



**ESTADO
DE LA EDUCACIÓN**

SEXTO INFORME ESTADO DE LA EDUCACION

Patrones de distribución territorial de los docentes de primaria en Costa Rica

Leonardo Sánchez Hernández
Colaboradores
Marcos Castillo Bastos.
Luis Zamora González

2016



Nota: Las cifras de las ponencias pueden no coincidir con las consignadas por el Sexto Informe Estado de la Educación (2017) en el tema respectivo, debido a revisiones posteriores. En caso de encontrarse diferencia entre ambas fuentes, prevalecen las publicadas en el Informe.

INDICE DE CONTENIDO

I INTRODUCCIÓN.....	3
II METODOLOGÍA.....	4
III. PRINCIPALES HALLAZGOS	6
a) Características del docente de primaria y su distribución en el territorio nacional	6
b) Factores asociados al rendimiento escolar público	9
c) Factores asociados al rendimiento escolar público: un enfoque local.	13
i) Factores asociados a la Deserción	13
ii) Factores asociados a la Repitencia.....	17
iii) Factores asociados a la aprobación	21
d) Análisis exploratorio de los tamaños de las escuelas públicas en Costa Rica: el caso de las escuelas unidocentes.	24
i) Concentración espacial de las escuelas unidocentes.....	24
ii) Problemas asociados a las escuelas unidocentes.....	25
iii) Patrones de localización de escuelas unidocentes: Un análisis de proximidad	27
iv) Patrones de cobertura (población de 6 a 12 años) de las escuelas unidocentes en Costa Rica.	30
IV. CONSIDERACIONES FINALES.....	34
ANEXOS.....	37
ANEXO 1: Características de las bases de datos utilizadas.	38
ANEXO 2: Metodología para la caracterización de los docentes y su distribución espacial	40
ANEXO 3: Metodología para estimar los efecto de las características de los profesores, del centro educativo y del entorno geográfico en el cual se localizan las escuelas sobre la deserción, aprobación y repitencia.	54
ANEXO 4: Metodología para el análisis exploratorio con modelos econométricos geográficamente ponderados a nivel de escuela.....	59

I INTRODUCCIÓN

El estudio tiene como objetivo analizar los patrones de distribución territorial de los docentes de primaria en Costa Rica en el marco del Sexto Informe del Estado de la Educación. El estudio trata de dar respuestas a las siguientes interrogantes:

- a) ¿Las características de los docentes relacionadas con su formación, experiencia, situación laboral, salarios e incentivos se distribuyen de manera homogénea en las escuelas del país? De no ser así, ¿cuál es esa distribución?, ¿presenta patrones de concentración en algunas zonas o cantones?
- b) ¿Qué implicaciones tiene sobre los porcentajes de deserción, aprobación y repitencia en escuelas el hecho de que no exista una distribución homogénea de las características de los profesores?, ¿Cómo afectan otras variables como la infraestructura y el entorno social de la escuela dichos indicadores de rendimiento?

Para realizar los aspectos anteriores, se analiza la base de docentes de primaria, indagando sobre sus características e integrándola a la megabase de datos georeferenciada y otras bases georeferenciadas sobre variables socioeconómicas, geográficas y de infraestructura. Se construyen mediante técnicas de Sistemas de Información Geográfica (SIG) variables relacionadas con el entorno de la escuela y se aprovechan bases de datos existentes sobre características del centro educativo para complementar los análisis.

Las variables obtenidas son analizadas mediante técnicas econométricas y métodos de estadística espacial, permitiendo la exploración de relaciones con indicadores de rendimiento educativo de la escuela. Los hallazgos obtenidos permiten corroborar hipótesis consistentes con la teoría.

La introducción de nuevas variables, la incorporación del componente territorial, así como las particularidades de cada escuela explican en parte estos resultados y permiten obtener análisis a nivel local (econometría espacial) y no solo global. Lo anterior, es novedoso en el país y plantea la discusión sobre el diseño de políticas educativas focalizadas para primaria en un sistema con diferentes niveles y realidades territoriales.

Adicionalmente el informe realiza análisis exploratorios sobre los patrones de distribución espacial y la localización de las escuelas unidocentes. Lo anterior permitió indagar sobre dos aspectos:

- a) El grado de proximidad entre escuelas
- b) la cobertura (demanda) que tienen dichas escuelas a población entre 6 y 12 años.

Los resultados obtenidos son un insumo valioso para la toma de decisiones en términos de inversión en infraestructura y amplía la discusión sobre la demanda de dichos centros en un contexto demográfico donde la participación de los menores de 12 años es cada vez menor dentro de la pirámide poblacional.

El documento se estructura de la siguiente manera: en la primera sección se plantea la metodología a utilizar de forma general, la cual se desarrolla en el anexo 1 del documento. La segunda sección expone los principales hallazgos encontrados en el estudio desagregándolos en cuatro subsecciones:

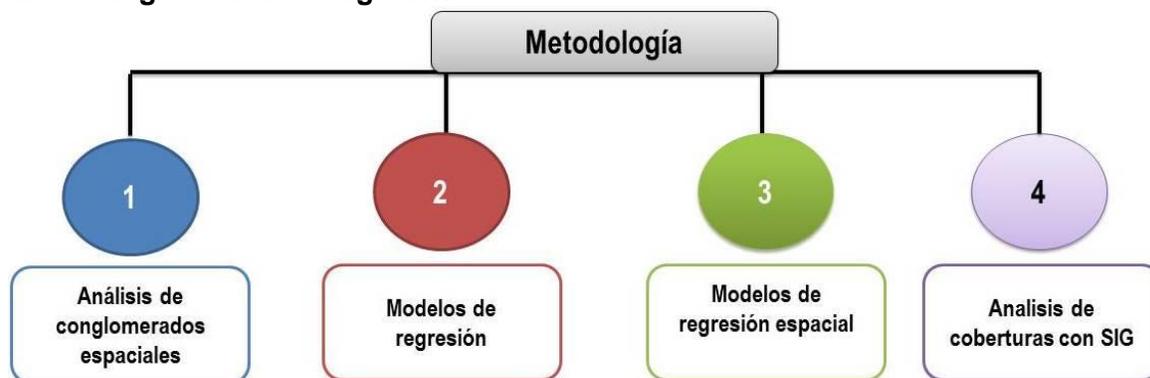
- a) Características del docente de primaria y su distribución en el territorio nacional
- b) Factores asociados al rendimiento escolar público
- c) Factores asociados al rendimiento escolar público: un enfoque local.
- d) Análisis exploratorio de los tamaños de las escuelas públicas en Costa Rica: el caso de las escuelas unidocentes.

En la tercera sección se exponen algunas consideraciones a la luz de los resultados encontrados y la cuarta sección muestra los anexos metodológicos de la ponencia.

II METODOLOGÍA

El abordaje metodológico¹ para la investigación propuesta está compuesto por cuatro elementos que se describen en el diagrama siguiente.

Diagrama 1:
Metodología de la investigación



- 1. Análisis de conglomerados espaciales²:** El objetivo con esta técnica es identificar escuelas donde hay presencia significativa de alguna

¹ El abordaje metodológico se desarrolla en los Anexos 1, 2,3 y 4.

² El desarrollo metodológico de esta herramienta se muestra en el anexo 2.

características del docente, que se encuentran rodeadas de escuelas con igual situación o bien, una escuela rodeada por otra escuela con situaciones diferentes; en ambos casos nos referiremos a ellas como zonas cluster (conglomerado).

Con este análisis, se trata de probar dos hipótesis, la primera establece que las diferentes características del docente de primaria evaluadas están distribuidas en forma aleatoria entre las escuelas del país. Por el contrario, la segunda hipótesis trata de probar que ese patrón de distribución de las características del docente no es aleatorio entre las escuelas, y por tanto tiende agruparse (cluster) en zonas específicas del país. Para aplicar la técnica se utiliza la base³ de datos suministrada por el Programa Estado de la Educación sobre docentes de primaria en escuelas públicas para el año 2015 del Ministerio de Educación Pública (MEP) y las estimaciones se realizan con Sistemas de Información Geográfica (SIG).

2. Modelos de regresión⁴: La metodología busca identificar factores asociados al rendimiento escolar (deserción, repitencia y aprobación) en escuelas públicas del país. Se desarrollan modelos econométricos con Mínimos Cuadrados Generalizados para estimar las relaciones. El análisis no consiste en abordar una función de producción educativa, sino, tratar de identificar relaciones estadísticas entre las características del docente, del centro educativo (infraestructura) e indirectamente con las características del hogar al incluir las zonas de pobreza (bajo clima educativo y bajos ingresos), así como la distancia del centro educativo al principal centro urbano que introduce el tema urbano-rural, accesibilidad y acceso a bienes y servicios. Se analizaron 25 variables introduciendo en los modelos solo aquellas que resultaron significativas y tenían una fuerte relación con la teoría.

