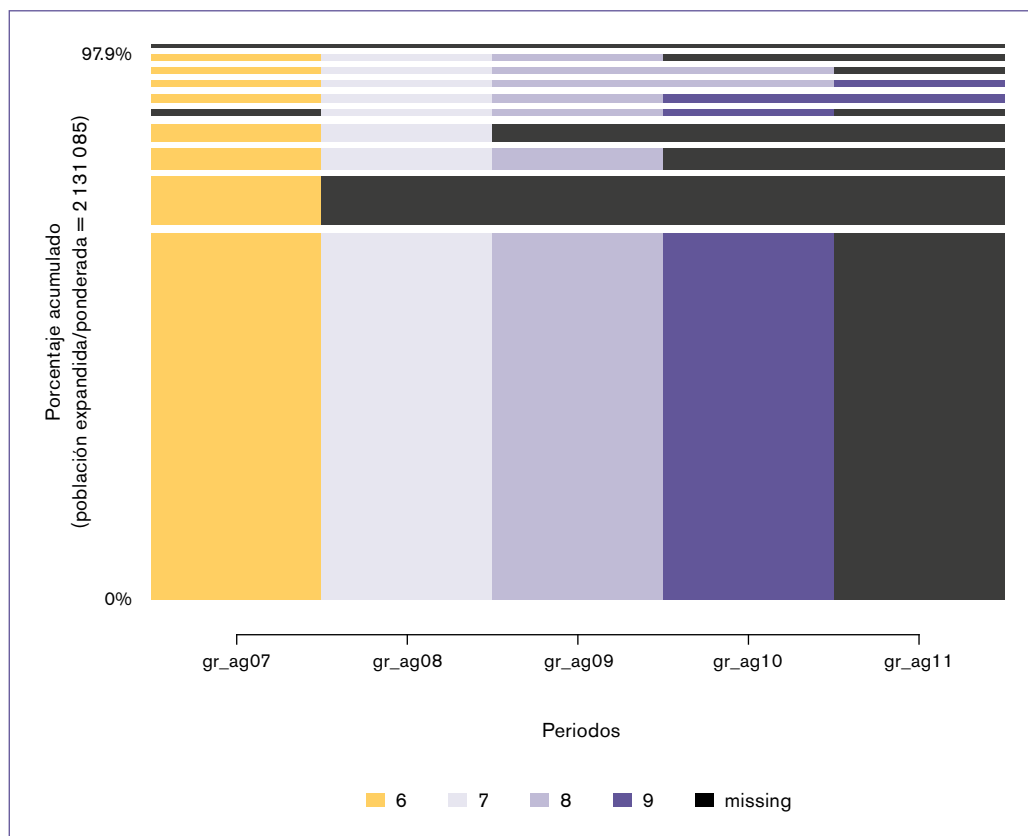
La quinta trayectoria disminuye mucho en su participación; sólo aporta 1.1%, pero puede interpretarse, en términos de la trayectoria educativa en secundaria, como normativa, ya que sigue desde agosto de 2008 a agosto de 2011 una secuencia regular y lineal. Se diferencia de la primera trayectoria normativa en que inicia en una fase $t-1$ (agosto 2007) con un no registro. En este sentido, se cree conveniente no interpretar esta ausencia como una interrupción propiamente dicha, ya que, como se señaló en el apartado metodológico, la base de RNA en el ciclo 2007-2008 presenta ausencia de registros incluso en entidades federativas completas.

Cuadro 22 Distribución de trayectorias educativas

Secuencias	Base expandida		Base sin expandir	
	Casos	%	Casos	%
6-7-8-9*	1 618 526	75.9	3 797	76.0
6*-*-*	202 370	9.5	475	9.5
6-7-8-*	87 485	4.1	205	4.1
6-7-*-*	65 933	3.1	155	3.1
-7-8-9	23 199	1.1	54	1.1
6-7-8-9-9	21 786	1.0	51	1.0
6-7-8-8-9	21 127	1.0	50	1.0
6-7-8-8*	15 028	0.7	35	0.7
6-7-7-*	13 587	0.6	32	0.6
6-7-7-8-9	12 166	0.6	29	0.6
6*-7-*	6 450	0.3	15	0.3
6*-7-8-9	5 160	0.2	12	0.2
6-7-7-8*	5 160	0.2	12	0.2
6-7-8*-8	4 730	0.2	11	0.2
6-7*-8-9	4 300	0.2	10	0.2
6*-*-7*	3 010	0.1	7	0.1
-7-8-	2 580	0.1	6	0.1
6-7-8-8-8	2 580	0.1	6	0.1
6*-7-8*	2 150	0.1	5	0.1
-7--*	1 720	0.1	4	0.1
6*-*-7-8	1 720	0.1	4	0.1
6-7*-8*	1 720	0.1	4	0.1
6*-*-8-9	1 290	0.1	3	0.1
6*-*-9	860	0.0	2	0.0
--*7	430	0.0	1	0.0
*-*9*	430	0.0	1	0.0
*-*7-*	430	0.0	1	0.0
*-*7-8*	430	0.0	1	0.0
-7-7-	430	0.0	1	0.0
*-7-8-8-9	430	0.0	1	0.0
6*-*-7	430	0.0	1	0.0
6*-7-7*	430	0.0	1	0.0
6*-7-8-8	430	0.0	1	0.0
6-7*-7*	430	0.0	1	0.0
6-7*-7-7	430	0.0	1	0.0
6-7*-7-8	430	0.0	1	0.0
6-7-7-7*	430	0.0	1	0.0
6-7-7-7-8	430	0.0	1	0.0
6-7-7-8-8	430	0.0	1	0.0
Totales	2 131 087	100.0	4 999	100.0

Fuente: elaboración propia con muestra de cohorte secundaria.

Gráfica 8 Las 10 secuencias más frecuentes en secundaria



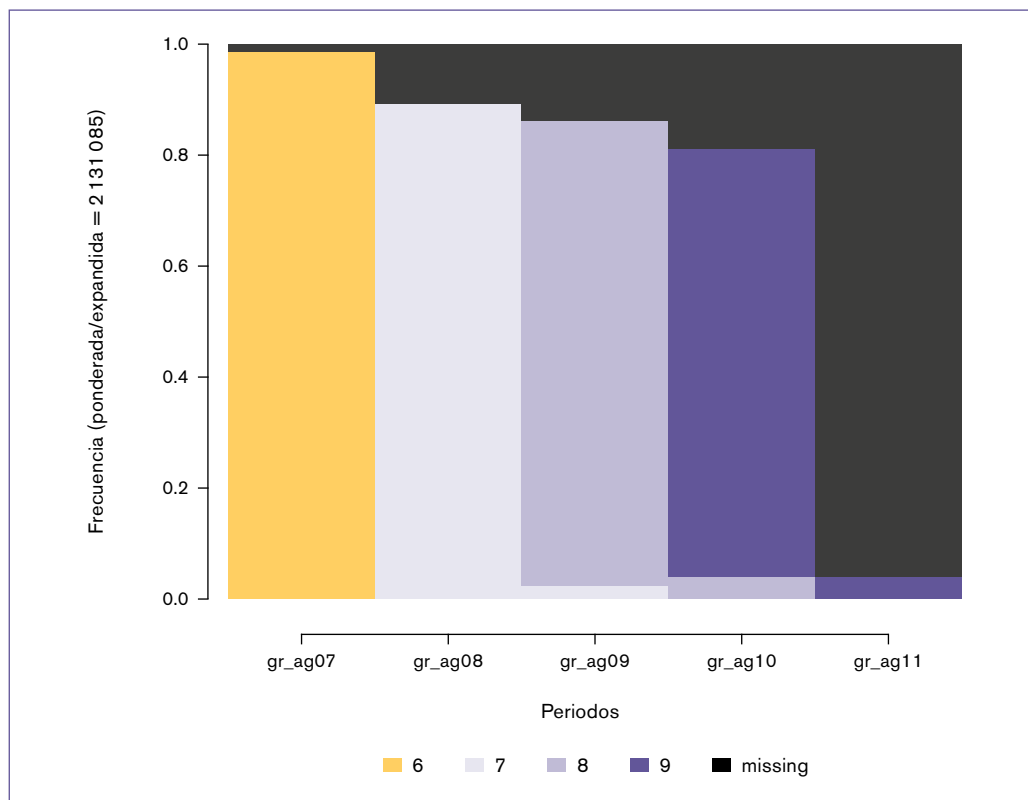
Fuente: elaboración propia con muestra ponderada de cohorte secundaria.

La gráfica 9 presenta la distribución de estados por periodos de observación (histograma), y permite ver que las interrupciones (no registros) en los diferentes ciclos educativos son más frecuentes en cada periodo que las repeticiones de grado.

Cuando se observa la distribución de secuencias por sexo tal como se presentan en el cuadro 23, se encuentran diferencias tanto en las frecuencias que arrojan trayectorias particulares como en su ordenamiento general.

Las mujeres presentan mayor proporción de trayectorias normativas y regulares que los varones, pues concentran 79% de experiencias individuales en esta modalidad de trayectoria frente a 73% alcanzado por los varones. En la gráfica 10 se representan las 10 secuencias más frecuentes por sexo; se hace evidente que los varones tienen mayor proporción de casos en las trayectorias con repeticiones e interrupciones, e incluso en la trayectoria que reporta la no transición de la primaria a la secundaria.

Gráfica 9 Histograma de secuencias en secundaria



Fuente: elaboración propia con muestra ponderada de cohorte secundaria.

La composición de las trayectorias educativas por edad al inicio de la secundaria muestra diferencias muy importantes entre los distintos grupos. Si bien en el caso de la cohorte de ingreso a secundaria se encuentran individuos menores a la edad normativa, éstos representan un grupo muy reducido y clasifican como menores por un número reducido de meses. De esta manera, si bien se presentan las trayectorias educativas de este grupo en el cuadro 24, el análisis se centrará en los grupos mayoritarios, que son los que se encuentran en edad normativa y los que presentan extraedad al inicio de la secundaria; por otro lado, se considera que los comportamientos del grupo de menores a la edad normativa quizás pueden asimilarse a los de quienes están en edad normativa.

Aquellos individuos que inician secundaria en agosto de 2008 en edad normativa presentan los niveles más altos de trayectorias regulares (84.1%). Asimismo, presentan los niveles más bajos en las trayectorias educativas que señalan la no transición de la primaria a la secundaria (codificada como 6-**-**) y en las que reportan interrupciones y repeticiones.

Cuadro 23 Distribución de trayectorias educativas en secundaria por sexo

Mujeres			Varones		
Secuencias	Casos	%	Secuencias	Casos	%
6-7-8-9*	1943	79.0	6-7-8-9*	1854	73.0
6*-7-8*	217	8.8	6*-7-8*	258	10.2
6-7-8-*	98	4.0	6-7-8-*	107	4.2
6-7-7-*	66	2.7	6-7-7-*	89	3.5
-7-8-9	28	1.1	6-7-8-8-9	37	1.5
6-7-8-9-9	18	0.7	6-7-8-9-9	33	1.3
6-7-8-8-9	13	0.5	*-7-8-9*	26	1.0
6-7-7-*	12	0.5	6-7-8-8-*	25	1.0
6-7-8-8-*	10	0.4	6-7-7-8-9	21	0.8
6*-7-7-*	9	0.4	6-7-7-*	20	0.8
6-7-7-8-9	8	0.3	6-7-7-8-*	8	0.3
6*-7-8-9	5	0.2	6-7-8-8-8	8	0.3
6-7*-8-9	5	0.2	6*-7-8-9	7	0.3
6-7-7-8-*	4	0.2	6*-7-7-*	6	0.2
-7-8-	3	0.1	6-7-8-8-8	6	0.2
6*-7-7-*	3	0.1	6-7*-8-9	5	0.2
6-7-8-8-*	3	0.1	6*-7-7-*	4	0.2
-7-7-	2	0.1	6*-7-8-*	4	0.2
6*-7-8	2	0.1	*-7-8-*	3	0.1
6*-7-8-9	2	0.1	6-7*-8-*	3	0.1
-7-7-	1	0.0	*-7-7-*	2	0.1
-7-8-	1	0.0	6*-7-7-8	2	0.1
6*-7-7-*	1	0.0	*-7-9-*	1	0.0
6*-7-8-*	1	0.0	*-7-7-*	1	0.0
6*-7-8-8	1	0.0	*-7-8-9	1	0.0
6-7*-8-*	1	0.0	6*-7-8-9	1	0.0
			6*-7-7-*	1	0.0
			6-7*-7-*	1	0.0
			6-7*-7-7	1	0.0
			6-7*-7-8	1	0.0
			6-7-7-7-*	1	0.0
			6-7-7-7-8	1	0.0
			6-7-7-8-8	1	0.0
Totales	2458	100.0	Totales	2541	100.0

Fuente: elaboración propia con muestra de cohorte secundaria.

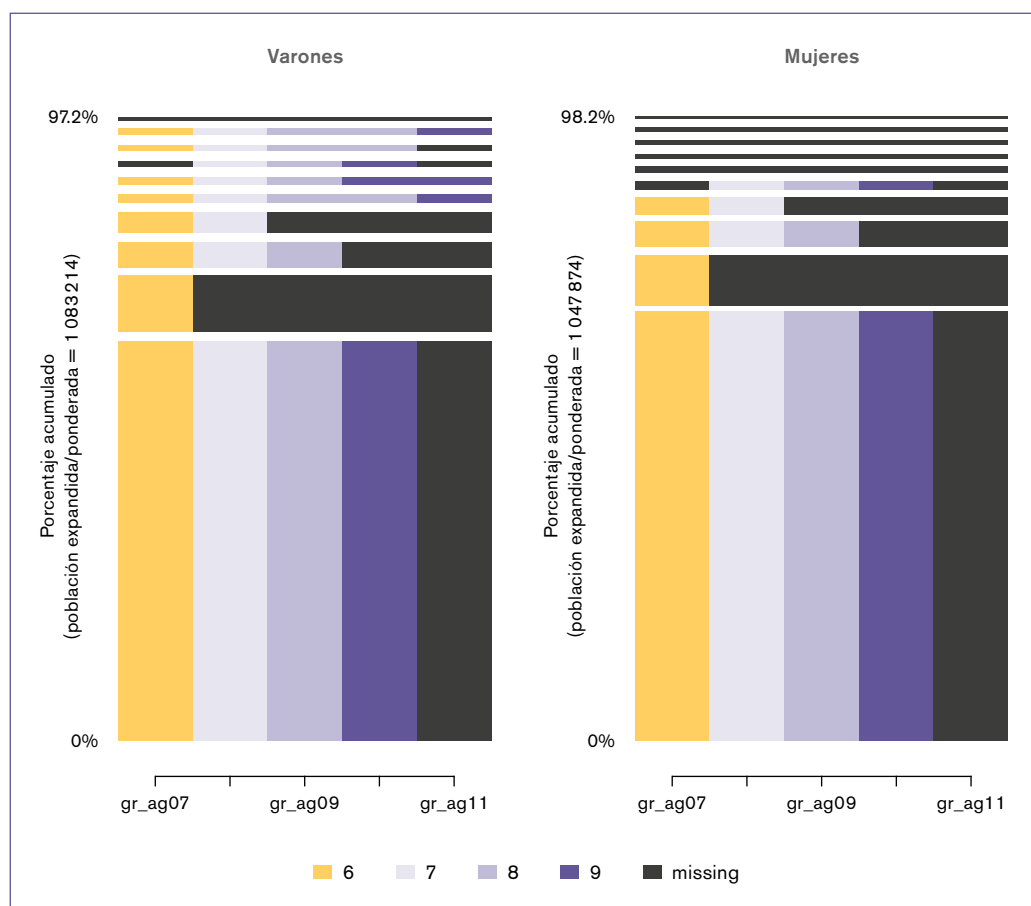
Cuadro 24 Distribución de trayectorias educativas por edad al inicio de la primaria

Menores a edad normativa			En edad normativa			Extraedad		
Secuencias	Casos	%	Secuencias	Casos	%	Secuencias	Casos	%
6-7-8-9-*	10	62.5	6-7-8-9-*	2462	84.1	6-7-8-9-*	1325	64.5
6-*-*-*	2	12.5	6-*-*-*	132	4.5	6-*-*-*	341	16.6
6-7-8-*	2	12.5	6-7-8-*	71	2.4	6-7-8-*	132	6.4
6-*-*-7-*	1	6.2	6-7-*-*-*	45	1.5	6-7-*-*-*	110	5.4
6-*-7-*	1	6.2	*-7-8-9-*	37	1.3	6-7-7-*	21	1.0
Totales	16	100.0	6-7-8-8-9	36	1.2	*-7-8-9-*	17	0.8
			6-7-8-9-9	35	1.2	6-7-8-9-9	16	0.8
			6-7-8-8-*	21	0.7	6-7-7-8-9	15	0.7
			6-7-7-8-9	14	0.5	6-7-8-8-*	14	0.7
			6-7-7-*	11	0.4	6-7-8-8-9	14	0.7
			6-*7-8-9	9	0.3	6-*7-*	8	0.4
			6-7-*8-9	8	0.3	6-7-7-8-*	5	0.2
			6-7-7-8-*	7	0.2	6-7-8-*8	4	0.2
			6-7-8-*8	7	0.2	*-7-8-*	3	0.1
			6-*7-*	6	0.2	6-*-*7-*	3	0.1
			6-7-8-8-8	4	0.1	6-*7-8-*	3	0.1
			-7-8-	3	0.1	6-*7-8-9	3	0.1
			6-*-*-7-*	3	0.1	*-7-*-*-*	2	0.1
			-7--*	2	0.1	6-*-*7-8	2	0.1
			6-*-*7-8	2	0.1	6-*-*8-9	2	0.1
			6-*7-8-*	2	0.1	6-7-*8-*	2	0.1
			6-7-*8-*	2	0.1	6-7-*8-9	2	0.1
			--*-7	1	0.0	6-7-8-8-8	2	0.1
			--9-*	1	0.0	*-*7-*	1	0.0
			6-*-*-9	1	0.0	*-*7-8-*	1	0.0
			6-*-*8-9	1	0.0	*-7-7-*	1	0.0
			6-*7-8-8	1	0.0	*-7-8-8-9	1	0.0
			6-7-*7-*	1	0.0	6-*-*7	1	0.0
			6-7-*7-8	1	0.0	6-*-*-9	1	0.0
			6-7-7-7-8	1	0.0	6-*7-7-*	1	0.0
			6-7-7-8-8	1	0.0	6-7-*7-7	1	0.0
			Totales	2928	100.0	6-7-7-7-*	1	0.0
						Totales	2055	100.0

Fuente: elaboración propia con muestra de cohorte secundaria.