3. Modelos de regresión espacial⁵: Los modelos de regresión expuestos en el punto anterior son modelos a nivel global, es decir estiman parámetros para todo el conjunto de escuelas del país y permiten identificar relaciones entre diferentes factores (determinantes) y los indicadores de rendimiento educativo. Sin embargo, el valor de los parámetros se debe interpretar como un valor promedio a nivel nacional, es decir, se parte del supuesto que un factor identificado afecta en promedio de manera similar a todos los centros educativos lo cual resulta en un supuesto muy fuerte dada las diferencias territoriales.

Dado lo anterior, se estiman modelos econométricos geográficamente ponderados con el fin de identificar como varían en el espacio los parámetros o factores estadísticamente significativos encontrados en los modelos globales.

³ Las características de la base de datos se muestran en el anexo 2

⁴ El desarrollo metodológico de esta herramienta se muestra en el anexo 3.

⁵ El desarrollo metodológico de esta herramienta se muestra en el anexo 4.

Para estimar dichos modelos se construye una ecuación distinta para cada escuela al incorporar las variables dependientes y explicativas de las escuelas que caen dentro del ancho de banda de cada entidad de destino. Para determinar la forma y la extensión del ancho de banda se utilizaron diferentes criterios (Kernel (núcleo), Ancho de banda, Distancia y Cantidad de vecinos). La aplicación de este tipo de modelos es posible gracias a la gran cantidad de datos disponibles (3.749 escuelas) y a que los mismos se encuentran georreferenciados.

4. **Analisis de coberturas⁶:** Esta metodología busca explorar la distribución espacial de los centros educativos unidocentes en el país (cercanía entre centros) y la demanda de niños en edad escolar que presentan. Para ello se realiza un análisis con sistemas de información geográfica para estimar las distancias entre escuelas y determinar los patrones de distribución espacial de proximidad. Por otro lado, se estiman áreas de cobertura para cada escuela para diferentes distancias con el fin de identificar la cantidad de niños de 6 a 12 años de edad en cada área de cobertura.

III. PRINCIPALES HALLAZGOS

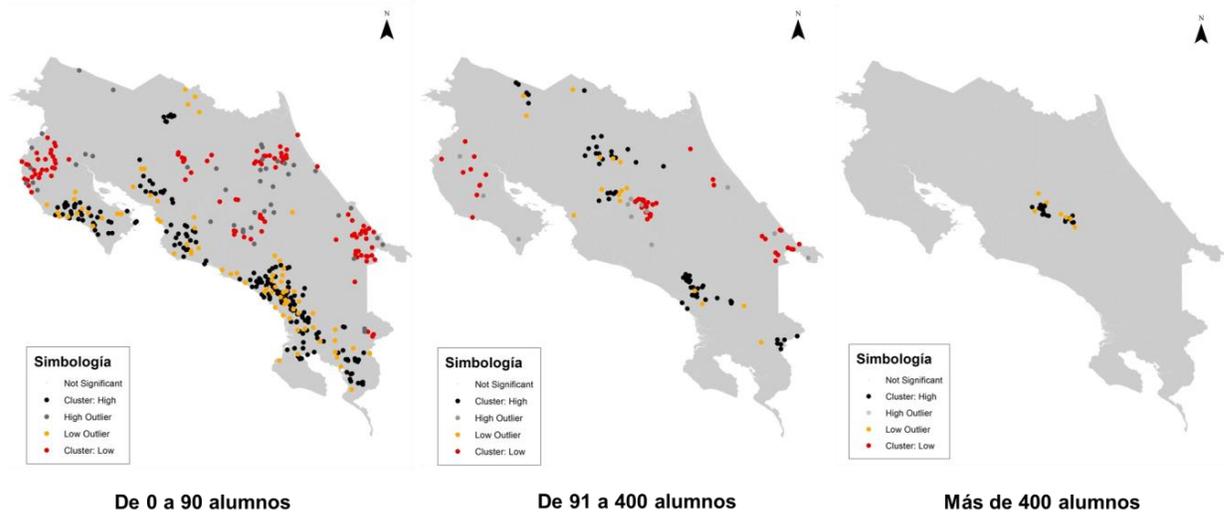
a) **Características del docente de primaria y su distribución en el territorio nacional⁷**

- En Costa Rica los docentes de primaria en escuelas públicas no se distribuyen en muchos casos de forma homogénea en el territorio nacional según sus características. Los resultados de esta investigación muestran que existen concentraciones en cantones y regiones específicas. En otras palabras, se puede concluir estadísticamente que no todas las escuelas públicas tienen las mismas posibilidades de acceder a docentes con perfiles similares creando desigualdades territoriales. Por características del docente entiéndase sexo, edad, grado académico, clase de puesto, grado profesional, categoría de formación, salarios e incentivos percibidos.
- Los docentes con grado profesional más alto (PT6 Y VT6) se concentran en escuelas específicas ubicadas en la península de Nicoya, Pacífico Central y Brunca. Este fenómeno no sucede en otras zonas del país. Por ejemplo, en los cantones de Guanacaste y la región Atlántica predominan docentes con grados profesionales menores. (Ver figura 1).

⁶ El desarrollo metodológico de esta herramienta se muestra en la última sección del apartado de Hallazgos...

⁷ Para más detalle ver anexo 2.

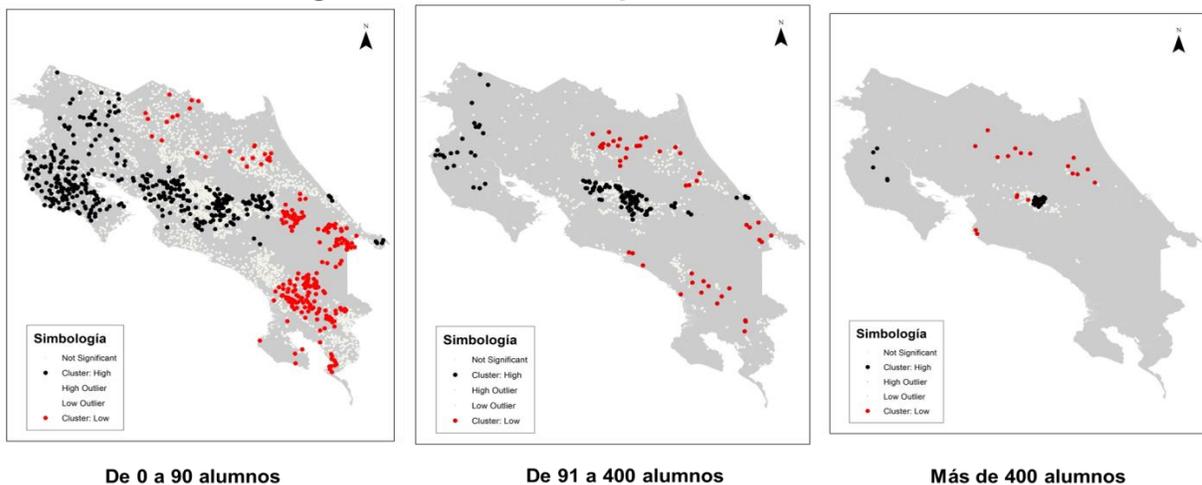
Figura 1
Clústeres de escuelas según grado profesional PT6 del docente por tamaño de la escuela, 2015.



Elaboración propia con información del MEP y PEE, 2016

- En línea con lo anterior, los resultados muestran que el sistema educativo costarricense tiende a concentrar a los docentes de mayor edad (especialmente mayores de 40 años) en la región Chorotega y GAM y a los más jóvenes (menores de 30) en las regiones Brunca, Atlántica y Huetar Norte. (Ver figura 2).

Figura 2
Clústeres de escuelas según edad del docente y tamaño de la escuela, 2015.

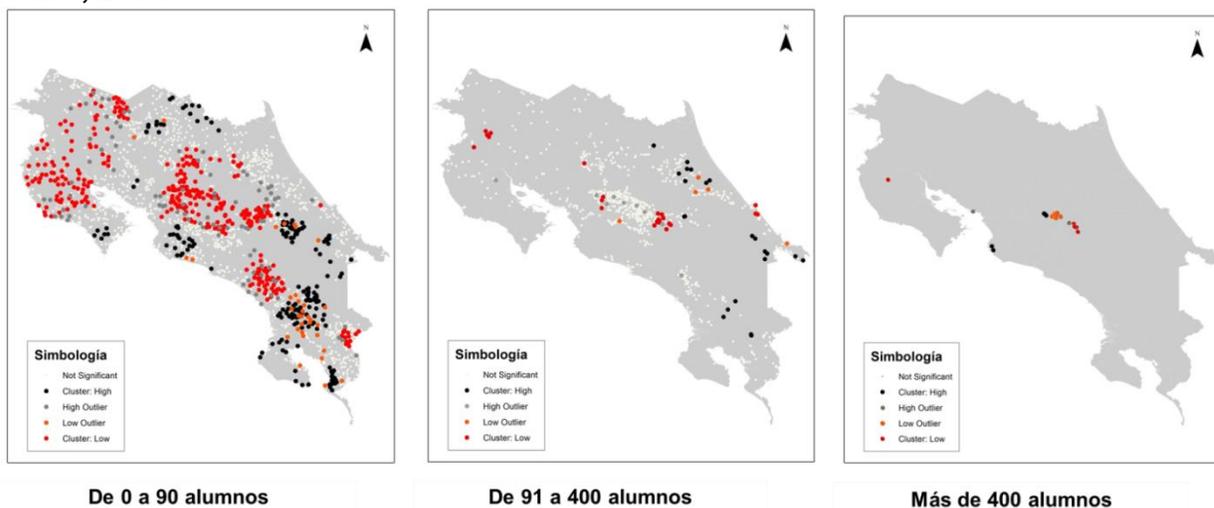


Elaboración propia con información del MEP y PEE, 2016

- De igual forma, se identificó que los docentes en condición de interinato, que no han alcanzado la propiedad, son distribuidos geográficamente en zonas muy específicas. Si bien es cierto imparten lecciones en diversos cantones del país el análisis de conglomerados ubico 13 cantones donde existe una fuerte

presencia de este perfil docente. Específicamente en Puntarenas, Aguirre, Garabito y Parrita en la región Pacífico Central, así como Osa, Golfito y Buenos Aires en la región Brunca. De igual forma Limón y Talamanca en la región Atlántica y otros cantones como Los Chiles, Guatuso, Abangares y Turrialba.

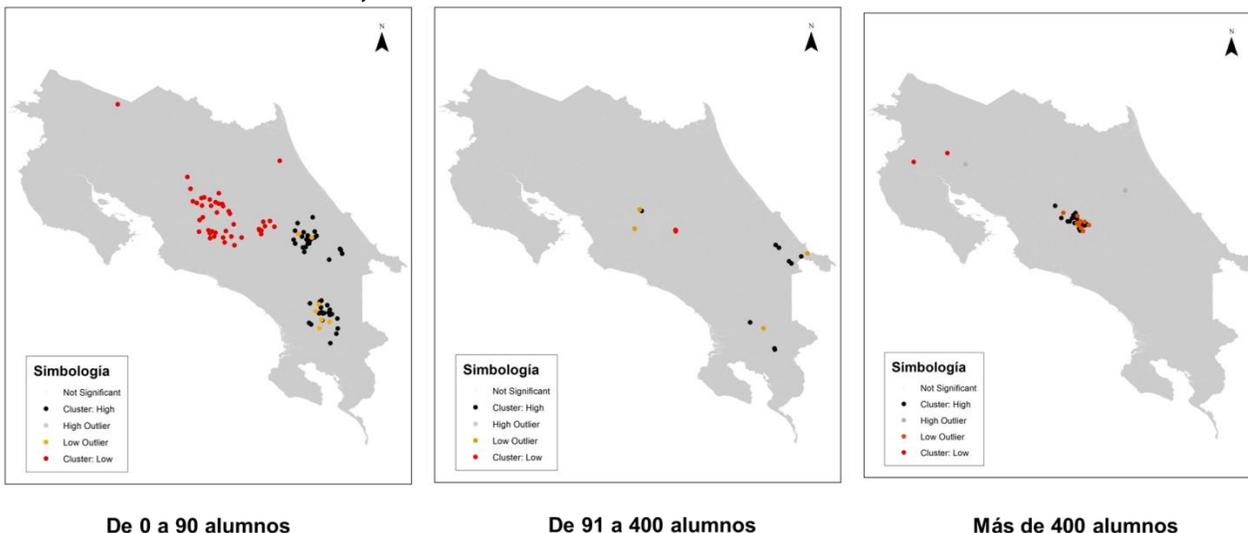
Figura 3
Clústeres de escuelas según condición de interinato del docente por tamaño del centro educativo, 2015.



Elaboración propia con información del MEP y PEE, 2016

– Otro hallazgo identificado es que los docentes en condición de Aspirantes se ubican especialmente en 5 cantones del país según el análisis de conglomerados realizado. Estas zonas tienen en la mayoría de los casos presencia de Territorios Indígenas. Estos cantones son: Turrialba, Talamanca, Matina, Limón y Buenos Aires. (Ver figura 4).

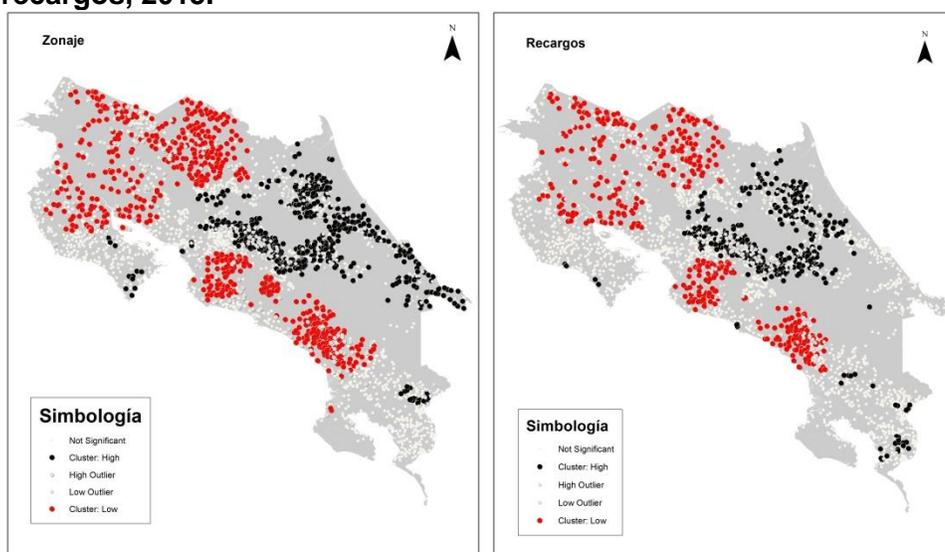
Figura 4
Clústeres de escuelas según el número de docentes en condición de aspirantes por tamaño del centro educativo, 2015.



– El pago de recargo por IDS y Zonaje debería por su definición estar concentrado en distritos y escuelas específicas del país, sin embargo, ambos casos no mostraron patrones de concentración y existe una distribución muy homogénea entre las escuelas públicas. (Ver figura 5).

En otras palabras, el recargo no parece estar respondiendo a criterios de focalización. Estadísticamente el pago del recargo en función del territorio tiende a ser aleatorio. Sin embargo, el análisis de conglomerados realizado muestra que si existen diferencias espaciales respecto al monto pagado por ambos recargos. Los docentes que trabajan en cantones de la GAM y la región Huetar Atlántica perciben un monto monetario mayor que el resto de regiones del país por este rubro. (Ver figura 5).

Figura 5
Clústeres de escuelas según salario promedio pagado a los docentes por concepto de zonaje y recargos, 2015.



Elaboración propia con información del MEP y PEE, 2016

b) Factores asociados al rendimiento escolar público⁸

Los datos a nivel de centros educativos de primaria sobre rendimiento educativo muestran diferencias en variables como repitencia, aprobación y deserción. Se identificaron una serie de factores, para determinar la medida en que esas diferencias son o no expresión de inequidades, relacionadas con el contexto inmediato del centro educativo y, por lo tanto, externas al sistema (desigualdad externa), o bien están asociadas a inequidades relacionadas con la dotación de

⁸ Para más detalle ver anexo 3.

factores con que los centros cuentan para cumplir con sus funciones (desigualdad interna al sistema educativo) o diferencias en las características de los docentes y su distribución espacial.

Es importante mencionar algunas limitaciones del presente análisis. La primera es que en educación primaria las tasas de repitencia, deserción y aprobación presentan valores relativamente bajos por lo que no existe una variabilidad entre centros educativos al igual que sucede con secundaria. Adicionalmente, el rendimiento educativo es producto de una serie de fenómenos que no siempre es captado por las variables existentes, en este caso no se cuenta con datos del estudiante, las familias o la dinámica del aula entre otros factores que la teoría menciona como relevantes. De igual forma, el análisis se realiza solo para centros educativos públicos ya que las bases de datos no estaban completas para todo el sistema educativo.