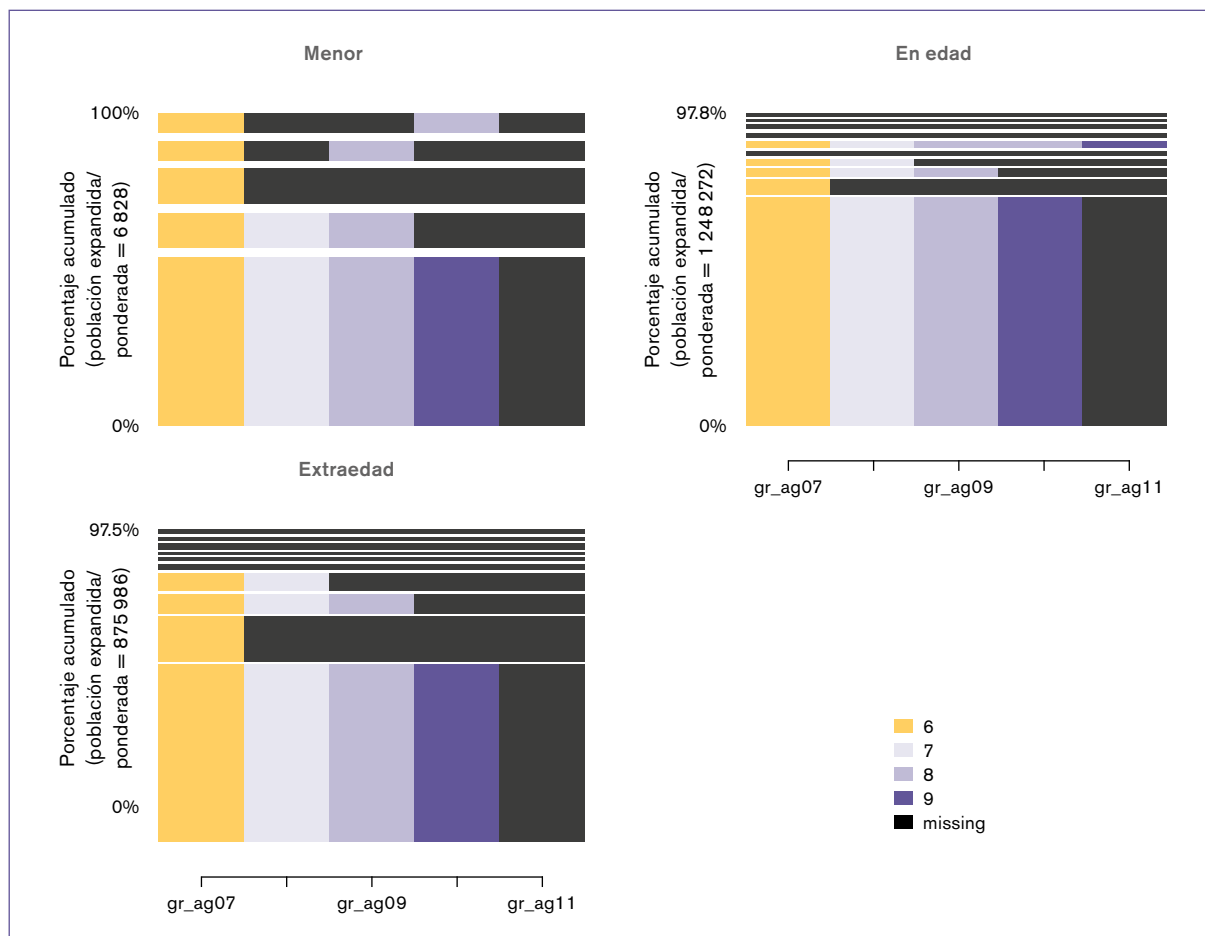
Sin embargo, el grupo que inicia secundaria con extraedad presenta trayectorias educativas con distribuciones muy diferentes al grupo en edad normativa. Sólo 64.5% presenta trayectorias regulares y normativas, y una gran proporción, que alcanza 16.5%, no realiza la transición de primaria a secundaria. Además en este grupo las trayectorias de interrupciones en primero y segundo grados de secundaria triplican las del grupo en edad normativa al inicio de la secundaria (algo muy notorio en la gráfica 11). Definitivamente, la extraedad en el momento de iniciar secundaria en agosto de 2008 indica que las trayectorias pasadas se tornan factores de peso en las posibilidades de un tránsito exitoso en la trayectoria educativa posterior en este grupo específico de población.

Gráfica 10 Las 10 secuencias más frecuentes en secundaria, por sexo



Fuente: elaboración propia con muestra ponderada cohorte primaria.

Gráfica 11 Las 10 secuencias más frecuentes en secundaria, por sexo



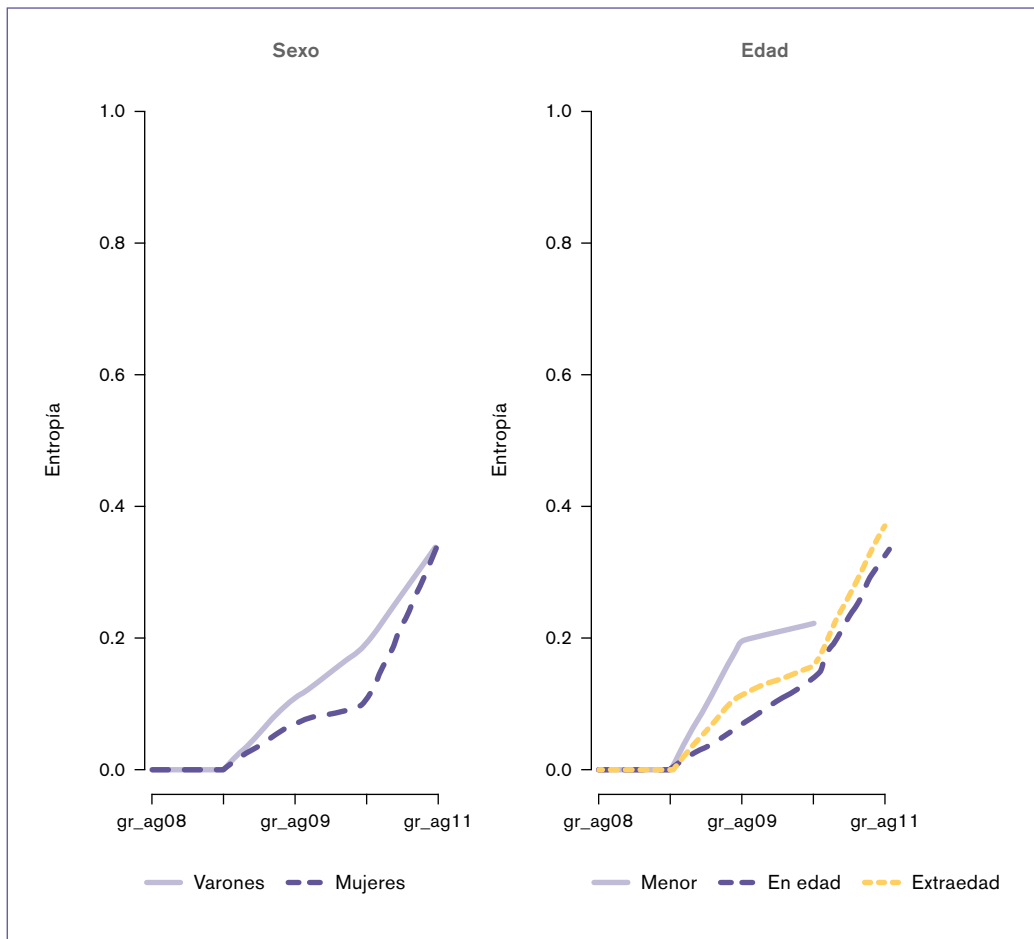
Fuente: elaboración propia con muestra ponderada de cohorte primaria.

En la gráfica 12 se presentan los resultados de la entropía transversal por sexo y edad al inicio de la secundaria; tales comportamientos entrópicos no sólo resumen las diferencias entre estas subpoblaciones, sino que también permiten mostrar los puntos de quiebre en las trayectorias educativas en secundaria, donde las interrupciones y repeticiones diferencian paulatinamente las experiencias de la cohorte analizada, lo cual propicia, por tanto, procesos de heterogeneidad transversal en los ciclos educativos.

Por la restricción que impone trabajar con una cohorte de inicio de secundaria en agosto de 2008, se parte de una entropía 0 en ese periodo de observación, ya que todos los individuos están en primer grado de secundaria y se obtiene la total homogeneidad entre estados. En el siguiente ciclo educativo, que inició en agosto de 2009, la heterogeneidad comienza a

manifestarse con incrementos en los niveles de entropía. En este sentido, en concordancia con lo que ya se pudo observar en la distribución de secuencias por covariantes, son los varones y los que están en extraedad quienes manifiestan mayor heterogeneidad en esta fase respecto a sus contrapartes: mujeres y aquellos en edad normativa al inicio de la secundaria respectivamente.

Gráfica 12 Índice de entropía transversal en trayectorias de primaria temprana, por sexo y edad al inicio de la primaria



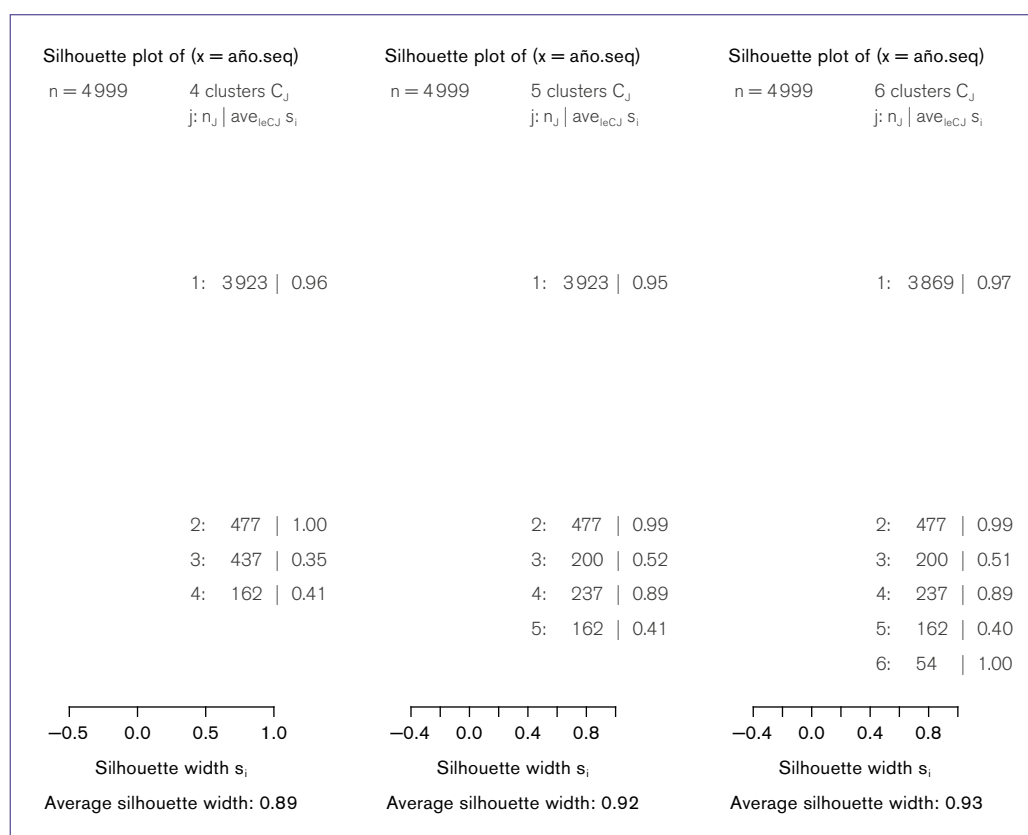
Fuente: elaboración propia con muestra ponderada de cohorte secundaria.

De todas formas, la brecha entre sexos es mayor en el tercer ciclo educativo (agosto de 2010), lo que muestra la dificultad de los varones por alcanzar tercer grado de secundaria por los mayores niveles de interrupciones en ciclos anteriores y las repeticiones de primer y segundo grado de secundaria.

Análisis estadístico de trayectorias

Si bien se ha detectado un nivel importante de concentración de experiencias en torno a un grupo reducido de trayectorias educativas, una cuarta parte de ellas mantienen diferencias importantes entre sí. Una forma de comparar secuencias, estimar distancias entre secuencias y agruparlas en torno a patrones de trayectorias es mediante *optimal matching*. Este método permitió alinear el conjunto completo de secuencias en torno a la estimación de una matriz de distancia interindividual de secuencias, para la cual se utilizó una matriz de costos y sustitución basada en las tasas de transición entre estados. Una vez estimada la matriz de distancia, se usó como insumo para agrupar y clasificar las trayectorias en grupos o tipologías mediante un análisis de clúster (jerárquico con criterio *ward*). Al igual que lo realizado para la cohorte que iniciaba primaria, se utilizó la técnica *silhouettes* para determinar el número conveniente de grupos que se usarían para el análisis de clúster. En la figura 4 se presentan los resultados del análisis de *silhouettes* para las soluciones de clúster de $k=4$, $k=5$ y $k=6$.

Figura 4 Salidas gráficas del análisis *silhouette* para soluciones de clúster 4, 5 y 6

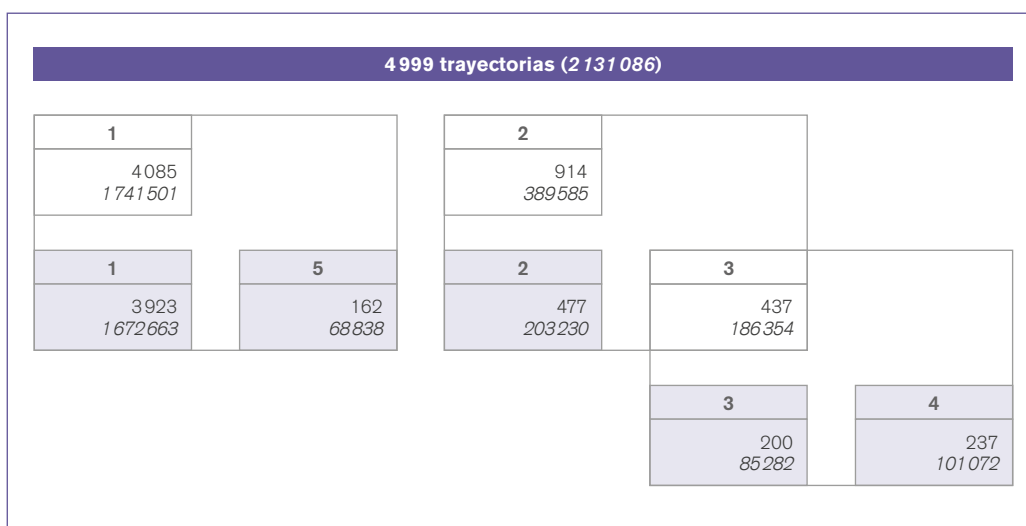


Fuente: elaboración propia con muestra de cohorte secundaria.