Estos aspectos ayudan a entender porque los modelos econométricos en primaria tienden a presentar un nivel de predicción menor en comparación con secundaria, sin embargo, como ejercicio, ayuda a entender como algunos factores afectan en mayor o menor magnitud el rendimiento en las escuelas del país. Algunos hallazgos relevantes son:

- Los resultados sugieren que un aumento en **el salario** de los profesores no tiene efectos sobre el desempeño educativo de primaria (repitencia, deserción y aprobación), lo que puede estar reflejando, dada el sistema actual de determinación salarial, que no existe relación positiva entre la formación del docente, anualidades y recargos y el desempeño.

- Otro hallazgo relevante es la relación estadística significativa entre escuelas con predominio de docentes **menores de 30 años** y mayores niveles de repitencia así como menores niveles de aprobación. El dato parece indicar que la experiencia medida por años de impartir lecciones afecta indirectamente el rendimiento educativo de las escuelas en las dos variables expuestas. Esto a pesar de que existe una creencia que docentes más jóvenes pueden estar más motivados para impartir clases, y/o que han recibido una mejor capacitación, dada mejoras en el sistema educativo a través del tiempo y por la implementación de métodos más tecnológicos. Como se mencionó en el apartado anterior, existe una clara concentración de centros educativos con profesores menores de 30 años en cantones de las regiones Brunca, Atlántica y Huetar Norte, con lo cual, este factor podría estar reproduciendo brechas territoriales en el desempeño educativo. (Ver cuadro 1).

- Por otro lado, se identificó que aquellas escuelas públicas donde los docentes presentan predominio del **grado profesional más alto (PT6)** se asocian a mayores niveles de aprobación, pero también menores niveles de repitencia. En este sentido cabe mencionar que este tipo de perfil docente no se distribuye homogéneamente en el territorio nacional, sino que presentan una mayor concentración en escuelas específicas ubicadas en la península de Nicoya,

Pacífico Central y Brunca caso contrario a lo que ocurre en los cantones de Guanacaste y la región Atlántica donde predominan docentes con grados profesionales menores. (Ver cuadro 1).

Cuadro 1
Resultados de los modelos de regresión, 2016

Variable/	Modelo ^b		
	Repitencia	Deserción	Aprobación
Porcentaje docentes que se les paga Zonaje en la escuela	-0,630**		
Porcentaje de profesores con menos de 30 años	0,896**		-0,922*
Porcentaje de profesores entre 30 y 40 años	0,478*		
Porcentaje de profesores con titulación PT6	-4,237**		1,988*
Porcentaje de profesores con titulación PT5	-3,761**		1,829*
Porcentaje de profesores con titulación KT2	-2,223*		-2,774*
Porcentaje de profesores en condición de Aspirantes dentro del colegio	-2,831**		
Cantidad de estudiantes por aula	0,098**	0,073**	-0,091**
Tamaño	-0,113*		
Infraestructura	-0,324**	-0,259*	0,273**
Inglés	-0,661**	-0,384*	0,734**
Zona Rural	-0,410**		
Zona de Pobreza	0,029**		-0,042**

a/ Significancia**p<0,05; ***: p<0,01

b/ Se muestran solo las variables significativas

Elaboración Propia.

– La seguridad laboral de los docentes de primaria medida por el **grado de interinato o propiedad** no es un factor que repercuta en el rendimiento del centro educativo. No se encontró evidencia estadística en los modelos aplicados para concluir que una escuela con mayor número de docentes interinos o en propiedad resulte en mejores niveles de aprobación o menores porcentajes de repitencia y deserción. Sin embargo, los centros educativos con predominio de **docentes aspirantes** si mostraron relación estadística significativa con mayores porcentajes de repitencia. El punto anterior es relevante si se toma en cuenta que los aspirantes se concentran en 5 cantones del país (Turrialba, Talamanca, Matina, Limón y Buenos Aires), donde existe en la mayoría de los casos presencia de Territorios Indígenas.

– La **cantidad de estudiantes por aula** afecta el logro académico obtenido por los estudiantes. Los resultados muestran que incrementos en el número de estudiantes por escuelas se asocia a menor rendimiento del centro educativo tanto en deserción como en repitencia y aprobación. En este sentido podrían derivarse varias hipótesis del porque estos resultados, algunos aspectos que se podrían mencionar es que los procesos de enseñanza, aprendizaje y convivencia en el aula de clase se podrían estar viendo afectados de diferentes maneras cuando se trabaja con grupos de clase más grandes. De acuerdo con Legler (2000), a menor número de alumnos en una clase, el número de interrupciones y el ruido disminuyen permitiendo al docente realizar otro tipo de actividades pedagógicas en el aula. Con grupos de clase pequeños, el docente puede brindar una atención

más personalizada y centrarse en las necesidades específicas de cada estudiante en lugar de las del grupo como un todo. Además, el tamaño de grupo afecta el nivel de ajuste o acoplamiento de los menores al mismo; a menor número de estudiantes se pueden llevar cabo actividades que exijan la participación de cada individuo. Adicionalmente, el tamaño de clase más pequeño aporta otros beneficios como ofrecerle al maestro la oportunidad de trabajar mejor con aquellos estudiantes que tienen necesidades especiales y con aquellos que por diversas razones se incorporan a la escuela un poco atrasados con respecto a sus compañeros. El tamaño de clase reducido se refleja también en la posibilidad de establecer una mejor relación con los padres y mayor contacto, que lleve a un mejor desempeño de los alumnos y a una reducción en los problemas disciplinarios en el aula de clase.

– El efecto de la **infraestructura** sobre el rendimiento de los alumnos, aunque limitado, es significativo. Decimos que limitado porque no conocemos el uso y estado de la misma, y a pesar de ello la sola presencia o tenencia de infraestructura en escuelas como gimnasio, comedor, biblioteca, taller de artes industriales y laboratorio de informática dio significativo dentro del modelo y se asocia a menores niveles de repitencia y deserción, así como mayor porcentaje de aprobación. Lo anterior, probablemente tenga que ver con que alumnos que estudian en establecimientos educativos con mejores condiciones de infraestructura se sienten más interesados por asistir a clase que aquellos que lo hacen en instalaciones que no disponen de servicios básicos y atractivos adicionales. Por esta razón, los insumos con que se dote la escuela tendrán un impacto apreciable en el logro académico de los estudiantes.

– Relacionado con este punto, el análisis encontró que **el tamaño** de la escuela sí es de relevancia. El tamaño esta positivamente relacionado con el rendimiento en la variable repitencia, es decir, conforme aumenta el tamaño de la escuela sistemáticamente se encontró menores niveles de repitencia. Lo anterior, probablemente este asociado que escuelas más grandes cuentan, a menudo, con más recursos.

– Los centros educativos de primaria en **zonas rurales y urbanas** reportan resultados distintos en el desempeño educativo. Estadísticamente, los resultados del modelo muestran que las zonas rurales presentan mejores desempeños en repitencia respecto a las zonas urbanas. Las variables deserción y aprobación resultaron ser no significativas. Estos resultados no deben sorprender dado que las escuelas están ubicadas a lo largo de todo el territorio nacional, en contextos sociales y económicos muy diversos.

– Los **niveles de pobreza** donde se ubica la escuela resultó en una variable significativa para explicar la repitencia y aprobación. En el caso de la repitencia esta tiende a aumentar conforme la incidencia de la pobreza aumenta mientras que la aprobación sigue un efecto contrario (tiende a disminuir). No se identificó ninguna relación significativa con la deserción. Esto nos muestra que un contexto

socioeconómico adverso, con familias de escasos recursos y bajos niveles educativos termina afectando las comodidades y facilidades que tienen los alumnos para asistir a la escuela y cumplir con sus deberes académicos. En este sentido, resulta difícil esperar mejores desempeños educativos en primaria, si las condiciones en que tienen lugar no son favorables.

– Por último, se encontró una relación significativa entre las escuelas que imparten lecciones del idioma **inglés** y el desempeño educativo. Específicamente se identificó que la deserción y repitencia disminuyen, y la aprobación aumenta.

c) Factores asociados al rendimiento escolar público⁹: un enfoque local.

Los resultados mostrados en la sección anterior responden a modelos a nivel global, es decir estiman parámetros para todo el conjunto de escuelas del país y permiten identificar relaciones entre diferentes factores (determinantes) y los indicadores de rendimiento educativo (deserción, repitencia y aprobación).

Sin embargo, el valor de los parámetros se debe interpretar como un valor promedio a nivel nacional, es decir, se parte del supuesto que un factor identificado afecta en promedio de manera similar a todos los centros educativos lo cual resulta en un supuesto muy fuerte dada las diferencias territoriales.