El análisis de *silhouettes* permite identificar que a partir de 5 grupos se alcanza un nivel de *average silhouettes width* de 0.92, mientras en la solución de 4 grupos se mantenía por debajo de 0.90. La solución de 5 grupos presenta 3 de sus grupos con niveles de *sil* superiores a 0.9 y dos grupos que presentan mayor heterogeneidad interna. Al realizar el análisis de *silhouettes* con 6 grupos esta situación no cambia, ya que el nuevo grupo surge de una subdivisión del grupo 1 y no de aquellos con mayores niveles de heterogeneidad interna. Los casos separados del grupo 1 original en la solución 6 de clúster son las 54 trayectorias que ya se había observado anteriormente que compartían una secuencia lineal y normativa en los grados de secundaria, pero que partían de no haber cursado sexto de primaria el ciclo 2007-2008 (*-7-8-9-*). Por tales razones se trabajará con la solución de $k=5$.

En la figura 5 puede observarse la estructura de árbol de la conformación de grupos de trayectorias a la que se ha aludido anteriormente para la solución de clúster escogida, así como la distribución de trayectorias en cada grupo.

Figura 5 Estructura de árbol entre los ocho grupos de trayectorias



Fuente: elaboración propia con muestra de cohorte secundaria.

En el cuadro 25 puede observarse la distribución de trayectorias entre los 5 grupos, y en la figura 6 se grafican las secuencias individuales para cada uno de los 5 grupos de trayectorias. De esta manera, se observa que el primer grupo, que representa 78.5%, contiene todas las trayectorias identificadas como normativas que completan los tres grados de secundaria, incluidas las que inician sin haber cursado sexto grado de primaria en el ciclo anterior. El resto de las experiencias de la cohorte se distribuye entre los cuatro grupos restantes.

Cuadro 25 Distribución de grupos de trayectorias

	Casos	%
1	3923	78.5
2	477	9.5
3	200	4.0
4	237	4.7
5	162	3.2
Total	4999	100.0

Fuente: elaboración propia con muestra de cohorte secundaria.

Con base en la agrupación de trayectorias por grupo que se observa en las gráficas de la figura 6 y su composición por covariante, que se presenta en el cuadro 26, se realizó el análisis de las tipologías y clasificación de las trayectorias educativas.

Tipología 1: trayectorias educativas normativas (78.5%)

Esta tipología agrupa las experiencias de la gran mayoría de la cohorte analizada y representa, como ya se ha mencionado, la trayectoria normativa, regular y lineal. Respecto a la composición de este grupo, se observa que está relativamente balanceado entre sexos, pero mayormente compuesto por individuos en edad normativa al inicio de la secundaria (más de las dos terceras parte). Respecto a la modalidad de servicios, más de la mitad de las trayectorias pertenecen a escuelas con modalidad general. De todas maneras, existe presencia de trayectorias con modalidades de servicio técnica y telesecundaria.

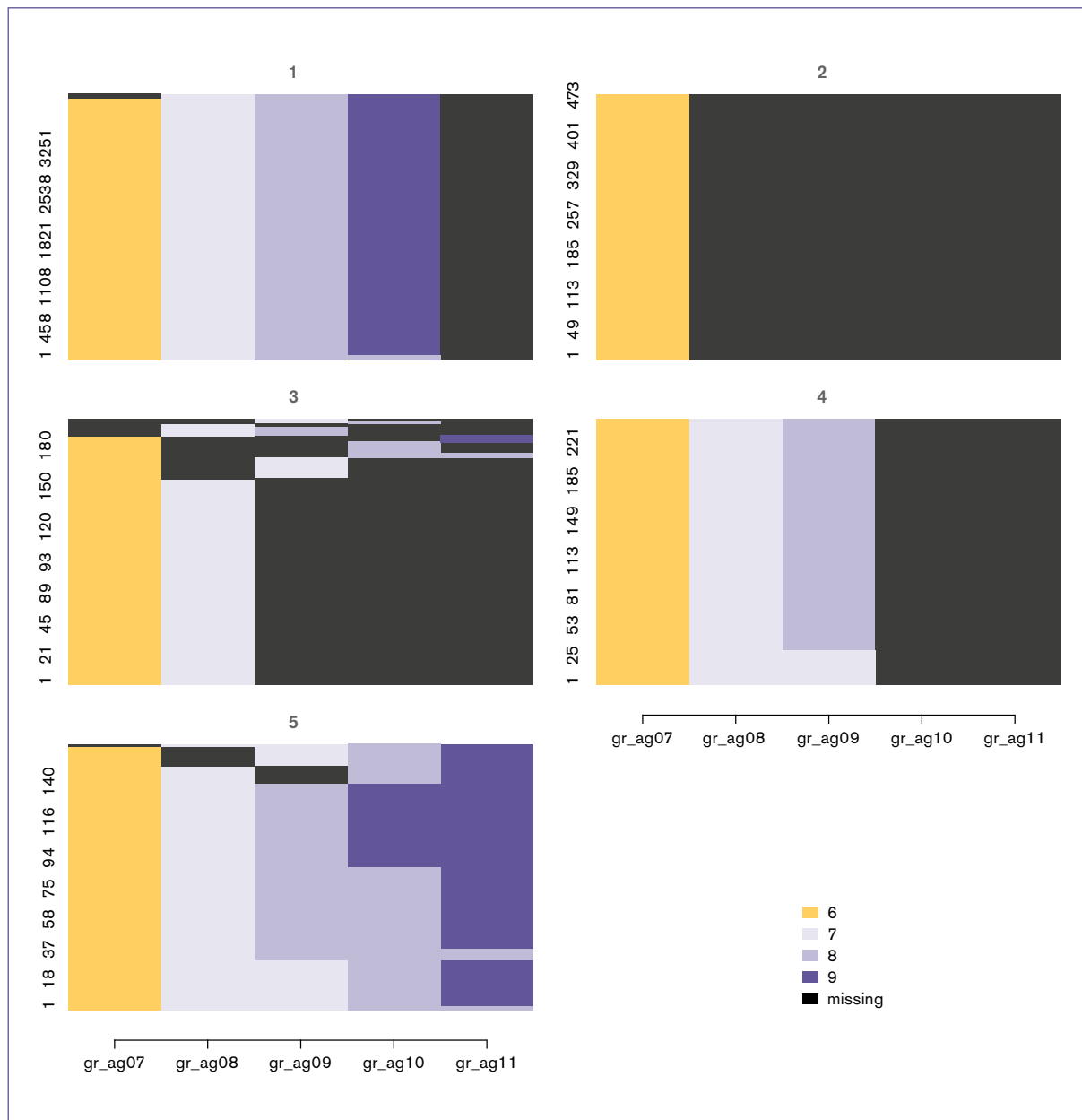
Tipología 2: trayectorias sin transición a secundaria (9.5%)

Esta tipología agrupa y distingue a quienes no ingresan a secundaria en agosto de 2008 aunque cursaron sexto de primaria en el ciclo anterior. Cabe señalar que por la forma en la que se seleccionó a la cohorte de inicio de secundaria en agosto de 2008, se excluía a aquellos que en ese mes ingresaron a sexto de primaria nuevamente, por lo cual no puede haber una confusión con casos de repetición en el último grado de primaria. Respecto a la composición de este grupo, puede observarse que hay más varones que mujeres, y que más de 70% de las trayectorias son de individuos que presentan extraedad al inicio de secundaria.

Tipología 3: trayectorias que alcanzan sólo primer grado de secundaria (4%)

Este grupo de trayectorias refleja experiencias de interrupciones o salidas de la escuela tras haber cursado sólo el primer año de secundaria. Respecto a su composición se integran por más varones que mujeres y un 67% presentan extraedad al inicio.

Figura 6 Trayectorias individuales por grupos de trayectorias (1 al 5)



Fuente: elaboración propia con muestra ponderada de cohorte secundaria.

Cuadro 26 Composición de los cinco grupos de trayectorias por covariantes

	1	2	3	4	5
Sexo					
Varones	49.3	51.1	55.5	53.6	69.1
Mujeres	50.7	48.9	44.5	46.4	30.9
Edad al inicio de secundaria					
Menor	0.3	0.4	1.0	0.8	0.0
En edad	64.7	27.9	32.0	34.6	67.3
Extraedad	35.0	71.7	67.0	64.6	32.7
Modalidad de servicio al inicio de la secundaria					
General	51.2	0.0	41.0	50.6	48.1
Técnica	27.4	0.0	24.5	26.2	30.9
Telesecundaria	17.8	0.0	17.0	17.3	9.3
Trabajo	0.3	0.0	1.0	0.8	1.2
NA	3.3	100.0	16.5	5.1	10.5

Fuente: elaboración propia con muestra de cohorte secundaria.

Tipología 4: trayectorias que alcanzan sólo segundo grado de secundaria (4.7%)

Esta tipología de trayectorias comparte con la anterior la presencia de interrupciones, pero en un nivel más avanzado de los estudios: a partir de haber cursado segundo grado de secundaria. Ninguno de los miembros de este grupo inicia tercer grado, y sólo una pequeña proporción repitió primer grado. Se compone por 7% más varones que mujeres, y casi 65% presenta extraedad al inicio de la secundaria.

Tipología 5: Trayectorias que completan los tres años secundaria con repeticiones (3.2%)

Este grupo de trayectorias comparte puntos de partida y llegada con la tipología 1, aunque no de forma lineal ni regular. Los individuos que lo conforman requieren un ciclo educativo más (agosto de 2011) para alcanzar lo que los miembros de la tipología 1 alcanzan un ciclo anterior. Cabe hacer una precisión en esta comparación. Para el caso de la tipología 1 puede constatar que en agosto de 2011 ninguno de sus miembros cursaba nuevamente tercer grado, lo que permite inferir promoción, pero en el caso de esta tipología 5 se requeriría de un periodo de observación más (2012-2013) para constatar si en el ciclo 2011-2012 efectivamente hay promoción. Respecto a su composición, 70% son varones, y 70% están en edad normativa al inicio de la secundaria. Este grupo se distingue en su composición con respecto al grupo de trayectorias de abandono justamente por la edad al inicio. Es la tipología que presenta la mayor prevalencia en escuelas técnicas, con 30%, y donde la participación de telesecundaria es muy baja.

Medidas resumen de las trayectorias educativas

En este apartado se presentan los resultados de tres medidas resumen estimadas para las trayectorias educativas de la cohorte secundaria. Igual que con la cohorte de primaria, se trabajó con la distribución por covariantes de duraciones en cada grado escolar —como una medida aproximada para dar cuenta de los procesos de repeticiones de grado—; la distribución por covariantes de las transiciones realizadas al final del periodo de observación, y la estimación de la matriz de transiciones para la población total y por sexo.

En los cuadros 27 y 28 se presentan los resultados de las duraciones en primero, segundo y tercer grado de secundaria. El periodo de observación impone para el caso de la cohorte secundaria la misma restricción que lo explicitado para la cohorte de primaria. A saber, que hay desigualdad de tiempo para observar oportunidades de repetición en los distintos grados escolares observados, por lo cual se recomendó no comparar las distribuciones entre los grados escolares.

En primer lugar es interesante destacar la desigualdad entre subgrupos (covariantes) de las duraciones 0 en los tres grados escolares analizados, que son aquellas que dan cuenta de la población de los distintos subgrupos que no alcanzan a cursar cada uno de los grados analizados. En el caso de primer grado de secundaria, esta duración 0 representa la no transición a la secundaria que se agrupa en la tipología 2 de trayectorias analizadas. Nuevamente, puede constatar que la brecha por edad al inicio de la secundaria es muy grande, pues la proporción de personas en extraedad que no realizan la transición de la escuela a la secundaria en agosto de 2008, aunque cursaron sexto de primaria en el periodo anterior, es cuatro veces mayor que otros casos.

Cuadro 27 Duraciones en cada grado educativo (de 7° a 9°) por sexo y edad (%)

Duraciones	Grado de secundaria					
	Primero		Segundo		Tercero	
Sexo	Varones	Mujeres	Varones	Mujeres	Varones	Mujeres
0	10.3	9.0	15.2	12.7	21.8	17.7
1	87.5	90.1	81.7	86.2	76.9	81.6
2	2.1	1.0	2.8	1.1	1.3	0.7
3	0.1	0.0	0.2	0.0	0.0	0.0
Totales	100.0	100.0	100.0	100.0	100.0	100.0
Edad al inicio	En edad	Extraedad	En edad	Extraedad	En edad	Extraedad
0	4.6	16.7	6.9	23.9	11.1	32.1
1	94.2	81.1	90.7	74.4	87.7	67.2
2	1.2	2.1	2.3	1.6	1.2	0.8
3	0.0	0.1	0.1	0.1	—	—
Totales	100.0	100.0	100.0	100.0	100.0	100.0

Fuente: elaboración propia con muestra de cohorte secundaria.

Cuadro 28 Duraciones en cada grado educativo (de 7° a 9°) por modalidad de servicio al inicio de secundaria (%)

Primer grado de secundaria				
Modalidad al inicio	General	Técnica	Telesecundaria	Trabajo
0	0.0	0.0	0.0	0.0
1	98.1	97.6	99.0	100.0
2	1.8	2.3	1.0	0.0
3	0.1	0.1	0.0	0.0
Totales	100.0	100.0	100.0	100.0
Segundo grado de secundaria				
0	4.2	5.2	4.4	10.5
1	93.5	92.1	94.2	84.2
2	2.1	2.8	1.4	5.3
3	0.3	0.0	0.0	0.0
Totales	100.0	100.0	100.0	100.0
Tercer grado de secundaria				
0	10.8	10.7	10.4	21.1
1	88.1	87.8	89.0	78.9
2	1.1	1.5	0.6	0.0
Totales	100.0	100.0	100.0	100.0

Fuente: elaboración propia con muestra de cohorte secundaria.

Las duraciones 0 en segundo y tercer grado son el resultado de dos mecanismos: el de las interrupciones (como las observadas en los grupos de trayectorias de las tipologías 3 y 4), y el de repeticiones de grados que provocan que la cohorte analizada no curse segundo o tercer grados de secundaria en el periodo de observación (situación representada en las trayectorias de la tipología 5). Si se pudiera seguir observando a esta cohorte, estas distribuciones en duraciones 0 podrían verse modificadas.

De todas maneras, se evidencia la desigualdad en las trayectorias en el periodo que abarca desde agosto de 2008 a agosto de 2011, donde en un igual número de años y ciclos educativos para una cohorte de inicio hay quienes no alcanzan los mismos grados de avance en comportamientos diferenciados por sexo, edad al inicio y modalidades de servicio escolar. Son muy grandes las brechas por edad al inicio de la secundaria entre quienes alcanzan a cursar cada uno de los grados escolares analizados y los que no (con diferencias que alcanzan a cuadruplicar las proporciones para quienes presentan extraedad al inicio). Respecto a las modalidades de servicio escolar, aquellos que cursan en modalidades de trabajo tienen mayor proporción de individuos que no alcanzan a cursar segundo y tercer grados de secundaria en el periodo analizado: 20% aproximadamente, frente a 10% de las otras modalidades.

Otra manera de ver procesos de rezago y diferenciación de trayectorias en una cohorte es mediante un análisis de las transiciones entre estados efectivamente realizadas por distintos subgrupos de población. En el cuadro 29 se presenta la distribución por sexo, edad y modalidad de servicio escolar al inicio de secundaria de las transiciones que se realizaron. En este caso, como se inicia la observación en un periodo $t-1$, que incluye la posibilidad o no de transitar de sexto de primaria a primero de secundaria, se debe tener ese reparo a la hora de analizar las distribuciones y compararlas con las tablas de distribución de frecuencias de trayectorias anteriormente presentadas. Por ejemplo, en la proporción de quienes no realizaron ninguna transición en el cuadro 29 están contabilizados los que no pasaron de sexto a séptimo (primer grado de secundaria) así como los que, sin haber cursado sexto de primaria en el periodo anterior, nunca pasaron de séptimo a octavo grado.

Cuadro 29 Distribución por covariantes de número de transiciones realizadas (%)

Sexo	Varones		Mujeres	
0	10.4		9.0	
1	5.0		3.9	
2	7.6		6.1	
3	77.0		81.0	
Totales	100.0		100.0	
Edad	En edad		En extraedad	
0	4.6		16.8	
1	2.4		7.3	
2	5.4		9.0	
3	87.6		66.9	
Totales	100.0		100.0	
Modalidad al inicio	General	Técnica	Telesecundaria	Trabajo
0	0.1	0.2	0.1	0.0
1	4.2	5.1	4.4	10.5
2	7.4	7.1	7.1	10.5
3	88.2	87.7	88.3	78.9
Totales	100.0	100.0	100.0	100.0

Fuente: elaboración propia con muestra cohorte primaria.

Respecto a las transiciones realizadas por la cohorte secundaria en el periodo analizado y su comportamiento por las covariantes consideradas, se destaca que las brechas más importantes se observan entre aquellos con edad normativa y los que presentan extraedad al inicio de la secundaria.

El análisis de distribución de duraciones por grado educativo y de transiciones realizadas por subgrupos poblacionales puede complementarse con la estimación de tasas de transición entre estados. Dichas tasas se presentan como matriz de transiciones entre estados en el cuadro 26, para la población total y por sexo. Así como se hizo para la cohorte primaria, se incorpora en la matriz de transición a los casos faltantes como un estado posible que da cuenta de quienes no están registrados en algunos periodos de observación; en la matriz se indican con asterisco (*). Su análisis es interesante como estado de destino y como estado de partida. Permiten identificar con cierta precisión el peso de la tendencia a salirse o interrumpir los estudios a partir de ciertos grados anteriores, así como de retornar a otros grados tras alguna interrupción anterior.

Para el caso particular de la cohorte de secundaria, es necesario hacer ciertas precisiones respecto al papel que cumplen los no registros luego de que los individuos hayan cursado tercer grado de secundaria (celda donde se intercepta la fila de 9 con la columna *). Cuando se consideraron los periodos de observación, se incluyó el estado de los individuos de la cohorte analizada en el inicio del ciclo educativo 2011-2012. Esto permitía observar si los que cursaron noveno grado en el 2010-2011 estaban cursándolo nuevamente en el ciclo 2011-2012, es decir, si habían repetido o estaban señalados como no registrados en ningún nivel de secundaria. Esta última circunstancia permitía inferir la aprobación del tercer grado de secundaria, y por lo tanto de la totalidad del nivel educativo. De esta forma, la tasas que señalan la transición de noveno grado a no registro (celdas señalan con azul) daría cuenta de los niveles de promoción y adquisición de la secundaria completa por parte de quienes efectivamente cursaron tercer grado de secundaria.

Cuadro 30 Matriz de transiciones entre estados para población total, varones y mujeres

Total	[6->]	[7->]	[8->]	[9->]	[*->]
[6->]	0.000	0.893	0.000	0.000	0.107
[7->]	0.000	0.018	0.931	0.000	0.051
[8->]	0.000	0.000	0.023	0.913	0.064
[9->]	0.000	0.000	0.000	0.013	0.987
[*->]	0.000	0.055	0.013	0.001	0.931
Varones	[6->]	[7->]	[8->]	[9->]	[*->]
[6->]	0.000	0.887			0.113
[7->]		0.024	0.917		0.059
[8->]			0.034	0.894	0.071
[9->]				0.017	0.983
[*->]	0.000	0.052	0.014	0.002	0.932
Mujeres	[6->]	[7->]	[8->]	[9->]	[*->]
[6->]	0.000	0.900			0.100
[7->]		0.011	0.946		0.043
[8->]			0.011	0.933	0.056
[9->]				0.009	0.991
[*->]	0.000	0.058	0.011	0.001	0.930

Fuente: elaboración propia con muestra de cohorte secundaria.

Una de las celdas de gran relevancia analítica en la matriz de transiciones es la diagonal señalada con color amarillo en el cuadro 30. Permite ver la proporción de individuos que, partiendo de un grado escolar, alcanzan el siguiente (una medida de aproximación a la promoción entre grados de la cohorte). Esta diagonal también permite observar el proceso de pérdida y reducción que suponen para la cohorte analizada las repeticiones de grados y las salidas e interrupciones en cada ciclo educativo consecutivo. En el caso de la población total, se observa que la primera transición entre estados (de sexto de primaria a primero de secundaria) deja un grupo de individuos fuera, mientras aproximadamente 89% de la cohorte alcanza séptimo grado. Por otro lado, las transiciones entre los estados definidos por grados escolares dentro de secundaria (7 a 8 y 8 a 9) presentan menores niveles de pérdida, con una retención aproximada de 93% y 91% respectivamente.

Esa reducción de la cohorte observada a través de las celdas amarillas se distribuye entre dos situaciones: los que quedan en el grado anterior, o sea, repiten, y los que se salen o interrumpen (representados con el asterisco). De esta forma, las celdas resaltadas con color rosa permiten ver las tasas de retención por grado educativo. En este sentido, el octavo grado, que representa segundo grado de secundaria, alcanza tasas mayores que los otros grados educativos, con 2.3%.

Las celdas resaltadas con morado indican los distintos niveles de atracción que tiene la interrupción desde los distintos grados de la primaria temprana. En este sentido, como ya se explicó anteriormente, no se considera en este mismo sentido la tasa de transición entre el noveno grado (tercero de secundaria) y el no registro, señalada por esta razón con color azul. No debe considerarse propiamente como una interrupción o salida de la escuela, sino como una medida aproximada de la tasa de aprobación de noveno grado (tercer grado de secundaria). Esta tasa señala que casi la totalidad (98.7%) de aquellos que alcanzan a cursar tercer grado de secundaria lo aprueban.

Si se miran las otras tasas de transición entre los distintos grados y los no registros puede verse la atracción relativa a la interrupción desde distintos puntos de partida. La transición de sexto de primaria al no registro (interrupción) alcanza la mayor tasa de transición —próxima a 11%—.

Si se observan todos los comportamientos referidos en las matrices de transición por sexo, se pueden constatar los patrones diferenciales entre varones y mujeres que se han resaltado a lo largo del análisis ya realizado para la cohorte de secundaria 2008-2009. En el caso de los varones, son siempre mayores las tasas de retención en el mismo grado (celdas rosas) que las de las mujeres; pero la brecha es mucho mayor para el caso de octavo grado, donde la tasa se triplica hasta alcanzar 3.4%. Asimismo, siempre son menores los niveles de pérdida de mujeres a medida que se avanza entre los ciclos educativos respecto a los que presentan los varones. En resumen, los varones presentan mayor tendencia a la repetición y la interrupción respecto a las mujeres. ■■

8

CONCLUSIONES

El análisis desarrollado en el marco de este proyecto ha permitido generar información muy relevante respecto a las trayectorias educativas de dos cohortes de alumnos de distintos niveles. La posibilidad de contar con un registro individualizado y continuado en el tiempo a escala nacional ha permitido reconstruir todas las trayectorias educativas de alumnos de dos cohortes —aquellos que inician primaria y quienes inician secundaria en el ciclo 2008-2009— y acompañarlos hasta los inicios del ciclo 2011-2012.

En términos generales, se pudo constatar que existe una diversidad relativamente acotada de trayectorias educativas, y que una proporción cercana a las tres cuartas partes de las cohortes analizadas desarrolla trayectorias normativas, regulares y lineales. Asimismo, se pudo observar que aquellos que no siguen estas trayectorias normativas y lineales representan aproximadamente la cuarta parte de la población de las cohortes analizadas; que se diferencian entre sí, y que es en ellos donde reposa la heterogeneidad de trayectorias observada en ambas cohortes. De igual importancia se consideran los hallazgos respecto a la distribución desigual de la población entre los distintos patrones de trayectorias educativas en ambas cohortes analizadas. Al considerar subgrupos de población, esa relación entre tres cuartas partes de las trayectorias normativas o regulares y una cuarta parte con trayectorias irregulares se modifica de manera evidente.

Para la cohorte específica de primaria se pudo constatar que existían dos fuentes de heterogeneidad: las repeticiones de grado y las interrupciones o salidas al menos durante todo el periodo de observación que se pudo seguir a los alumnos (inicio de ciclo en agosto de 2011). De todas maneras, la fuente principal de heterogeneidad entre las trayectorias radica en distintos patrones de repetición; los alumnos repiten una o dos veces, y además, en diferentes grados. En la construcción de tipologías, cuando se busca agruparlas y clasificarlas, estas situaciones particulares se distinguen entre sí en grupos de trayectorias diferenciadas (tipologías 2, 5, 6, 7 y 8). Estas tipologías concentran, a su vez, un mayor número de casos respecto a aquellas que representan patrones de interrupción. Los patrones de interrupción no sólo están representados en un número menor de trayectorias sino que, además, sólo se distinguen en dos grupos de trayectorias o tipologías (tipologías 3 y 4).

Otro elemento a resaltar en los hallazgos de trayectorias de la primaria temprana de la cohorte de inicio en agosto de 2008 tiene que ver con la composición sociodemográfica de las distintas tipologías. Su análisis permitió constatar que existe una distribución diferenciada de la población entre las trayectorias, y que el sexo, la edad al inicio de la primaria y las modalidades de servicio escolar al inicio de la primaria resultan factores relevantes a la hora de posicionar a unos y otros alumnos en los distintos patrones de trayectorias educativas.

De esta manera, los varones tienen mayor presencia en aquellos grupos de trayectorias o tipologías con repeticiones. Los alumnos con extraedad al inicio de la primaria tienen mayor presencia en tipologías con interrupciones. Se destaca que aquellos alumnos que sólo cursan primer grado de primaria conforman la tipología que concentra mayor número de alumnos de escuelas con modalidad de servicio indígena.

Respecto a la cohorte de inicio de secundaria analizada, se observó que tanto patrones de repetición como de interrupción son fuentes de heterogeneidad en sus trayectorias. Pero a diferencia de lo observado en la cohorte de primaria analizada, en secundaria la interrupción o salida de la escuela por todo el periodo observado resulta tener mayor presencia que las repeticiones.

Al igual que en la cohorte de primaria, el sexo es un factor que posiciona a los individuos ante oportunidades diferenciales de desarrollar trayectorias más o menos regulares. La repetición y la interrupción son fenómenos más marcados en varones que en mujeres. Asimismo, la extraedad al inicio de la secundaria también se muestra como un elemento de alta capacidad clasificatoria.

En términos generales, la extraedad al inicio de ambos niveles educativos analizados se muestra como un factor muy importante en la composición de trayectorias no regulares ni normativas. Este hallazgo parece muy relevante, porque más allá del carácter exploratorio de este estudio y las limitaciones de posibilidad de construcción de otras variables explicativas de las trayectorias educativas, en esta primera aproximación al trabajo longitudinal con la base de datos del Registro Nacional de Alumnos (RNA) la edad emerge como una fuente de heterogeneidad de las trayectorias muy evidente. En este sentido, cabe recordar que la extraedad no sólo indica el tiempo individual en el que están posicionados los individuos al inicio de un nivel educativo, sino que es una variable de contenido también longitudinal al dar cuenta de manera implícita de trayectorias anteriores a la analizada, que si bien son desconocidas para este estudio, indudablemente deben haber sido irregulares. A pesar de la imposibilidad de distinguir en nuestro estudio si la extraedad es el resultado de salidas de la escuela, repeticiones o ambos procesos a la vez, ésta posiciona a los individuos en estados y ciclos educativos con rezago respecto a quienes inician en edades normativas. Y en un juego de causación acumulativa, son estos alumnos los que aparecen como quienes desarrollan trayectorias recientes también irregulares. ■■

9

RECOMENDACIONES

A continuación se sintetizan las principales recomendaciones relacionadas con el mejoramiento del Registro Nacional de Alumnos (RNA) para fortalecerlo como fuente principal de futuros análisis longitudinales:

- Integrar más períodos de inicio y fin de ciclo de RNA para tener más años consecutivos de observación, y la posibilidad de construir trayectorias que permitan ver niveles educativos completos.
- Mejorar la recolección en el proceso administrativo de armado del Registro para que sus datos tengan mejor calidad.
- Mejorar la cobertura del Registro tanto en las entidades federativas como en las escuelas. En el registro de 2007 a 2011, por ejemplo, están subrepresentadas las escuelas comunitarias.
- Integrar al registro el año aprobado al fin de ese ciclo y no sólo el año cursado, así como las reprobaciones en secundaria. Esto permitiría construir indicadores más precisos de repetición, deserción y aprobación. ■■

BIBLIOGRAFÍA

- Abbott, A. (1983). Sequences of Social Events: Concepts and Methods for the Analysis of Order in Social Processes. *Historical Methods*, 4, pp. 129-147.
- (1995). Sequence Analysis: New Methods for Old Ideas. *Annual Review of Sociology*, 21, pp. 93-113.
- (2001). *Time Matters: On Theory and Method*. Chicago: University of Chicago Press.
- y De Viney, S. (1992). The Welfare State as Transnational Event: Evidence from Sequences of Policy Adoption. *Social Science History*, 16(2), pp. 245-274.
- y Hrycak, A. (1990). Measuring Resemblance in Sequence Data: An Optimal Matching Analysis of Musicians' Careers. *American Journal of Sociology*, 96, pp. 144-185.
- y Tsay, A. (2000). Sequence Analysis and Optimal Matching Methods in Sociology. *Review and Prospect. Sociological Methods and Research*, 29(1), pp. 3-33.
- Aisenbrey, S. (2000). *Optimal Matching Analyse*. Opladen: Leske-Budrich.
- Billari, F.C. (2001). The Analysis of Early Life Courses: Complex Description of the Transition to Adulthood. *Journal of Population Research*, 18(2), pp. 119-142.
- y Piccarreta, R. (2005). Analyzing Demographic Life Courses through Sequence Analysis. *Mathematical Population Studies*, 12, pp. 81-106.
- Elder, G.H. (1985). Perspectives on the Life Course. En Glen H. Elder Jr. (ed.). *Life Course Dynamics, Trajectories and Transitions* (pp. 23-49). Ithaca: Cornell University Press.
- (1999). *Children of the Great Depression: Social Change in Life Experience*. Boulder: Westview Press.
- Elzinga, C.H. (2003). Sequence Similarity: A Nonaligning Technique. *Sociological Methods and Research*, 32, pp. 3-29.
- (2006). *Sequence Analysis: Metric Representations of Categorical Time Series*. Amsterdam: Department of Social Science Research Methods-Vrije Universiteit.
- Fussell, E. (2006). Structuring the Transition to Adulthood: An Entropy Analysis of the Early Life Course in the United States, 1880 to 2000, en Congress de la Population Association of America Meetings, Los Ángeles, California, Estados Unidos.
- Gabadinho, A., Ritschard, G., Studer, M., y Müller, N. (2011). *Mining Sequence Data in R with the TraMineR Package: A User's Guide*. Ginebra: Department of Econometrics and Laboratory of Demography-University of Geneva.
- Gauthier, J. (2007). *Empirical Categorizations of Social Trajectories: A Sequential View on the Life Course*. Lausanne: University of Lausanne.
- , Widmer, E., Bucher, P., y Notredame, C. (2009). How Much Does It Cost? Optimization of Costs in Sequence Analysis of Social Science Data. *Sociological Methods and Research*, 38(1), pp. 197-231.
- , Widmer, E., Bucher, P., y Notredame, C. (2010). Multichannel Sequence Analysis Applied to Social Science Data. *Sociological Methodology*, 40, pp. 1-38.
- Han S.K. (2012). *Motif of Sequence, Motif in Sequence*. Lausanne: LaCOSA.

- Hollister, M. (2009). Is Optimal Matching Suboptimal? *Sociological Methods and Research*, 38, pp. 235-264.
- Lesnard, L. (2006). Optimal Matching and Social Sciences (CREST Working Papers 2006-01). Paris: INSEE.
- Robles, H. et al. (2008). *Panorama educativo de México, 2008: Indicadores del sistema educativo nacional*. México: INEE.
- ____ et al. (2009). La eficacia y eficiencia del sistema educativo mexicano para garantizar el derecho a la escolaridad básica. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 7(4), pp. 48-77.
- ____ (2010, julio-diciembre). El sistema de indicadores del Instituto Nacional para la Evaluación de la Educación de México. *Sinéctica*, 35. Recuperado de: <http://www.sinectica.iteso.mx>
- Rousseeuw, P.J. (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math.*, 20, pp. 53-65.
- UNESCO. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (1984). *Métodos estadísticos para mejorar la estimación de la repetición y el abandono escolar: dos estudios metodológicos*. París: autor.
- Valencia R. E. (2012). Breve revisión de la calidad estadística del Registro Nacional de Alumnos (RNA), inicios y fines de los ciclos escolares 2007-2008 a 2010-2011 e inicio del ciclo escolar 2011-2012. Subdirección de Análisis Estadístico, Dirección de Indicadores Educativos del Instituto Nacional para la Evaluación de la Educación.