Por ejemplo, el rendimiento (deserción, repitencia y aprobación) de una escuela ubicada en el pacífico sur del país puede ser afectado por los mismos factores que una escuela ubicada en la frontera norte del país, sin embargo, el valor del parámetro puede ser muy diferente, dada la variación que se presenta en las relaciones de los factores en cada contexto.

Dado lo anterior, dentro de la investigación se procedió a estimar modelos econométricos geográficamente ponderados con el fin de identificar como varían en el espacio los parámetros o factores estadísticamente significativos encontrados en el análisis anterior.

Esto permite identificar en que zonas del país y en que variables se podría prestar mayor atención para focalizar políticas dirigidas a mejorar el rendimiento educativo en los centros educativos de primaria. Para estimar dichos modelos se construyeron ecuaciones distintas para cada escuela. Algunos hallazgos importantes son:

i) Factores asociados a la Deserción

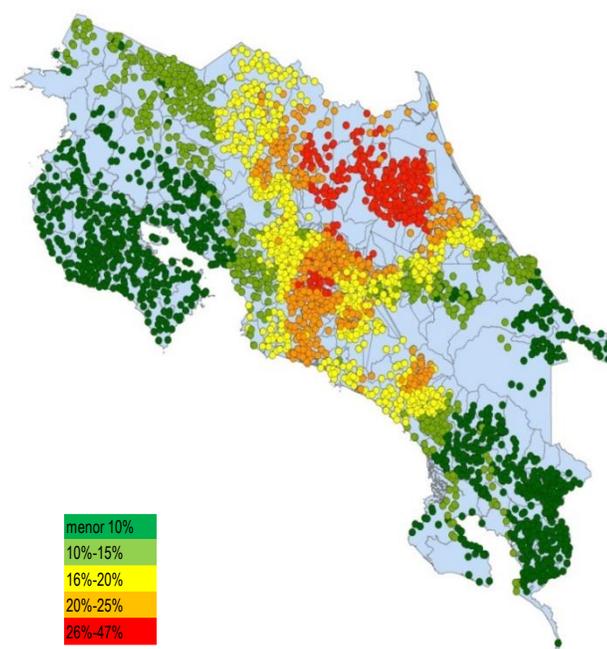
Con los datos existentes, los modelos diseñados para tratar de explicar el comportamiento de la deserción en las escuelas públicas del país tienden a ser un

⁹ Para más detalle ver anexo 4.

mejor predictor en cantones como Sarapiquí, San Carlos, Pococí, Mora, Acosta, Puriscal y Parrita, y menos explicativo en las regiones Chorotega, Brunca y la parte sur de la región Atlántica.

Lo anterior se observa en la figura 1 donde los R² más altos de los modelos locales se denotan con un color rojo en el mapa (el modelo explica entre el 26% y el 47% de la variabilidad en la deserción) y los R² más bajos y por tanto de menor predicción en color verde (explican menos del 10%). Esto confirma que el contexto geográfico juega un papel importante para entender el comportamiento del rendimiento escolar y que factores iguales, afectan de manera distinta el rendimiento según la zona geográfica del país en la que se ubique la escuela.

Figura 6
R² de los modelos locales para la variable deserción, 2016

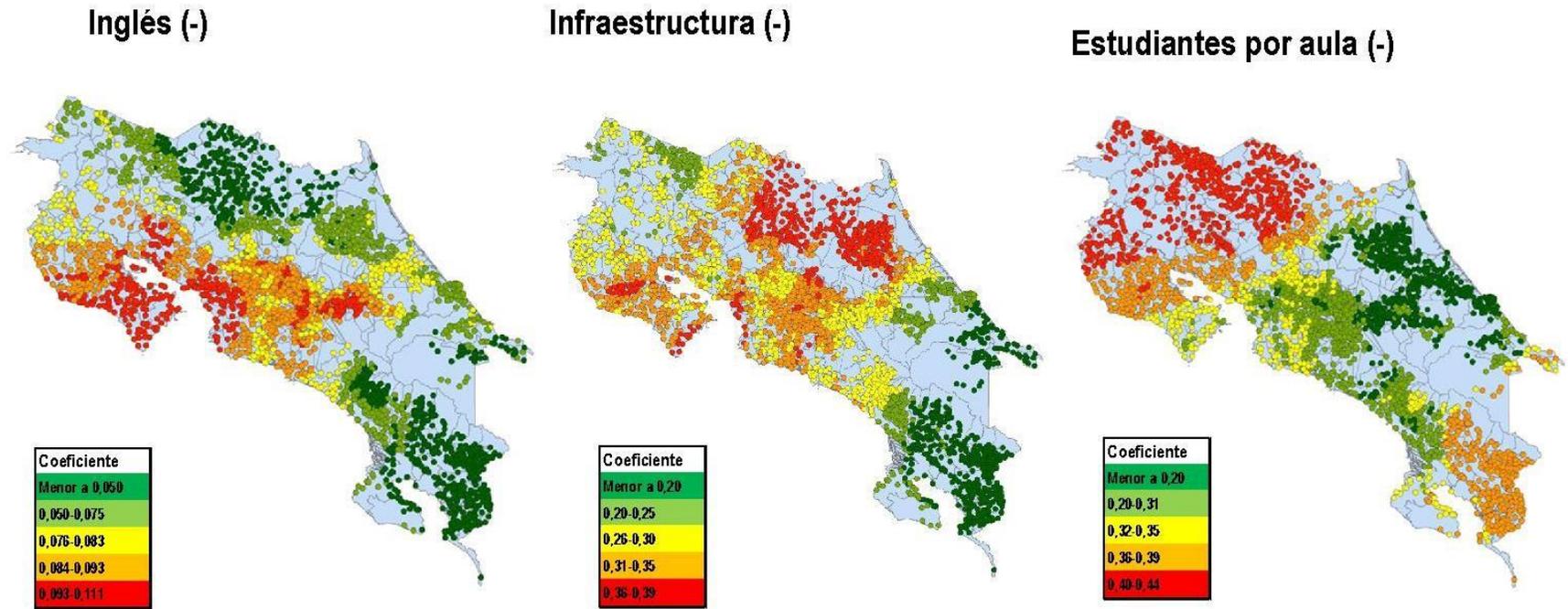


Elaboración propia con información del MEP y PEE, 2016

Los modelos desarrollados en la sección anterior identificaron en el caso de la deserción tres elementos o factores significativos estadísticamente, estos son: la infraestructura, la cantidad de estudiantes por aula y escuelas donde se imparten lecciones de inglés. Sin embargo, el efecto de dichas variables sobre la deserción no es homogéneo en el territorio, algunos aspectos que demuestran lo anterior son:

- Los resultados de los modelos locales desarrollados muestran que impartir **inglés** en las escuelas tienen a presentar un mayor efecto en la disminución de la deserción en centros educativos ubicados en la península de Nicoya, la GAM y la región Pacífico Central y una menor incidencia en las escuelas localizadas en los cantones de la región Huetar Norte, Brunca y Atlántica. (ver figura 7)
- En línea con el punto anterior, se identificó que la **Infraestructura** resulta ser más significativa para explicar la disminución en los niveles de deserción en las escuelas ubicadas en los cantones de San Carlos, Pococí, Sarapiquí, Hojancha, Nicoya, Puntarenas, Santa Barbara y Parrita. Por el contrario, en los cantones de Buenos Aires, Coto Brus, Corredores, Osa, Golfito, Limón, Talamanca y Upala el nivel explicativo tiende a ser mucho menor. (ver figura 7)
- Por último, si bien es cierto en el modelo global el número de **estudiantes por aula** resultó en una variable significativa para explicar los porcentajes en deserción, el modelo local muestra que esta variable es más relevante en escuelas ubicadas en cantones como San Carlos, Los Chiles, Upala, Guatuso y los cantones de la región Chorotega. Por el contrario, es menos significativa en cantones de la región Atlántica y Turrialba. (ver figura 7)

Figura 7
Coeficientes locales para la variable inglés, infraestructura y estudiantes por aula 2016