ANEXO

Códigos de programación en R

1 Cohorte primaria 2008-2009

```
#### Código de análisis de secuencia completo base muestra de Cohorte Primaria 2008/2009 ####
```

```
#### Karina Videgain ####
```

```
#### Noviembre 2014 ####
```

```
# Se abre la base de datos muestra para la cohorte que inicia primaria en 2008/2009
  ##Aclaración: habrá que modificar la ruta de acceso según su lugar en el disco.
prim<- read.dta("C:/INEE /Bases/muestra_primaria.dta")

# Identificación de casos duplicados en la muestra

anyDuplicated(prim$id)      ## No hay casos duplicados

# Establecer dónde se guardaran los resultados, bases y demás objetos
  ##Aclaración: habrá que modificar la ruta de acceso según su lugar en el disco.
setwd("C:/ INEE /Resultados//Muestra Primaria")
getwd()

# Creación de variables de interés
names(prim)
str(prim)

  ## Crear variable sexo

table(prim$sexo_i08)
## Hay casos sin declaración de sexo en la base de inicio de ciclo 2008
  ## Completar casos faltantes con información de otros ciclos

prim$sexo[prim$sexo_i08=="H" | prim$sexo_f08=="H" | prim$sexo_i09=="H" | prim$sexo_f09=="H" | prim$sexo_
i10=="H" | prim$sexo_f10=="H" | prim$sexo_i11=="H"] <- "H"
prim$sexo[prim$sexo_i08=="M" | prim$sexo_f08=="M" | prim$sexo_i09=="M" | prim$sexo_f09=="M" | prim$sexo_
i10=="M" | prim$sexo_f10=="M" | prim$sexo_i11=="M"] <- "M"

table(prim$sexo)
which(is.na(prim$sexo)) ## Se recupera información para todos los casos

class(prim$fecha_nac_i08)

## Declarar las variables de fecha de nacimiento como variables "date"

prim$nac_i08 <- as.Date(prim$fecha_nac_i08, "%d/%m/%Y")
prim$nac_f08 <- as.Date(prim$fecha_nac_f08, "%d/%m/%Y")
prim$nac_i09 <- as.Date(prim$fecha_nac_i09, "%d/%m/%Y")
prim$nac_f09 <- as.Date(prim$fecha_nac_f09, "%d/%m/%Y")
prim$nac_i10 <- as.Date(prim$fecha_nac_i10, "%d/%m/%Y")
prim$nac_f10 <- as.Date(prim$fecha_nac_f10, "%d/%m/%Y")
prim$nac_i11 <- as.Date(prim$fecha_nac_i11, "%d/%m/%Y")
```

```

summary(prim$nac_i08)
## Ver las nuevas variables en los 6 primeros casos de la base
head(prim$nac_i08)
head(prim$nac_i09)
head(prim$nac_i10)
head(prim$nac_i11)
## Hay edades que identificamos como problemas de registros, con nacimiento en 1911, 1953 y 2006.
## Podemos eliminar los casos, pero no queremos perder su trayectoria educativa
## Por lo tanto vamos a modificar las tres edades por lo que se cree podría ser la fecha correcta
## De todas formas, esto modifica solo para la variable edad y en 3 casos nada más, su trayectoria queda intacta.

prim$fnac <- as.Date(prim$fecha_nac_i08, "%d/%m/%Y")
prim$fnac[prim$nac_i08=="53"] <- "1993-07-12"
prim$fnac[prim$nac_i08=="1911-02-02"] <- "2001-02-02"
prim$fnac[prim$nac_i08=="2006-06-03"] <- "1996-06-03"
summary(prim$fnac)

# Crear variable que dé cuenta si está en un rango de edad indicado para el grado cursado
## En el ciclo 2008/2009 deben iniciar primero de primaria
## aquellos que hayan cumplido 6 años al 31 de diciembre del 2008,

## En edad (1): aquellos que cumplían 6 años, nacidos entre 01/01/2002 y 31/12/2002
## Extraedad (2): Nacidos antes del 01/01/2002
## Menor edad (0): Nacidos después del 31/12/2002

prim$edad <- "1"
prim$edad[prim$fnac<"2002-01-01"] <- "2"
prim$edad[prim$fnac>"2002-12-31"] <- "0"

table(prim$edad)

# En esta base de datos muestral proporcionada por el INEE ya están
## las variables de servicio, nivel, grado y secuencias creadas acorde
## con los códigos anteriormente establecidos por esta investigación.

#servicio: GEN = 2
#          IND = 3
#          " " = NA

table(prim$ser_i1)
table(prim$ser_i2)
table(prim$ser_i3)
table(prim$ser_i4)

# Hay datos faltantes en inicio que puede ser completado con servicio al final del mismo periodo

table(prim$ser_i1)
table(prim$ser_f1)
table(prim$ser_i1, prim$ser_f1)

prim$se_i1 <- prim$ser_i1
prim$se_i1[prim$ser_i1 == "NA" & prim$ser_f1 == "2"] <- "2"
prim$se_i1[prim$ser_i1 == "NA" & prim$ser_f1 == "3"] <- "3"
table(prim$ser_i1, prim$se_i1)
table(prim$se_i1)
table(prim$ser_i2)
table(prim$ser_f2)
table(prim$ser_i2, prim$ser_f2)

prim$se_i2 <- prim$ser_i2
prim$se_i2[prim$ser_i2 == "NA" & prim$ser_f2 == "2"] <- "2"
prim$se_i2[prim$ser_i2 == "NA" & prim$ser_f2 == "3"] <- "3"

table(prim$ser_i2, prim$se_i2)

```



```
table(prim$se_i2)
```

```
table(prim$ser_i3)  
table(prim$ser_f3)  
table(prim$ser_i3, prim$ser_f3)
```

```
prim$se_i3 <- prim$ser_i3  
prim$se_i3[prim$ser_i3 == "NA" & prim$ser_f3 == "2" ] <- "2"  
prim$se_i3[prim$ser_i3 == "NA" & prim$ser_f3 == "3" ] <- "3"
```

```
table(prim$ser_i3, prim$se_i3)  
table(prim$se_i3)
```

```
table(prim$ser_i4)  
prim$se_i4 <- prim$ser_i4
```

```
table(prim$se_i4)
```

```
table(prim$se_i1)  
table(prim$se_i2)  
table(prim$se_i3)  
table(prim$se_i4)
```

```
# Verificar distribución de variable que informa sólo del grado escolar cursado en cada período
```

```
table(prim$año_i1)  
table(prim$año_f1)  
table(prim$año_i2)  
table(prim$año_f2)  
table(prim$año_i3)  
table(prim$año_f3)  
table(prim$año_i4)
```

```
# Se encuentran información faltante respecto a años intermedios  
## que puede ser completada con la información del año anterior y el posterior.  
### A saber, se corrige con la siguiente codificación:
```

```
table(prim$año_i1)  
table(prim$año_i1, prim$año_f1 )  
prim$año_i1 <- prim$año_i1  
prim$año_i1[prim$año_i1 == "NA" & prim$año_f1 == "1"] <- "1"
```

```
table(prim$año_i1, prim$año_i1)  
table(prim$año_i1)
```

```
table(prim$año_i2)  
table(prim$año_f2)  
table(prim$año_i2, prim$año_f2)  
prim$año_i2 <- prim$año_i2  
prim$año_i2[prim$año_i2 == "NA" & prim$año_i1 == "1" & prim$año_i3 == "3" ] <- "2"  
prim$año_i2[prim$año_i2 == "NA" & prim$año_f1 == "1" & prim$año_i3 == "3" ] <- "2"  
prim$año_i2[prim$año_i1 == "1" & prim$año_i2 == "NA" & prim$año_f2 == "1" ] <- "1"  
prim$año_i2[prim$año_i1 == "1" & prim$año_i2 == "NA" & prim$año_f2 == "2" ] <- "2"  
prim$año_i2[prim$año_i2 == "NA" & prim$año_i1 == "1" & prim$año_f2 == "NA" & prim$año_f3 == "3" ] <- "2"  
prim$año_i2[prim$año_i2 == "NA" & prim$año_f2 == "NA" & prim$año_i1 == "1" & prim$año_i3 == "NA" &  
prim$año_i4 == "4" ] <- "2"  
prim$año_i2[prim$año_i2 == "NA" & prim$año_i1 == "1" & prim$año_i3 == "3" ] <- "2"  
prim$año_i2[prim$año_i2 == "NA" & prim$año_f1 == "1" & prim$año_f2 == "2" & prim$año_i3 == "NA" & prim$año_  
i4 == "4" ] <- "2"  
prim$año_i2[prim$año_i2 == "NA" & prim$año_f1 == "1" & prim$año_f3 == "3" ] <- "2"  
prim$año_i2[prim$año_i2 == "NA" & prim$año_f2 == "1" & prim$año_i3 == "2" ] <- "1"  
prim$año_i2[prim$año_i2 == "NA" & prim$año_f2 == "2" & prim$año_i3 == "NA" & prim$año_f3 == "3" ] <- "2"  
prim$año_i2[prim$año_i2 == "NA" & prim$año_i1 == "1" & prim$año_f2 == "2" & prim$año_i3 == "3" ] <- "2"  
prim$año_i2[prim$año_i2 == "NA" & prim$año_f2 == "2" & prim$año_i1 == "1" ] <- "2"  
prim$año_i2[prim$año_i2 == "NA" & prim$año_f2 == "2" & prim$año_i1 == "NA" & prim$año_f1 == "1" ] <- "2"
```

```

prim$añ_i2[prim$añ_i2 == "NA" & prim$añ_f1 == "1" & prim$añ_f2 == "1" & prim$añ_f3 == "2"] <- "1"
table(prim$añ_i2)
table(prim$añ_i2, prim$añ_i2)

table(prim$añ_i3)
table(prim$añ_f3)
table(prim$añ_i3, prim$añ_f3)
prim$añ_i3 <- prim$añ_i3
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_i2 == "2" & prim$añ_i4 == "4"] <- "3"
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_i1 == "1" & prim$añ_i2 == "NA" & prim$añ_f3 == "3"] <- "3"
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_i1 == "1" & prim$añ_i2 == "NA" & prim$añ_f3 == "2"] <- "2"
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_f3 == "NA" & prim$añ_i1 == "1" & prim$añ_i4 == "4"] <- "3"
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_i2 == "NA" & prim$añ_f3 == "3" & prim$añ_i4 == "4"] <- "3"
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_f3 == "NA" & prim$añ_i4 == "4" & prim$añ_f2 == "2"] <- "3"
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_f3 == "NA" & prim$añ_i2 == "2" & prim$añ_i4 == "4"] <- "3"
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_f3 == "3" & prim$añ_i4 == "NA" & prim$añ_i2 == "2"] <- "3"
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_f3 == "3" & prim$añ_f2 == "2" & prim$añ_i4 == "4"] <- "3"
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_f3 == "3" & prim$añ_i4 == "4" & prim$añ_f2 == "2"] <- "3"
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_i1 == "1" & prim$añ_i2 == "NA" & prim$añ_f2 == "NA" &
prim$añ_f3 == "1"] <- "1"
prim$añ_i3[prim$añ_i3 == "2" & prim$añ_i2 == "2" & prim$añ_f3 == "3" & prim$añ_i4 == "4"] <- "3"
prim$añ_i3[prim$añ_i3 == "3" & prim$añ_i2 == "2" & prim$añ_f3 == "2" & prim$añ_i4 == "3"] <- "2"
prim$añ_i3[prim$añ_i3 == "3" & prim$añ_i2 == "2" & prim$añ_f3 == "2" & prim$añ_i4 == "2"] <- "2"
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_i1 == "1" & prim$añ_i2 == "1" & prim$añ_f3 == "2"] <- "2"
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_i1 == "1" & prim$añ_i2 == "2" & prim$añ_f3 == "2"] <- "2"
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_f1 == "1" & prim$añ_i2 == "1" & prim$añ_f3 == "1"] <- "1"
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_f1 == "1" & prim$añ_f2 == "2" & prim$añ_f3 == "3"] <- "3"
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_f2 == "2" & prim$añ_f3 == "2"] <- "2"
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_f1 == "1" & prim$añ_i2 == "NA" & prim$añ_f3 == "3"] <- "3"
prim$añ_i3[prim$añ_i3 == "NA" & prim$añ_f1 == "1" & prim$añ_f2 == "1" & prim$añ_f3 == "2"] <- "2"
table(prim$añ_i3, prim$añ_i3)
table(prim$añ_i3)

table(prim$añ_i4)
prim$añ_i4 <- prim$añ_i4

table(prim$añ_i1)
table(prim$añ_i2)
table(prim$añ_i3)
table(prim$añ_i4)

# Re-crear la variable que informa de año y modalidad de servicio (i1 a i4) con las variables input corregidas

table(prim$i1)
table(prim$f1)
table(prim$i2)
table(prim$f2)
table(prim$i3)
table(prim$f3)
table(prim$i4)
## Crear estado por periodo que combine año y modalidad de servicio del centro

prim$as_i1 <- paste(prim$añ_i1, prim$sse_i1)
prim$as_i1[prim$añ_i1 == "NA"] <- "NA_NA"
table(prim$i1)
table(prim$as_i1)

prim$as_i2 <- paste(prim$añ_i2, prim$sse_i2)
prim$as_i2[prim$añ_i2 == "NA"] <- "NA_NA"
table(prim$i2)
table(prim$as_i2)

prim$as_i3 <- paste(prim$añ_i3, prim$sse_i3)
prim$as_i3[prim$añ_i3 == "NA"] <- "NA_NA"
table(prim$i3)

```

```

table(prim$as_i3)

prim$as_i4 <- paste(prim$año_i4, prim$se_i4)
prim$as_i4[prim$año_i4 == "NA"] <- "NA_NA"
table(prim$i4)
table(prim$as_i4)

# Generar variable con secuencia de modalidades para ver la presencia de cambios intermodalidad

prim$servicios <- paste(prim$se_i1, prim$se_i2, prim$se_i3, prim$se_i4)
tabla <- table(prim$servicios)
ser <- prop.table(tabla)

#convertir el resultado en archivo formato tabla
write.table(ser, file = "servicio.csv", append = FALSE, quote = TRUE, row.names = TRUE, col.names = TRUE)

# Generar variable que informe a la modalidad de servicio al inicio de primaria
prim$serv_inicio <- prim$se_i1

# Distribución de movimientos entre modalidades de servicio
## Pasaje de primer año de primaria a segundo año primaria

table(prim$se_i2)
with(prim, table(prim$serv_inicio, prim$se_i2)) -> serv1.serv2.table
prop.table(serv1.serv2.table, 1)

## Distribución de secuencias de modalidades desde el inicio a los años siguientes de primaria

prim$serv_prim <- paste(prim$se_i1, prim$se_i2, prim$se_i3, prim$se_i4)
tabla.servsercu <- table(prim$serv_prim)
tabla.servsercu

# Extraer de la clave CCT información de entidad federativa donde se localiza el centro escolar de las trayectorias

prim$ent_i08 <- (substr(prim$clavecct_i08, start=1, stop=2))
prim$ent_i09 <- (substr(prim$clavecct_i09, start=1, stop=2))
prim$ent_i10 <- (substr(prim$clavecct_i10, start=1, stop=2))
prim$ent_i11 <- (substr(prim$clavecct_i11, start=1, stop=2))

prim$ent1 <- as.integer(prim$ent_i08)
prim$ent2 <- as.integer(prim$ent_i09)
prim$ent3 <- as.integer(prim$ent_i10)
prim$ent4 <- as.integer(prim$ent_i11)

# Generar base igual a la actual pero sólo con variables de interés para el análisis (covariantes)
## Antes se debe cambiar nombre a variables para que las etiquetas que use en gráficos sean más evidentes
respecto a lo      ###que representan.

prim$ms_ag08 <- prim$se_i1
prim$ms_ag09 <- prim$se_i2
prim$ms_ag10 <- prim$se_i3
prim$ms_ag11 <- prim$se_i4

prim$gr_ag08 <- prim$año_i1
prim$gr_ag09 <- prim$año_i2
prim$gr_ag10 <- prim$año_i3
prim$gr_ag11 <- prim$año_i4

prim$am_ag08 <- prim$as_i1
prim$am_ag09 <- prim$as_i2
prim$am_ag10 <- prim$as_i3
prim$am_ag11 <- prim$as_i4

cov <- prim[,c("id", "sexo", "edad", "serv_inicio", "samplingweight", "ms_ag08", "ms_ag09", "ms_ag10", "ms_ag11",
"gr_ag08", "gr_ag09", "gr_ag10", "gr_ag11", "am_ag08", "am_ag09", "am_ag10", "am_ag11")]