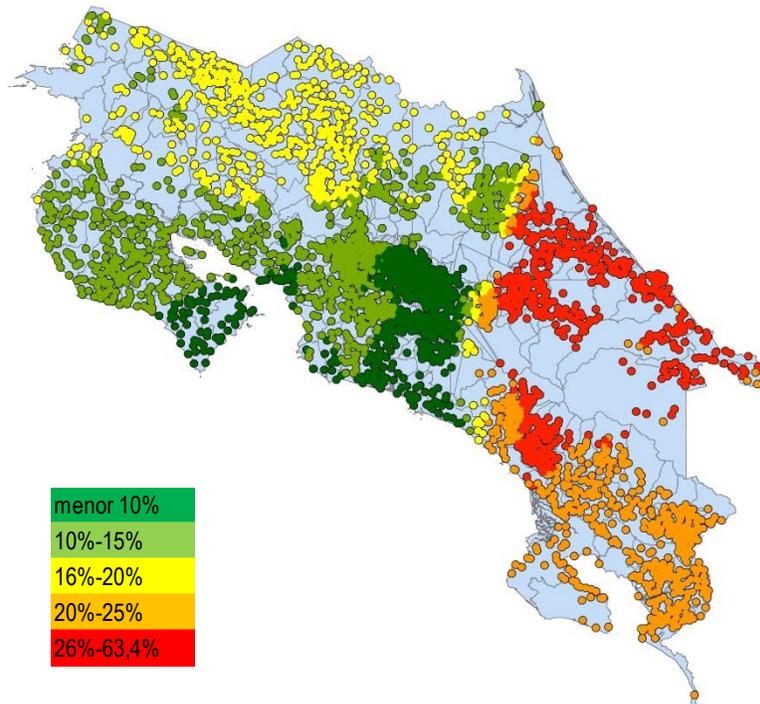
Elaboración propia con información del MEP y PEE, 2016

ii) Factores asociados a la Repitencia

Dentro del modelo global aplicado se pudieron identificar varios factores relacionados con la repitencia. Con el modelo local fue posible determinar en qué zonas del país resultaban ser más importantes o significativas dichas variables. Algunos hallazgos importantes son:

- El modelo Local aplicado para indagar sobre factores asociados a la repitencia tiende a ser un mejor predictor (R^2) en los cantones de las regiones Brunca y Atlántica y menos explicativo en Pacífico Central, el suroeste de la región Central y la península de Nicoya. (Ver figura 8).

Figura 8
 R^2 de los modelos locales para la variable repitencia, 2016



Elaboración propia con información del MEP y PEE, 2016

- El factor “**docentes aspirantes**” para explicar la repitencia es un fenómeno que toma mayor importancia en los cantones de la región Brunca, Talamanca, Santa Cruz, Nicoya, Hojanca, Abangares y Liberia. (Ver figura 9).
- La relación de “**estudiantes por aula**” si bien es cierto es una variable que dio significativa a nivel general los resultados locales muestran que es en los

cantones de la región Brunca y en Talamanca donde mayor incidencia tiene sobre la repitencia. (Ver figura 9).

– En el caso de la “**infraestructura**” los resultados muestran un efecto mayor sobre la disminución de la repitencia en los cantones de la región Chorotega y región Huetar Norte (menos San Carlos). En el caso de los centros educativos donde se imparte “**inglés**” el efecto sobre la reducción de la repitencia es mayor en cantones como San Carlos, Sarapiquí, Pococí, Carrillo, Liberia y Bagaces. (Ver figura 9).

– La “**pobreza**” resulto en una variable significativa asociada a mayores niveles de repitencia. Sin embargo, este factor toma mayor relevancia en los cantones de la Península de Nicoya, cantones de Cartago y Limón, así como la zona Norte del país. (Ver figura 10).

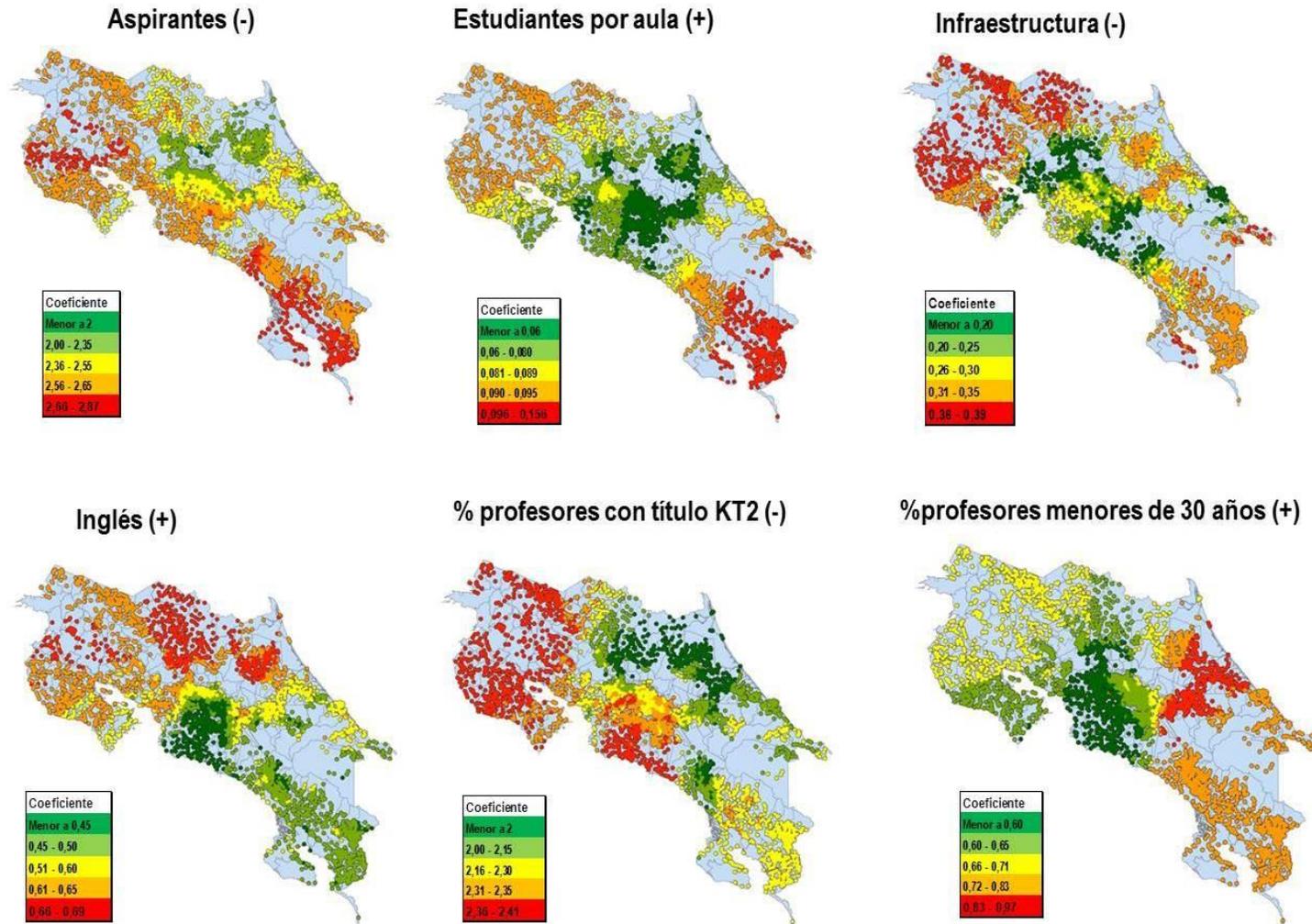
– Una mayor presencia de docentes con **grado profesional PT6** en centros educativos de primaria está asociado de acuerdo al modelo global a menores niveles de repitencia. Este factor toma mayor importancia en los cantones de la GAM, en Garabito, Esparza, San Mateo y Orotina en la región Pacífico Central, así como en los cantones de Turrubares, Puriscal, Sarapiquí y Pococí. (Ver figura 10).

– **La ruralidad** resulto en otro factor asociado a escuelas donde se presentan menor incidencia de la repitencia. Sin embargo, el modelo local nos indica que no es en todas las zonas rurales del país sino en algunos cantones de occidente de la región Central como San Ramón, Zarcero, Naranjo, Atenas. De igual forma los cantones de San Mateo, Esparza y Garabito en la región Pacífico Central, cantones de la región Brunca así como Turrialba y Talamanca. (Ver figura 10).

– Por último, **el tamaño** del centro educativo se identificó dentro del modelo global como un factor que tiene una relación positiva con los niveles de repitencia, es decir, centros educativos con mayor incidencia de repitencia están asociados a escuelas de mayor tamaño sistemáticamente. En este sentido, el modelo local permite identificar en que zonas del país este factor muestra mayor efecto sobre la repitencia dentro de los cuales están los cantones de la región Huetar Norte, Montes de Oro y Esparza en la región Pacífico Central, San Ramón, Zarcero, Naranjo y Valverde Vega en la zona de Occidente de la región Central, así como en Talamanca, Buenos Aires y Coto Brus. (Ver figura 9).