```

```

# Generación de variables como factor: sexo, edad, modalidad de servicio al inicio

table(cov$edad)
cov$edadf <- factor(cov$edad, levels=0:2, labels=c("Menor", "En edad", "Extraedad"))
table(cov$edadf)

table(prim$sexo)
cov$sexof <- factor(cov$sexo, labels=c("Varones", "Mujeres"))
table(cov$sexof)

table(cov$serv_inicio)
cov$servinif <- factor(cov$serv_inicio, labels=c("General", "Indígena"))
table(cov$servinif)

table(cov$sexof, cov$edadf)

# Crear secuencia de modalidad de servicios
## Determinar etiquetas y códigos del alfabeto de la secuencia

sec3.labels <- c("GEN", "IND", "NA")
sec3.scode <- c("GEN", "IND", "NA")

head(cov)
ser.seq <- seqdef(cov, var=6:9, states=sec3.scode, labels=sec3.labels, gaps=NA, right=NA, left=NA, missing=FALSE,
missing.color="black")

## Sacar distribución de frecuencias de las secuencias
seqtab(ser.seq, tlim=1:10, weighted=TRUE, format="STS")
seqtab(ser.seq, tlim=1:100, weighted=TRUE, format="STS")

## Valorar los cambios entre modalidades como transiciones entre estados
## como indicador de niveles de movilidad o estabilidad en modalidades de servicio
cambio_ms <- seqtransn(ser.seq, pweight=FALSE)
table(cambio_ms)

## Los cambios son menores y se debe tomar en cuenta que los casos faltantes por salidas
## de la escuela aparecen como un cambio de modalidad, de la que estaba a NA.

# Determinar colores a los estados

cpal(ser.seq) <- c("steelblue", "tan4", "gray20")

# Gráficos
seqfplot(ser.seq, withlegend=TRUE, title="10 secuencias más frecuentes", cex.lab = 0.6, cex.main = 0.7, cex.
legend=0.6)
seqlplot(ser.seq, withlegend=TRUE, title="Frecuencias", sortv = "from.start", xlab="periodos", cex.lab=0.8, cex.main=0.8,
cex.legend=0.6)
seqdplot(ser.seq, withlegend=TRUE, title="Histograma", xlab="Periodos", cex.lab=0.7, cex.legend=0.6)

## Guardar gráficos como archivos .png
png("secuencia Servicios 10masF.png")
par(mfrow = c(1,1))
seqfplot(ser.seq, withlegend=TRUE, xlab="Periodos", cex.lab = 0.7, cex.main = 0.7, cex.legend=0.8)
dev.off()

png("Frecuencias Ind Servicios.png")
par(mfrow = c(1,1))
seqlplot(ser.seq, withlegend=TRUE, sortv = "from.start", xlab="Periodos", cex.lab=0.8, cex.main=0.8, cex.legend=0.8)
dev.off()
png("Histo secuencia Servicios.png")
par(mfrow = c(1,1))
seqdplot(ser.seq, withlegend=TRUE, xlab="Periodos", cex.lab=0.8, cex.main=0.8, cex.legend=0.8)
dev.off()

# Crear secuencia de grado cursado al inicio de cada periodo de registro desde 2008 a 2011

```

```

## Determinar etiquetas y códigos del alfabeto de la secuencia

sec2.labels <- c("1", "2", "3", "4")
sec2.scode <- c("1", "2", "3", "4")

año.seq <- seqdef(cov, var=10:13, states= sec2.scode, labels=sec2.labels, weights=cov$samplingweight, gaps=
NA, right=NA, left=NA, missing=NA, missing.color="black")

# Determinar colores a los estados

cpal(año.seq) <- c("darkolivegreen1", "darkolivegreen3", "darkolivegreen4", "darkolivegreen")
#Gráficos guardados como archivos formato .png

png("secuencia año 10masF.png")
par(mfrow = c(1,1))
seqfplot(año.seq, withlegend=TRUE, xlab="Periodos", cex.lab = 0.7, cex.main = 0.8, cex.legend=0.8)
dev.off()

png("secuencia año ind.png")
par(mfrow = c(1,1))
seqlplot(año.seq, withlegend=TRUE, sortv = "from.start", xlab="periodos",cex.lab=0.7,cex.main=0.8, cex.legend=0.8)
dev.off()

png("secuencia año histo.png")
par(mfrow = c(1,1))
seqdplot(año.seq, withlegend=TRUE, with.missing=TRUE, xlab="Periodos",cex.lab=0.8, cex.main=0.8, cex.
legend=0.8)
dev.off()
png("Entropía por Servicio.png")
seqplot.tentrop(año.seq, group=cov$servinif, title = "Entropía por modalidad de inicio", ylim = c(0, 1), cex = 0.7, cex.
lab = 0.4, cex.main = 0.6)
dev.off()

png("Índice de entropía por sexo.png")
par(mfrow = c(1,2))
seqplot.tentrop(año.seq, group=cov$sexof, title = "Sexo", ylim = c(0, 1), cex = 0.6, cex.lab = 0.4, cex.main = 0.6)
seqplot.tentrop(año.seq, group=cov$edadf, title = "Edad", ylim = c(0, 1), cex = 0.6, cex.lab = 0.4, cex.main = 0.6)
dev.off()

png("año_10masF por sexo.png")
seqfplot(año.seq, group = cov$sexof, title = "Sexo", cex = 1, cex.lab = 0.8, cex.main = 0.9, cex.legend = 0.8)
dev.off()
png("año_10masF por edad.png")
seqfplot(año.seq, group = cov$edadf, cex = 1, cex.lab = 0.8, cex.main = 0.8, cex.legend = 0.8)
dev.off()

# Distribución de secuencias en tabla
#Distribución de rangos de edad por sexo

table(cov$edad, cov$sexo)

## Extraer sub-bases por sexo para obtener frecuencias de la población total y por covariantes

table(cov$sexo)
dim(cov)
cov.H <- cov[cov$sexo == "H", ]
dim(cov.H)
cov.M <- cov[cov$sexo == "M", ]
dim(cov.M)
#Crear secuencias por sexo

añoH.seq <- seqdef(cov.H, var=10:13, states= sec2.scode, labels=sec2.labels, weights=cov.H$samplingweight,
gaps= NA, right=NA, left=NA, missing=NA, missing.color="black")
añoM.seq <- seqdef(cov.M, var=10:13, states= sec2.scode, labels=sec2.labels, weights=cov.M$samplingweight,
gaps= NA, right=NA, left=NA, missing=NA, missing.color="black")

```

```

#Distribución de frecuencias de secuencias por sexo con base expandida y sin expandir

seqtab(año.seq, tlim=1:100, weighted=TRUE, format="STS")
seqtab(año.seq, tlim=1:100, weighted=FALSE, format="STS")

seqtab(añoM.seq, tlim=1:100, weighted=TRUE, format="STS")
seqtab(añoM.seq, tlim=1:100, weighted=FALSE, format="STS")

seqtab(añoH.seq, tlim=1:100, weighted=TRUE, format="STS")
seqtab(añoH.seq, tlim=1:100, weighted=FALSE, format="STS")

seqtab(añoM.seq, tlim=1:10, weighted=TRUE, format="STS")
seqtab(añoM.seq, tlim=1:10, weighted=FALSE, format="STS")

## Frecuencia de secuencias por edad al inicio de la primaria

table(cov$edadf)

cov.edad <- cov[cov$edadf == "En edad", ]
dim(cov.edad)
cov.extra <- cov[cov$edadf == "Extraedad", ]
dim(cov.extra)

añoedad.seq <- seqdef(cov.edad, var=10:13, states= sec2.scode, labels=sec2.labels, gaps= NA, right=NA, left=NA,
missing=NA, missing.color="black")
añoextra.seq <- seqdef(cov.extra, var=10:13, states= sec2.scode, labels=sec2.labels, gaps= NA, right=NA, left=NA,
missing=NA, missing.color="black")

seqtab(añoedad.seq, tlim=1:100, weighted=FALSE, format="STS")
seqtab(añoextra.seq, tlim=1:100, weighted=FALSE, format="STS")
## Guardar frecuencias como archivo formato tabla

seqtab<- seqtab(año.seq, tlim=1:100, format="STS")
seqtab

write.table(seqtab, file = "seqtab.csv", append = FALSE, quote = TRUE, row.names = TRUE, col.names = TRUE)

# Calculo de duraciones por grado educativo
## Primero se calcula la distribución longitudinal (distribución de estado en cada secuencia)
## Se guardan estas distribuciones como base de datos

distlong<- data.frame(seqistatd(año.seq, with.missing=TRUE))

## Luego se crea objeto con lista de id para identificar las distlong en base de datos de individuos

id<- cov$id
head(id)
str(id)
id_distlong <- data.frame(cbind(id, distlong))

## Luego se pega id_distlong con base cov

cov_dl <- merge(id_distlong, cov, by="id", all=TRUE)
# Frecuencias de duraciones de 0 a 3 para cada grado educativo para población total y por covariantes
## 0: no haber alcanzado ese grado
## 1: haberlo cursado una vez
## 2: haberlo cursado dos veces
## 3: Haberlo cursado tres veces
## 4: Haberlo cursado cuatro veces

table(cov_dl$X1)
table(cov_dl$X2)
table(cov_dl$X3)
table(cov_dl$X4)
table(cov_dl$X.)

```

```

## Por Sexo
table(cov_d1$X1, cov_d1$sexo)
table(cov_d1$X2, cov_d1$sexo)
table(cov_d1$X3, cov_d1$sexo)
table(cov_d1$X4, cov_d1$sexo)

## Por Modalidad de Servicio al inicio de primaria
table(cov_d1$X1, cov_d1$servinif)
table(cov_d1$X2, cov_d1$servinif)
table(cov_d1$X3, cov_d1$servinif)
table(cov_d1$X4, cov_d1$servinif)

## Por Edad al inicio de primaria
table(cov_d1$X1, cov_d1$edadf)
table(cov_d1$X2, cov_d1$edadf)
table(cov_d1$X3, cov_d1$edadf)
table(cov_d1$X4, cov_d1$edadf)

# Cálculo de las transiciones realizadas
trans<- seqtransn(año.seq, pweight=FALSE)

## transiciones por covariantes
table(trans)
table(trans, cov$sexo)
table(trans, cov$edadf)
table(trans, cov$servinif)

# Cálculo de matriz de transiciones
## Arroja una matriz con tasas de transiciones entre estados a partir de las secuencias.
## Las tasas de transiciones son probabilidades de transición de un estado a otro observado en las secuencias
## Estas tasas fueron las utilizadas como matriz de sustitución de costos para estimar la matriz de distancia

trate<- seqtrate(año.seq, with.missing=TRUE, weighted=TRUE)
round(trate, 3)

## Estimar la matriz de transición por sexo

trate<- seqtrate(añoH.seq, with.missing=TRUE, weighted=TRUE)
round(trate, 3)

trate<- seqtrate(añoM.seq, with.missing=TRUE, weighted=TRUE)
round(trate, 3)

# Construir tipologías de trayectorias a partir de una matriz de costos, matriz de distancia y clúster
## Crear una matriz de costos (costos de sustitución)

subcosts<- "TRATE" #los costos de sustitución utilizan las tasas de transición entre estados
ccosts<- seqsubm(año.seq, method = subcosts, cval=2, with.miss=TRUE)

## Calculo de matriz de distancia por medio de Optimal Matching Analysis (OMA)

mat<- NULL
mat<- seqdist(año.seq, method = "OM", sm = ccosts, with.miss=TRUE)

## Ver la matriz de distancia interindividual (distancia entre trayectorias)

mat[1:10, 1:10]

## Aplicar cluster análisis a la matriz de distancia para producir tipos de secuencias
## con la función AGNES que es Agglomerative Nesting (Hierarchical Clustering)

año.clusterward<- agnes(mat, method = "ward", diss=TRUE)

#Guardar coo png dendograma

```

```

png("Dendo.png")
plot(año.clusterward, ask = F, which.plots = 2)
dev.off()

# Evaluar con una medida el número de grupos a seleccionar del clúster
## Ver distintas soluciones de cluster, de 2 a 4 grupos
### Para cada una de esas soluciones se estiman los índices de Silhouette individuales y promedios

año.clu.nclu_2 <- cutree(año.clusterward, k = 2)
año.clu.nclu_3 <- cutree(año.clusterward, k = 3)
año.clu.nclu_4 <- cutree(año.clusterward, k = 4)
año.clu.nclu_5 <- cutree(año.clusterward, k = 5)
año.clu.nclu_6 <- cutree(año.clusterward, k = 6)
año.clu.nclu_7 <- cutree(año.clusterward, k = 7)
año.clu.nclu_8 <- cutree(año.clusterward, k = 8)

#índice de Sil
si2 <- silhouette(año.clu.nclu_2, mat)
si3 <- silhouette(año.clu.nclu_3, mat)
si4 <- silhouette(año.clu.nclu_4, mat)
si5 <- silhouette(año.clu.nclu_5, mat)
si6 <- silhouette(año.clu.nclu_6, mat)
si7 <- silhouette(año.clu.nclu_7, mat)
si8 <- silhouette(año.clu.nclu_8, mat)

# Guardar como archivo .png los gráficos que arroja a técnica de Silhouette_
png("Silhouette_345.png")
par(mfrow = c(1,3))
plot(si3, cex.names=0.3)
plot(si4)
plot(si5)
dev.off()

png("Silhouette_678.png")
par(mfrow = c(1,3))
plot(si6)
plot(si7)
plot(si8)
dev.off()

## Guardar como objeto el valor del índice de cada trayectoria
si8 <- wcSilhouetteObs(mat, año.clu.nclu_8, measure="ASWw")
## Graficar las secuencias ordenadas por el silhouette width y guardar como archivo .png
###(los mejores clasificados de cada grupo quedan graficados arriba)

png("TI_sil8.png", width = 600, height = 600, units = "px")
par(mfcol=c(2,3), mar=c(2.2,2,0.9,0.9))
seqplot(año.seq, group=año.clu.nclu_8, sortv=sil8)
dev.off()

## Obtener el árbol de distribución de grupos

st<- as.seqtree(año.clusterward, seqdata=año.seq, diss=mat, weighted=FALSE, ncluster=8)
print(st)
st<- as.seqtree(año.clusterward, seqdata=año.seq, diss=mat, weighted=TRUE, ncluster=8)
print(st)
### Dado el análisis a partir de los índice de Silhouette se decide quedarse con 8 grupos
### Guardar como gráfica en formato .png las trayectorias individuales ordenas por el inicio común para cada grupo
png("TI_8.png", width = 600, height = 600, units = "px")
par(mfcol=c(2,3), mar=c(2.2,2,0.9,0.9))
seqplot(año.seq, group=año.clu.nclu_8, sortv="from.start")
dev.off()

## Gráficar y guardar como .png los histogramas de cada grupo
png("Histo_8.png", width = 600, height = 600, units = "px")

```



```

par(mfcol=c(2,3), mar=c(2.2,2,0.9,0.9))
seqdplot(año.seq, group=año.clu.nclu_8, with.missing=TRUE)
dev.off()

## Evaluar composición de los 8 grupos del clúster
## Pegar identificador de clúster en base individuos

id_clu <- data.frame(cbind(id, año.clu.nclu_8))
## Pegar en la base de las trayectorias individuales el grupo de cluster al que pertenece

cov_i_clu <- merge(id_clu, cov, by="id", all=TRUE)
head(cov_i_clu)
dim(cov_i_clu)

## Distribución por sexo, edad y modalidad servicio en cada grupo de trayectorias

cov_i_clu$stray_f <- factor(cov_i_clu$año.clu.nclu_8, levels=1:8, labels=c("Tipo 1", "Tipo 2", "Tipo 3", "Tipo 4", "Tipo 5",
"Tipo 6", "Tipo 7", "Tipo 8"))

table(cov_i_clu$stray_f, cov_i_clu$año.clu.nclu_8)
table(cov_i_clu$stray_f)
table_sexo <- table(cov_i_clu$sexof, cov_i_clu$stray_f)
table_edad <- table(cov_i_clu$edadf, cov_i_clu$stray_f)
table_servinif <- table(cov_i_clu$servinif, cov_i_clu$stray_f)
table3 <- table(cov_i_clu$edadf, cov_i_clu$stray_f, cov_i_clu$sexof)
prop.table(table_sexo, 2)
prop.table(table_edad, 2)
prop.table(table_servinif, 2)

## Guardar como archivo csv todo lo anterior

write.table(table_sexo, file = "sex.csv", sep = ",", col.names = NA, qmethod = "double")
write.table(table_edad, file = "edad.csv", sep = ",", col.names = NA, qmethod = "double")
write.table(table_servinif, file = "servini.csv", sep = ",", col.names = NA, qmethod = "double")

```

2 Cohorte secundaria 2008-2009

Código de análisis de secuencia completo base muestra de Cohorte Secundaria 2008/2009

Karina Videgain

Noviembre 2014

```

# Se abre la base de datos muestra para la cohorte que inicia secundaria en 2008/2009
##Aclaración: habrá que modificar la ruta de acceso según su lugar en el disco.
secu<- read.dta("C:/INEE /Bases/muestra_secundaria.dta")

# Identificación de casos duplicados en la muestra

anyDuplicated(secu$id)

secu[-c(3563), ]

## Eliminar el caso duplicado y por eso de ahora en más la muestra tendrá 4999.
secu<- secu[-c(3563), ]

dim(secu)
anyDuplicated(secu$id)

# Establecer dónde se guardaran los resultados, bases y demás objetos
##Aclaración: habrá que modificar la ruta de acceso según su lugar en el disco.

```

```

setwd("C:/ INEE /Resultados//Muestra Secundaria")
getwd()

# Creación de variables de interés
names(secu)
str(secu)

# Crear variable sexo

table(secu$sexo_i07)
## Hay casos sin declaración de sexo en la base de inicio de ciclo 2007
## Completar casos faltantes con información de otros ciclos

secu$sexo[secu$sexo_i07=="H" | secu$sexo_f07=="H" | secu$sexo_i08=="H" | secu$sexo_f08=="H" | secu$sexo_i09=="H" | secu$sexo_f09=="H" | secu$sexo_i10=="H" | secu$sexo_f10=="H" | secu$sexo_i11=="H"] <- "H"
secu$sexo[secu$sexo_i07=="M" | secu$sexo_f07=="M" | secu$sexo_i08=="M" | secu$sexo_f08=="M" | secu$sexo_i09=="M" | secu$sexo_f09=="M" | secu$sexo_i10=="M" | secu$sexo_f10=="M" | secu$sexo_i11=="M"] <- "M"
table(secu$sexo)
which(is.na(secu$sexo))
## Se recupera información para todos los casos

class(secu$fecha_nac_i07)

## Declarar las variables de fecha de nacimiento como variables "date"

secu$nac_i07 <- as.Date(secu$fecha_nac_i07, "%d/%m/%Y")
secu$nac_f07 <- as.Date(secu$fecha_nac_f07, "%d/%m/%Y")
secu$nac_i08 <- as.Date(secu$fecha_nac_i08, "%d/%m/%Y")
secu$nac_f08 <- as.Date(secu$fecha_nac_f08, "%d/%m/%Y")
secu$nac_i09 <- as.Date(secu$fecha_nac_i09, "%d/%m/%Y")
secu$nac_f09 <- as.Date(secu$fecha_nac_f09, "%d/%m/%Y")
secu$nac_i10 <- as.Date(secu$fecha_nac_i10, "%d/%m/%Y")
secu$nac_f10 <- as.Date(secu$fecha_nac_f10, "%d/%m/%Y")
secu$nac_i11 <- as.Date(secu$fecha_nac_i11, "%d/%m/%Y")

class(secu$nac_i07)

## Ver las nuevas variables en los 6 primeros casos de la base

head(secu$nac_i07)
head(secu$nac_i08)
head(secu$nac_i09)
head(secu$nac_i10)

## Hay mucha información faltante, por lo que se extraerá la información de la fecha contenida en variable ID para considerarla.

# Crear una variable fecha de nacimiento a partir de la ID

secu$nac <- (substr(secu$id, start=5, stop=10))

class(secu$nac)
head(secu$nac)
secu$fnac <- as.Date(secu$nac, "%y%m%d")
head(secu$fnac)
class(secu$fnac)
summary(secu$fnac)
## Arroja valores más acertados que el registro de las fechas de cada periodo de registro de RNA y no presenta casos perdidos

# Crear variable que dé cuenta si está en un rango de edad indicado para el grado cursado
## En el ciclo 2008/2009 deben iniciar primero de secundaria (séptimo grado) aquellos que hayan cumplido 12 años al ### 31 de diciembre del 2008,
## o a sus efectos 11 años en el fin del ciclo primario,
## que es cuando se inicia la observación de la secuencia.

```

```

## En edad (1): aquellos que cumplían 12 años entre 01/01/96 y 31/12/96
## Extradad (2): Nacidos antes del 01/01/96
## Menor edad (0): Nacidos después del 31/12/96

```

```

secu$edad <- "1"
secu$edad[secu$fnac < "1996-01-01"] <- "2"
secu$edad[secu$fnac > "1996-12-31"] <- "0"

```

```

# En esta base de datos muestral proporcionada por el INEE ya están
# las variables de servicio, nivel, grado y secuencias creadas con mis códigos anteriores.

```

```

#servicio: COM = 1
#             GEN = 2
#             IND = 3
#             TEC = 4
#             TEL = 5
#             TRA = 6
#             MIG = 5
#             " " = NA

```

```

# Ver si hay datos faltantes en el proceso que se pueda recuperar al tener la información organizada en forma
longitudinal

```

```

# Hay datos faltantes en inicio que puede ser completado con servicio al final del mismo periodo

```

```

table(secu$ser_i1)
table(secu$ser_f1)
table(secu$ser_i1, secu$ser_f1)

```

```

secu$se_i1 <- secu$ser_i1
secu$se_i1[secu$ser_i1 == "NA" & secu$ser_f1 == "1"] <- "1"
secu$se_i1[secu$ser_i1 == "NA" & secu$ser_f1 == "2"] <- "2"
secu$se_i1[secu$ser_i1 == "NA" & secu$ser_f1 == "3"] <- "3"
table(secu$ser_i1, secu$se_i1)
table(secu$se_i1)

```

```

table(secu$ser_i2)
table(secu$ser_f2)
table(secu$ser_i2, secu$ser_f2)

```

```

secu$se_i2 <- secu$ser_i2
secu$se_i2[secu$ser_i2 == "NA" & secu$ser_f2 == "2"] <- "2"
secu$se_i2[secu$ser_i2 == "NA" & secu$ser_f2 == "4"] <- "4"
secu$se_i2[secu$ser_i2 == "NA" & secu$ser_f2 == "5"] <- "5"
secu$se_i2[secu$ser_i2 == "NA" & secu$ser_f2 == "6"] <- "6"

```

```

table(secu$ser_i2, secu$se_i2)
table(secu$se_i2)

```

```

table(secu$ser_i3)
table(secu$ser_f3)
table(secu$ser_i3, secu$ser_f3)

```

```

secu$se_i3 <- secu$ser_i3
secu$se_i3[secu$ser_i3 == "NA" & secu$ser_f3 == "2"] <- "2"
secu$se_i3[secu$ser_i3 == "NA" & secu$ser_f3 == "4"] <- "4"
secu$se_i3[secu$ser_i3 == "NA" & secu$ser_f3 == "5"] <- "5"
secu$se_i3[secu$ser_i3 == "NA" & secu$ser_f3 == "6"] <- "6"

```

```

table(secu$ser_i3, secu$se_i3)
table(secu$se_i3)

```

```

table(secu$ser_i4)
table(secu$ser_f4)
table(secu$ser_i4, secu$ser_f4)

```

```

secu$se_i4 <- secu$ser_i4
secu$se_i4[secu$ser_i4 == "NA" & secu$ser_f4 == "2" ] <- "2"
secu$se_i4[secu$ser_i4 == "NA" & secu$ser_f4 == "4" ] <- "4"
secu$se_i4[secu$ser_i4 == "NA" & secu$ser_f4 == "5" ] <- "5"
secu$se_i4[secu$ser_i4 == "NA" & secu$ser_f4 == "6" ] <- "6"

```

```

table(secu$ser_i4, secu$se_i4)
table(secu$se_i4)

```

```

table(secu$ser_i5)
secu$se_i5 <- secu$ser_i5
table(secu$se_i5)

```

```

table(secu$se_i1)
table(secu$se_i2)
table(secu$se_i3)
table(secu$se_i4)
table(secu$se_i5)

```

```

## Con esta recodificación se recuperaron información faltante, sobre toda aquella
## que respondía a falta de registro y que con registros subsiguientes de trayectoria educativa
## se pudo inferir y completar.
## Quedan igual casos faltantes en períodos de registro que dan cuenta de
## aquellos casos de interrupciones en el proceso educativo.

```

```

# Verificar distribución de variable que informa del grado escolar cursado en cada período de registro

```

```

table(secu$año_i1)
table(secu$año_f1)
table(secu$año_i2)
table(secu$año_f2)
table(secu$año_i3)
table(secu$año_f3)
table(secu$año_i4)
table(secu$año_f4)
table(secu$año_i5)

```

```

## Se encuentran información faltante respecto a años intermedios
## que puede ser completada con la información del año anterior y el posterior.

```

```

## A saber, se corrige con la siguiente codificación:

```

```

table(secu$año_i1)
secu$añ_i1 <- secu$año_i1
table(secu$año_i1, secu$añ_i1)

```

```

table(secu$año_i2)
table(secu$año_f2)
table(secu$año_i2, secu$año_f2)
secu$añ_i2 <- secu$año_i2
secu$añ_i2[secu$año_i2 == "NA" & secu$año_i1 == "6" & secu$año_i3 == "8" ] <- "7"
secu$añ_i2[secu$año_i2 == "NA" & secu$año_i1 == "NA" & secu$año_f2 == "7" ] <- "7"
secu$añ_i2[secu$año_i2 == "NA" & secu$año_i1 == "6" & secu$año_i4 == "9" ] <- "7"
secu$añ_i2[secu$año_i2 == "NA" & secu$año_i1 == "6" & secu$año_i3 == "NA" & secu$año_f3 == "8" ] <- "7"
secu$añ_i2[secu$año_i2 == "NA" & secu$año_f2 == "7" & secu$año_i3 == "NA" ] <- "7"
secu$añ_i2[secu$año_i2 == "NA" & secu$año_f2 == "7" & secu$año_i3 == "8" ] <- "7"
secu$añ_i2[secu$año_i2 == "NA" & secu$año_f2 == "NA" & secu$año_i3 == "8" ] <- "7"
table(secu$añ_i2)
table(secu$año_i2, secu$añ_i2)

```

```

table(secu$año_i3)
table(secu$año_f3)
table(secu$año_i3, secu$año_f3)
secu$añ_i3 <- secu$año_i3
secu$añ_i3[secu$año_i3 == "NA" & secu$año_i2 == "7" & secu$año_i4 == "9" ] <- "8"
secu$añ_i3[secu$año_i3 == "NA" & secu$año_i2 == "NA" & secu$año_f2 == "7" & secu$año_i4 == "9" ] <- "8"

```

```

secu$añ_i3[secu$año_i3 == "NA" & secu$año_i1 == "6" & secu$año_f3 == "8" & secu$año_i4 == "NA" ] <- "8"
secu$añ_i3[secu$año_i3 == "NA" & secu$año_f3 == "8" & secu$año_i4 == "NA" ] <- "8"
secu$añ_i3[secu$año_i3 == "NA" & secu$año_f3 == "NA" & secu$año_i4 == "9" & secu$año_i2 == "7" ] <- "8"
secu$añ_i3[secu$año_i3 == "NA" & secu$año_i1 == "6" & secu$año_i4 == "9" ] <- "8"
secu$añ_i3[secu$año_i3 == "7" & secu$año_i2 == "7" & secu$año_f3 == "8" & secu$año_i4 == "9" ] <- "8"
secu$añ_i3[secu$año_i3 == "7" & secu$año_i2 == "7" & secu$año_f3 == "8" & secu$año_i4 == "7" ] <- "7"
secu$añ_i3[secu$año_i3 == "8" & secu$año_i2 == "7" & secu$año_f3 == "7" & secu$año_i4 == "7" ] <- "7"
table(secu$año_i3, secu$añ_i3)
table(secu$añ_i3)

table(secu$año_i4)
table(secu$año_f4)
secu$añ_i4 <- secu$año_i4
secu$añ_i4[secu$año_i4 == "NA" & secu$año_i3 == "8" & secu$año_i5 == "9" ] <- "9"
secu$añ_i4[secu$año_i4 == "NA" & secu$año_i3 == "8" & secu$año_f4 == "9" & secu$año_i5 == "NA" ] <- "9"
secu$añ_i4[secu$año_i4 == "NA" & secu$año_i3 == "8" & secu$año_f4 == "9" & secu$año_i5 == "9" ] <- "9"
secu$añ_i4[secu$año_i4 == "8" & secu$año_i3 == "8" & secu$año_f4 == "9" ] <- "9"
secu$añ_i4[secu$año_i4 == "9" & secu$año_i3 == "8" & secu$año_f4 == "8" ] <- "8"
secu$añ_i4[secu$año_i4 == "9" & secu$año_i3 == "7" & secu$año_f4 == "8" ] <- "8"
table(secu$año_i4, secu$añ_i4)

secu$añ_i5 <- secu$año_i5

table(secu$añ_i1)
table(secu$añ_i2)
table(secu$añ_i3)
table(secu$añ_i4)
table(secu$añ_i5)

# Re-crear la variable que informa de año y modalidad de servicio (i1 a i5) con las variables input corregidas

table(secu$i1)
table(secu$f1)
table(secu$i2)
table(secu$f2)
table(secu$i3)
table(secu$f3)
table(secu$i4)
table(secu$f4)
table(secu$i5)
### Crear estado por periodo que combine año y modalidad de servicio del centro

secu$as_i1 <- paste(secu$añ_i1, secu$se_i1)
secu$as_i1[secu$añ_i1 == "NA" ] <- "NA_NA"
table(secu$i1)
table(secu$as_i1)

secu$as_i2 <- paste(secu$añ_i2, secu$se_i2)
secu$as_i2[secu$añ_i2 == "NA" ] <- "NA_NA"
table(secu$i2)
table(secu$as_i2)

secu$as_i3 <- paste(secu$añ_i3, secu$se_i3)
secu$as_i3[secu$añ_i3 == "NA" ] <- "NA_NA"
table(secu$i3)
table(secu$as_i3)
secu$as_i4 <- paste(secu$añ_i4, secu$se_i4)
secu$as_i4[secu$añ_i4 == "NA" ] <- "NA_NA"
table(secu$i4)
table(secu$as_i4)

secu$as_i5 <- paste(secu$añ_i5, secu$se_i5)
secu$as_i5[secu$añ_i5 == "NA" ] <- "NA_NA"
table(secu$i5)
table(secu$as_i5)

```

```

# Generar variable con secuencia de modalidades para ver la presencia de cambios intermodalidad

secu$servicios <- paste(secu$se_i1, secu$se_i2, secu$se_i3, secu$se_i4)
tabla<- table(secu$servicios)
ser<- prop.table(tabla)
ser
#Convertir y guardar el resultado en archivo formato tabla de excel

write.table(ser, file = "servicio.csv", append = FALSE, quote = TRUE, row.names = TRUE, col.names = TRUE)

# Generar variable que informe a la modalidad e servicio que entró en su primer año de secundaria

secu$serv_inicio <- secu$se_i2

# Distribución de movimientos entre modalidades de servicio
## Pasaje de primaria a secundaria

table(secu$se_i1)

with(secu, table(secu$se_i1, secu$serv_inicio)) -> serv0.serv1.table
prop.table(serv0.serv1.table, 1) # With respect to row marginal sums.

## Pasaje de primer año de secundaria a segundo año secundaria

table(secu$se_i2)
with(secu, table(secu$serv_inicio, secu$se_i3)) -> serv1.serv2.table
prop.table(serv1.serv2.table, 1) # With respect to row marginal sums.

## Distribucion de secuencias de modalidades desde el inicio a los años de secundaria siguiente

secu$serv_secu <- paste(secu$se_i2, secu$se_i3, secu$se_i4)
tabla.servsercu <- table(secu$serv_secu)
tabla.servsercu

## Distribución de modalidades de inicio de secundaria por sexo

tabla.servinicio.sex <- table(secu$serv_inicio, secu$sexo)
tabla.servinicio.sex

prop.table(tabla.servinicio.sex, 1)
prop.table(tabla.servinicio.sex, 2)

# Generar base igual a la muestra con todos los casos pero sólo con variables de interés para el análisis (covariantes)
## Antes cambio nombre a variables para que las etiquetas que use en gráficos sean más evidentes respecto a lo
que ####representan

secu$ms_ag07 <- secu$se_i1
secu$ms_ag08 <- secu$se_i2
secu$ms_ag09 <- secu$se_i3
secu$ms_ag10 <- secu$se_i4
secu$ms_ag11 <- secu$se_i5

secu$gr_ag07 <- secu$añ_i1
secu$gr_ag08 <- secu$añ_i2
secu$gr_ag09 <- secu$añ_i3
secu$gr_ag10 <- secu$añ_i4
secu$gr_ag11 <- secu$añ_i5

secu$am_ag07 <- secu$as_i1
secu$am_ag08 <- secu$as_i2
secu$am_ag09 <- secu$as_i3
secu$am_ag10 <- secu$as_i4
secu$am_ag11 <- secu$as_i5