Figura 9

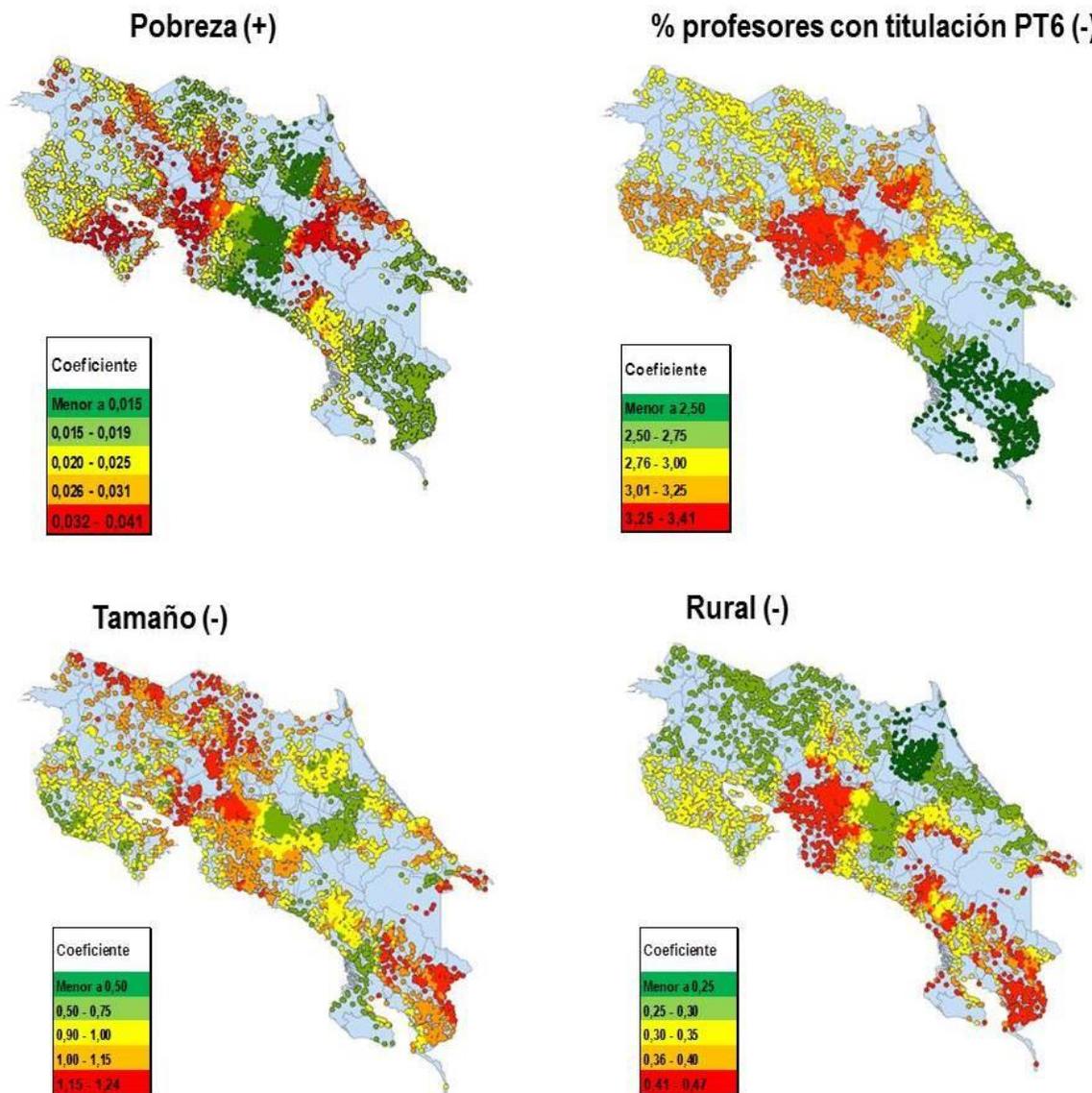
Coeficientes locales para las variables docentes aspirantes, estudiantes por aula, infraestructura, inglés, porcentaje de profesores con título kt2 y porcentaje de profesores menores de 30 años, 2016.



Elaboración propia con información del MEP y PEE, 2016

Figura 10

Coefficientes locales para las variables pobreza, porcentaje de profesores con titulación PT6, tamaño del centro educativo y zona (rural-urbana), 2016.



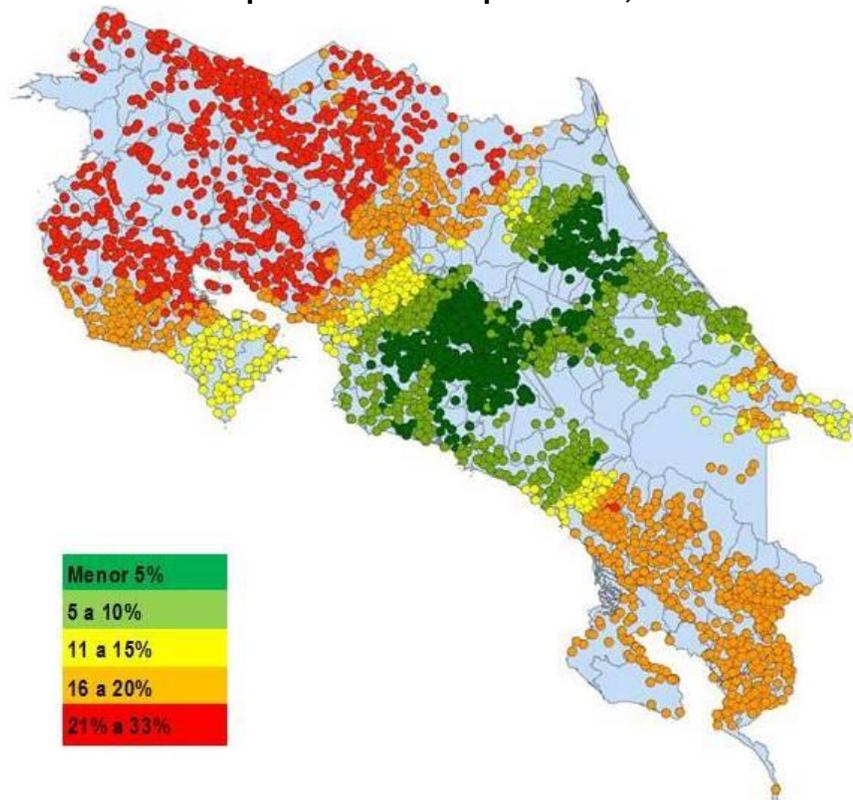
Elaboración propia con información del MEP y PEE, 2016

iii) Factores asociados a la aprobación

Dentro del modelo global aplicado se pudieron identificar varios factores relacionados con la aprobación en los centros educativos de primaria. Dentro los factores encontrados están la cantidad de estudiantes por aula, la infraestructura del centro educativo, impartir lecciones de inglés, el porcentaje de profesores con titulación KT2 y PT6, el porcentaje de docentes con menos de 30 años y las condiciones de pobreza donde se ubica la escuela. El objetivo de estimar el modelo econométrico local consiste en determinar las zonas del país donde resultaban ser más importantes o significativos los factores identificados. Algunos hallazgos importantes son:

– El modelo Local aplicado para indagar sobre factores asociados a la aprobación tiende a ser un mejor predictor (R^2) en los cantones de las regiones Chorotega y Huetar Norte (puntos color rojo del mapa) presentando los valores R^2 más altos. y menos explicativo en los cantones de la GAM y Región Atlántica. (Ver figura 11).

Figura 11
 R^2 de los modelos locales para la variable aprobación, 2016



Elaboración propia con información del MEP y PEE, 2016

– La relación de “**estudiantes por aula**” si bien es cierto es una variable que dio significativa a nivel general los resultados locales muestran que es en los

cantones de la región Brunca, Talamanca y Abangares donde mayor peso explicativo tiene sobre la aprobación, caso contrario el modelo tiende a ser menos predictivo (aunque significativo) en los cantones de Upala, Liberia y Carrillo. (Ver figura 12).

– En el caso de la “**infraestructura**” los resultados muestran un efecto mayor sobre el aumento de la aprobación en los cantones de San Carlos, Los Chiles, Guatuso, Sarapiquí y Tilarán. En el caso de los centros educativos donde se imparte “**inglés**” el efecto sobre la aprobación es mayor en cantones de la región Huetar Norte, Tilarán, Grecia, Talamanca, Limón y Pérez Zeledón. (Ver figura 12).