```

```

cov <- secu[,c("id","sexo", "edad", "serv_inicio", "samplingweight", "ms_ag07", "ms_ag08", "ms_ag09", "ms_ag10",
"ms_ag11", "gr_ag07", "gr_ag08", "gr_ag09", "gr_ag10", "gr_ag11", "am_ag07", "am_ag08", "am_ag09", "am_ag10",
"am_ag11")]
head(cov)
dim(cov)

# Generación de variables como factor: sexo, edad, modalidad servicio al inicio

table(cov$edad)
cov$edadf <- factor(cov$edad, levels=0:2, labels=c("Menor", "En edad", "Extraedad"))
table(cov$edadf)

table(secu$sexo)
cov$sexof <- factor(cov$sexo, labels=c("Varones", "Mujeres"))
table(cov$sexof)

table(cov$serv_inicio)
cov$servinif <- factor(cov$serv_inicio, labels=c("General", "Técnica", "Telesecundaria", "Trabajo", "NA"))
table(cov$servinif)

table(cov$sexof, cov$edadf)

# Crear secuencia de modalidad de servicios
## Determino etiquetas y códigos del alfabeto de la secuencia

sec3.labels <- c("COM", "GEN", "IND", "TEC", "TEL", "TRA", "NA")
sec3.scode <- c("COM", "GEN", "IND", "TEC", "TEL", "TRA", "NA")
head(cov)
ser.seq <- seqdef(cov, var=6:10, states= sec3.scode, labels=sec3.labels, gaps= NA, right=NA, left=NA,
missing=FALSE, missing.color="black")
head(ser.seq)

seqtab(ser.seq, tlim=1:10, weighted=TRUE, format="STS")
seqtab(ser.seq, tlim=1:100, weighted=TRUE, format="STS")

# Determinar colores a los estados

cpal(ser.seq) <- c("green", "steelblue", "tan4", "orangered", "violetred", "yellow3", "gray20")

# Graficas

seqfplot(ser.seq, withlegend=TRUE, title="10 secuencias más frecuentes", cex.lab = 0.6, cex.main = 0.7, cex.
legend=0.6)
seqlplot(ser.seq, withlegend=TRUE, title="Frecuencias", sortv = "from.start", xlab="periodos", cex.lab=0.8, cex.main=0.8,
cex.legend=0.6)
seqdplot(ser.seq, withlegend=TRUE, title="Histograma", xlab="Periodos", cex.lab=0.7, cex.legend=0.6)

# Guardar gráficas como archivos con formato .png

png("Secuencia Servicios 10masF.png")
par(mfrow = c(1,1))
seqfplot(ser.seq, withlegend=TRUE, xlab="Periodos", cex.lab = 0.7, cex.main = 0.7, cex.legend=0.8)
dev.off()

png("Frecuencias Ind Servicios.png")
par(mfrow = c(1,1))
seqlplot(ser.seq, withlegend=TRUE, sortv = "from.start", xlab="Periodos", cex.lab=0.8, cex.main=0.8, cex.legend=0.8)
dev.off()

png("Histo Secuencia Servicios.png")
par(mfrow = c(1,1))
seqdplot(ser.seq, withlegend=TRUE, xlab="Periodos", cex.lab=0.8, cex.main=0.8, cex.legend=0.8)
dev.off()

# Creo secuencia de grado al inicio de cada ciclo

```

```

## Determino etiquetas y codigos del alfabeto de la secuencia
sec2.labels <- c("6", "7", "8", "9")
sec2.scode <- c("6", "7", "8", "9")

año.seq <- seqdef(cov, var=11:15, states= sec2.scode, labels=sec2.labels, weights=cov$samplingweight, gaps=
NA, right=NA, left=NA, missing=NA, missing.color="gray27")

# Determino colores a los estados

cpal(año.seq) <- c("darkolivegreen3", "skyblue3", "steelblue", "royalblue4")
# Guardar gráficas como archivos con formato .png

png("Secuencia año 10masF.png")
par(mfrow = c(1,1))
seqfplot(año.seq, withlegend=TRUE, xlab="Periodos", cex.lab = 0.8, cex.main = 0.8, cex.legend=0.8)
dev.off()

png("Secuencia año ind.png")
par(mfrow = c(1,1))
seqfplot(año.seq, withlegend=TRUE, sortv = "from.start", xlab="periodos",cex.lab=0.8,cex.main=0.8, cex.legend=0.8)
dev.off()

png("Secuencia año histo.png")
par(mfrow = c(1,1))
seqdplot(año.seq, withlegend=TRUE, with.missing=TRUE, xlab="Periodos",cex.lab=0.8, cex.main=0.8, cex.
legend=0.8)
dev.off()

#graficas de secuencias por covariantes

seqplot.tentrop(año.seq, group=cov$servinif, title = "E x modalidad de inicio", ylim = c(0, 1), cex = 0.7, cex.lab = 0.4,
cex.main = 0.6)

png("Índice de entropía por sexo.png")
par(mfrow = c(1,2))
seqplot.tentrop(año.seq, group=cov$sexof, title = "Sexo", ylim = c(0, 1), cex = 0.6, cex.lab = 0.4, cex.main = 0.6)
seqplot.tentrop(año.seq, group=cov$edadf, title = "Edad", ylim = c(0, 1), cex = 0.6, cex.lab = 0.4, cex.main = 0.6)
dev.off()

png("año_10masF por sexo.png")
seqfplot(año.seq, group = cov$sexof, title = "Sexo", cex = 1, cex.lab = 0.8, cex.main = 0.8, cex.legend = 0.8)
dev.off()

png("año_10masF por edad.png")
seqfplot(año.seq, group = cov$edadf, cex = 1, cex.lab = 0.9, cex.main = 0.9, cex.legend = 0.8)
dev.off()

# Distribución de secuencias en tabla
##distribución de rangos de edad por sexo

table(cov$edad, cov$sexo)

## Extraer sub-bases por sexo para obtener frecuencias de la población total y por covariantes

table(cov$sexo)
dim(cov)
cov.H <- cov[cov$sexo == "H", ]
dim(cov.H)
cov.M <- cov[cov$sexo == "M", ]
dim(cov.M)
head(cov)
añoH.seq <- seqdef(cov.H, var=11:15, states= sec2.scode, labels=sec2.labels, weights=cov.H$samplingweight,
gaps= NA, right=NA, left=NA, missing=NA, missing.color="gray27")
añoM.seq <- seqdef(cov.M, var=11:15, states= sec2.scode, labels=sec2.labels, weights=cov.M$samplingweight,
gaps= NA, right=NA, left=NA, missing=NA, missing.color="gray27")

```



```

# Frecuencias de la población total y por covariantes

seqtab(año.seq, tlim=1:100, weighted=TRUE, format="STS")
seqtab(año.seq, tlim=1:100, weighted=FALSE, format="STS")

seqtab(añoM.seq, tlim=1:100, weighted=TRUE, format="STS")
seqtab(añoM.seq, tlim=1:100, weighted=FALSE, format="STS")

seqtab(añoH.seq, tlim=1:100, weighted=TRUE, format="STS")
seqtab(añoH.seq, tlim=1:100, weighted=FALSE, format="STS")

seqtab(añoM.seq, tlim=1:10, weighted=TRUE, format="STS")
seqtab(añoM.seq, tlim=1:10, weighted=FALSE, format="STS")

## Frecuencia de secuencias por edad al inicio de la secundaria
table(cov$edadf)

dim(cov)
cov.menor <- cov[cov$edadf == "Menor", ]
dim(cov.menor)
cov.edad <- cov[cov$edadf == "En edad", ]
dim(cov.edad)
cov.extra <- cov[cov$edadf == "Extraedad", ]
dim(cov.extra)

año.menor.seq <- seqdef(cov.menor, var=1:15, states= sec2.scode, labels=sec2.labels, gaps= NA, right=NA,
left=NA, missing=NA, missing.color="gray27")
año.edad.seq <- seqdef(cov.edad, var=1:15, states= sec2.scode, labels=sec2.labels, gaps= NA, right=NA, left=NA,
missing=NA, missing.color="gray27")
año.extra.seq <- seqdef(cov.extra, var=1:15, states= sec2.scode, labels=sec2.labels, gaps= NA, right=NA, left=NA,
missing=NA, missing.color="gray27")

seqtab(año.menor.seq, tlim=1:100, weighted=FALSE, format="STS")
seqtab(año.edad.seq, tlim=1:100, weighted=FALSE, format="STS")
seqtab(año.extra.seq, tlim=1:100, weighted=FALSE, format="STS")

## Guardar frecuencias como archivo en formato tabla

seqtab <- seqtab(año.seq, tlim=1:100, format="STS")

write.table(seqtab, file = "seqtab.csv", append = FALSE, quote = TRUE, row.names = TRUE, col.names = TRUE)

# Calculo de duraciones por grado educativo
## Primero se calcula la distribución longitudinal (distribución de estado en cada secuencia)
## Se guardan estas distribuciones como base de datos

distlong <- data.frame(seqistatd(año.seq, with.missing=TRUE))

## Crear objeto con lista de id para identificar las distlong en base de datos de individuos

id <- cov$id
head(id)
str(id)

id_distlong <- data.frame(cbind(id, distlong))

## Pegar id_distlong y luego a base cov
cov_dl <- merge(id_distlong, cov, by="id", all=TRUE)

dim(cov_dl)
head(cov_dl, 20)
summary(cov_dl)

```

```

#Frecuencias de duraciones de 0 a 3
## 0: no haber alcanzado ese grado
## 1: haberlo cursado una vez
## 2: haberlo cursado dos veces
## 3: Haberlo cursado tres veces

table(cov_d1$X7)
table(cov_d1$X8)
table(cov_d1$X9)

## Por Sexo
table(cov_d1$X7,cov_d1$sexo)
table(cov_d1$X8, cov_d1$sexo)
table(cov_d1$X9, cov_d1$sexo)

## Por Modalidad de Servicio al inicio de secundaria

table(cov_d1$X7, cov_d1$servinif)
table(cov_d1$X8, cov_d1$servinif)
table(cov_d1$X9, cov_d1$servinif)

## Por Edad al inicio de secundaria

table(cov_d1$X7,cov_d1$edadf)
table(cov_d1$X8, cov_d1$edadf)
table(cov_d1$X9, cov_d1$edadf)

# Cálculo de las transiciones realizadas

trans<- seqtransn(año.seq, pweight=FALSE)

## transiciones por covariantes

table(trans)
table(trans, cov$sexo)
table(trans, cov$edadf)
table(trans, cov$servinif)

# Cálculo de matriz de transiciones
## Arroja una matriz con tasas de transiciones entre estados a partir de las secuencias.
## Las tasas de transiciones son probabilidades de transición de un estado a otro observado en las secuencias
## Estas tasas fueron las utilizadas como matriz de sustitución de costos para estimar la matriz de distancia

trate<- seqtrate(año.seq, with.missing=TRUE, weighted=TRUE)
round(trate, 3)

## Estimar la matriz de transición por sexo

trate<- seqtrate(añoH.seq, with.missing=TRUE, weighted=TRUE)
round(trate, 3)
trate<- seqtrate(añoM.seq, with.missing=TRUE, weighted=TRUE)
round(trate, 3)

## Estimar la matriz de transición por edad
trate<- seqtrate(añoedad.seq, with.missing=TRUE, weighted=TRUE)
round(trate, 3)

# Construir tipologías de trayectorias a partir de una matriz de costos, matriz de distancia y clúster
## Crear una matriz de costos (costos de sustitución)

subcosts<- "TRATE" #los costos de sustitución utilizan las tasas de transición entre estados
ccosts<- seqsubm(año.seq, method = subcosts, cval=2, with.miss=TRUE)
ccosts

## Calculo de matriz de distancia por medio de Optimal Matching Analysis (OMA)

```

```

mat<- NULL
mat<- seqdist(año.seq, method = "OM", sm = ccosts, with.miss=TRUE)

## Ver la matriz de distancia interindividual (distancia entre trayectorias)

mat[1:10, 1:10]

## Aplicar clúster análisis a la matriz de distancia para producir tipos de secuencias
## con función AGNES que es Agglomerative Nesting (Hierarchical Clustering)

año.clusterward <- agnes(mat, method = "ward", diss=TRUE)

# Guardar como archivo .png el dendograma

png("Dendo.png")
plot(año.clusterward, ask = F, which.plots = 2)
dev.off()

# Evaluar con una medida el número de grupos a seleccionar del clúster

## Ahora vamos a ver distintas soluciones de clúster, de 2 a 4 grupos
## Para cada una de esas soluciones podremos estimar los índices de
## silhouette individuales y promedios

año.clu.nclu_2 <- cutree(año.clusterward, k = 2)
año.clu.nclu_3 <- cutree(año.clusterward, k = 3)
año.clu.nclu_4 <- cutree(año.clusterward, k = 4)
año.clu.nclu_5 <- cutree(año.clusterward, k = 5)
año.clu.nclu_6 <- cutree(año.clusterward, k = 6)

#índice de Sil
si2 <- silhouette(año.clu.nclu_2, mat)
si3 <- silhouette(año.clu.nclu_3, mat)
si4 <- silhouette(año.clu.nclu_4, mat)
si5 <- silhouette(año.clu.nclu_5, mat)
si6 <- silhouette(año.clu.nclu_6, mat)

# Guardar como archivos .png gráficos que arroja la técnica de Silhouette

png("Silhouette_456.png")
par(mfrow = c(1,3))
plot(si4)
plot(si5)
plot(si6)
dev.off()

## Guardo como objeto el valor del índice de cada trayectoria

sil<- wcSilhouetteObs(mat, año.clu.nclu_5, measure="ASWw")

## Guardar como archivo .png el gráfico de secuencias individuales ordenadas por el silhouette width y
## los mejores clasificados de cada grupo quedan graficados arriba
png("Sil5.png")
seqplot(año.seq, group=año.clu.nclu_5, sortv=sil)
dev.off()

## Dado el análisis a partir de los índice de Silhouette se decide quedar con 5 grupos

nclu<- 5

## Obtener el árbol de distribución de grupos
st<- as.seqtree(año.clusterward, seqdata=año.seq, diss=mat, weighted=FALSE, ncluster=5)
print(st)
st<- as.seqtree(año.clusterward, seqdata=año.seq, diss=mat, weighted=TRUE, ncluster=5)

```

```

print(st)
seqtreedisplay(tree=st, type="l", border=NA)

# Corto el dendrograma para crear el número deseado de grupos
año.clu.nclu <- cutree(año.clusterward, k = nclu)

# Guardar gráficas por tipologías como archivo .png
png("Tray individuales_5.png")
par(mfcol=c(2,3), mar=c(2.2,2,0.9,0.9))
seqplot(año.seq, group=año.clu.nclu_5, sortv="from.start", xlab="Periodos",cex.lab=0.8,cex.main=0.9)
dev.off()

png("Histograma_5.png")
par(mfcol=c(2,3), mar=c(2.2,2,0.9,0.9))
seqplot(año.seq, group=año.clu.nclu_5, with.missing=TRUE, xlab="Periodos",cex.lab=0.8,cex.main=0.9)
dev.off()

# Evaluar composición de los 5 grupos del clúster
seqtab(año.seq, group=año.clu.nclu_5)

## Pegar identificador de clúster en base individuos
id_clu <- data.frame(cbind(id, año.clu.nclu))

## Pegar en la base de las trayectorias individuales el grupo de clúster al que pertenece
head(cov)
dim(cov)
cov_i_clu <- merge(id_clu, cov, by="id", all=TRUE)

## Distribución por sexo, edad y modalidad servicio en cada grupo de trayectorias
cov_i_clu$stray_f <- factor(cov_i$año.clu.nclu, levels=1:5, labels=c("Tipo 1", "Tipo 2", "Tipo 3", "Tipo 4", "Tipo 5"))

table(cov_i_clu$stray_f, cov_i_clu$año.clu.nclu)
table_sexo <- table(cov_i_clu$sexof, cov_i_clu$stray_f)
table_edad <- table(cov_i_clu$edadf, cov_i_clu$stray_f)
table_servinif <- table(cov_i_clu$servinif, cov_i_clu$stray_f)
prop.table(table_sexo, 2)
prop.table(table_edad, 2)
prop.table(table_servinif, 2)

## Guardar como archivo csv todo lo anterior
write.table(table_sexo, file = "sex.csv", sep = ",", col.names = NA, qmethod = "double")
write.table(table_edad, file = "edad.csv", sep = ",", col.names = NA, qmethod = "double")
write.table(table_servinif, file = "servini.csv", sep = ",", col.names = NA, qmethod = "double")

```

DIRECTORIO

Junta de Gobierno

Sylvia Irene Schmelkes del Valle
Consejera Presidenta

Margarita María Zorrilla Fierro
Consejera

Teresa Bracho González
Consejera

Gilberto Ramón Guevara Niebla
Consejero

Eduardo Backhoff Escudero
Consejero

Unidades administrativas

Francisco Miranda López
Unidad de Normatividad y Política Educativa

Jorge Antonio Hernández Uralde
Unidad de Evaluación del Sistema Educativo Nacional

Agustín Caso Raphael
Unidad de Información y Fomento de la Cultura de la Evaluación

Luis Salvador Castillo Montes
Unidad de Planeación, Coordinación y Comunicación Social

Miguel Ángel de Jesús López Reyes
Unidad de Administración

Luis Felipe Michel Díaz
Contraloría Interna

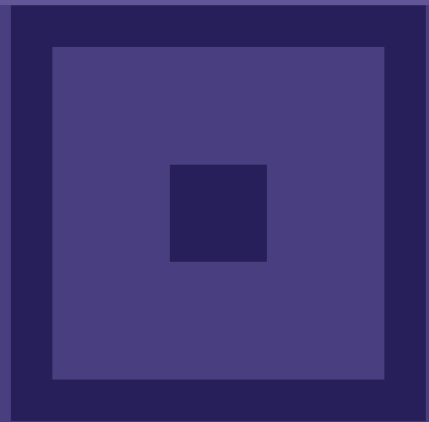


**Indicadores
educativos**

ANÁLISIS LONGITUDINAL DEL REGISTRO NACIONAL DE ALUMNOS
SOBRE TRAYECTORIAS EDUCATIVAS.

Es una publicación digital del Instituto Nacional para la Evaluación de la Educación.
Para su formación se empleó la familia tipográfica
Berthold Akzidenz Grotesk.

Septiembre, 2015



Comuníquese
con nosotros



Visite nuestro
portal