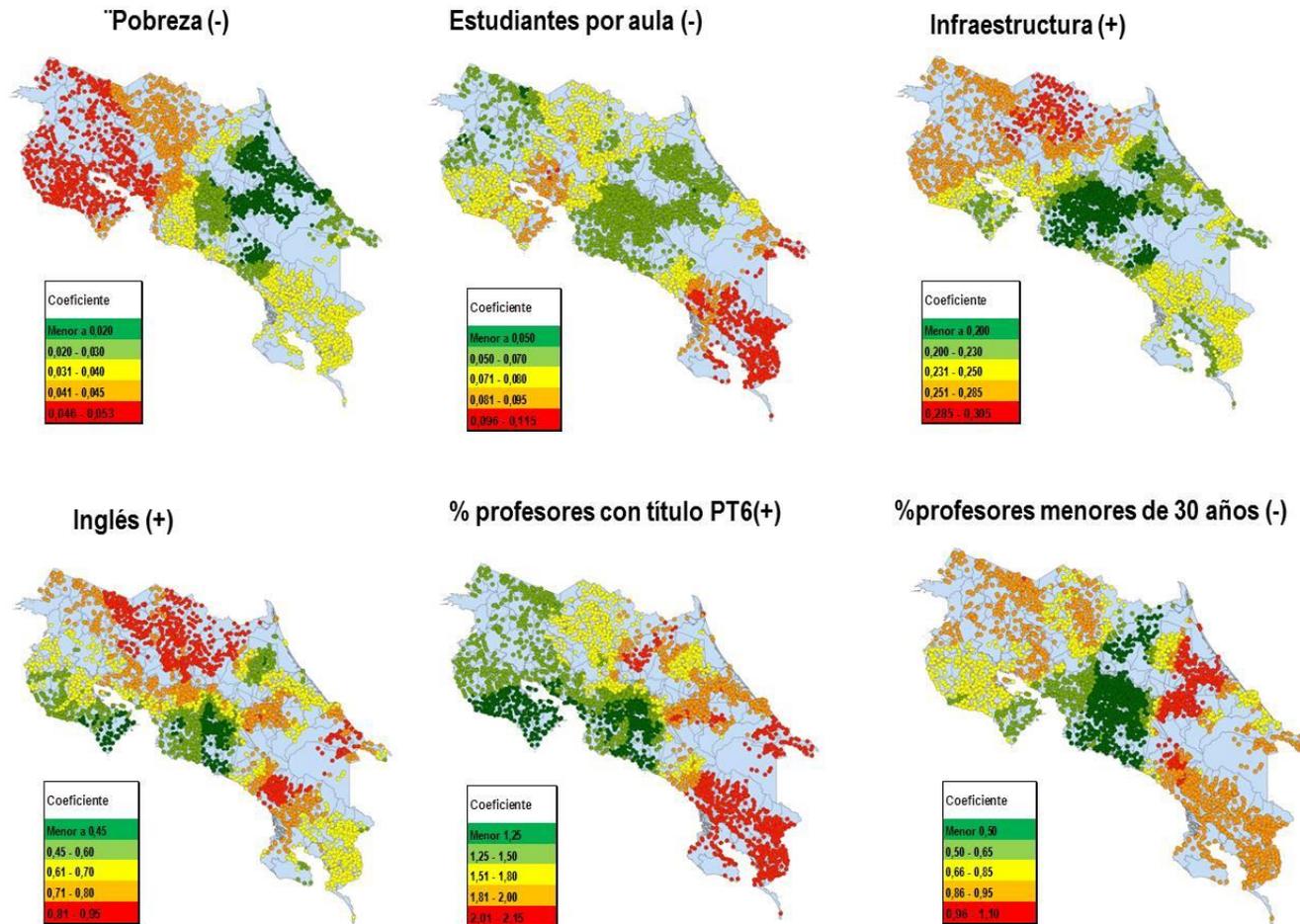
– La “**pobreza**” resultó en una variable significativa asociada a menores niveles de aprobación. Sin embargo, este factor toma mayor relevancia en los cantones de la región Chorotega, Upala, Guatuso, Puntarenas, Montes de Oro y Esparza y un menor grado de asociación en los cantones de la región Atlántica y la provincia de Cartago. (Ver figura 12).

– Una mayor presencia de docentes con **grado profesional PT6** en centros educativos de primaria está asociado de acuerdo al modelo global a mayores niveles de aprobación. Este factor toma mayor importancia en los cantones de la región Brunca, Limón, Talamanca y Sarapiquí, así como en Grecia y San Carlos y los cantones de la provincia de Cartago (excepto Paraíso y el Guarco). La predicción del modelo es menor (aunque significativa) en algunos cantones de la periferia sur y oeste de la región Central como Turrubares, Acosta, León Cortes, Puriscal, Aserrí, Mora, Santa Ana y Desamparados. De igual forma en cantones de Guanacaste como Nicoya, Nandayure y Hojanca así como en Puntarenas, Esparza, Parrita, Aguirre y Garabito en la región Pacífico Central. (Ver figura 12).

– Los resultados del modelo global muestran que el predominio de docentes menores de 30 años en las escuelas está relacionada a menores porcentajes de aprobación. Este hallazgo toma mayor relevancia (más significativo) en los cantones de la GAM, la periferia sur de la región Central (Turrubares, Puriscal, Acosta, Tarrazú y León Cortes) y en cantones de Puntarenas como Garabito, Parrita y Aguirre. De igual forma en Sarapiquí, Grecia y San Carlos presento los mayores coeficientes. Por el contrario, esta variable tiende a ser menos significativa en los cantones de la provincia de Cartago y Limón (Pococí, Guacimo y Siquirres). (Ver figura 12).

Figura 12

Coeficientes locales para las variables pobreza, porcentaje de profesores con titulación PT6, tamaño del centro educativo y zona (rural-urbana), 2016.



Elaboración propia con información del MEP y PEE, 2016

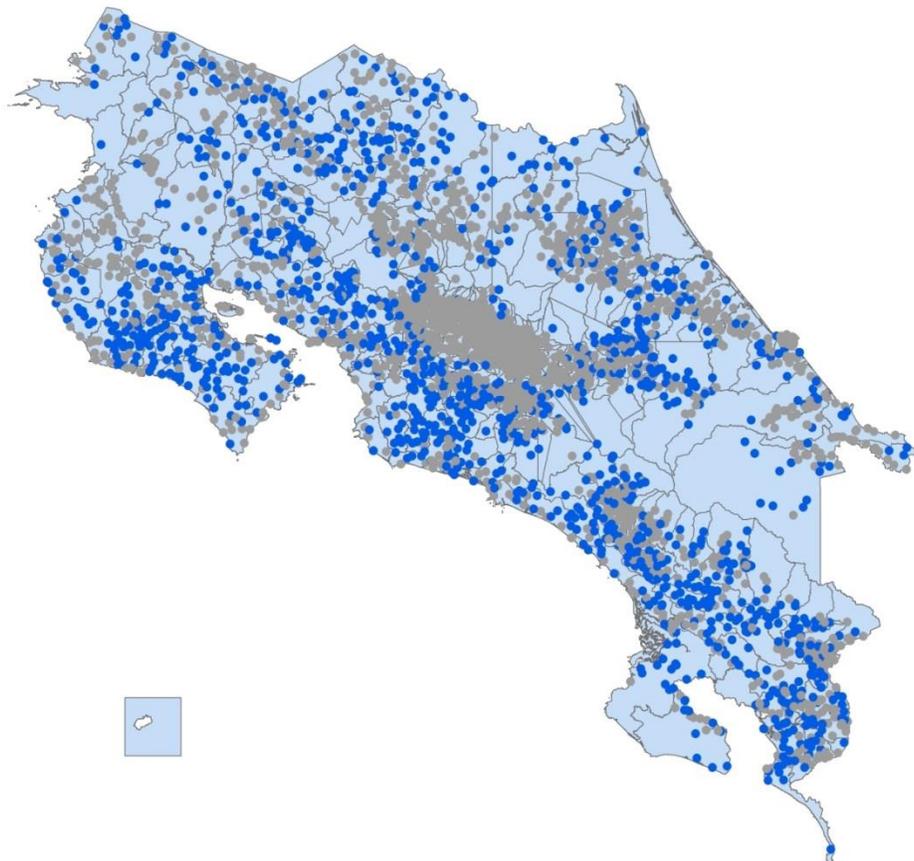
d) Análisis exploratorio de los tamaños de las escuelas públicas en costa rica: el caso de las escuelas unidocentes.

i) Concentración espacial de las escuelas unidocentes

Actualmente, el 37% de las escuelas del país son unidocentes, es decir, presentan menos de 30 estudiantes (97,4% son escuelas públicas). Este tipo de escuelas conforman el segundo grupo más importante en cuanto a cantidad de escuelas del país. Sin embargo, solo agrupan el 4% del total de matrícula del país.

La distribución espacial de las escuelas entre los cantones del país muestra concentración en 12 cantones que agrupan el 42% de todas las escuelas. Estos cantones son San Carlos, Pérez Zeledón, Buenos Aires Puntarenas, Pococí, Turrialba, Limón, Upala, Coto Brus, Golfito, Sarapiquí y Nicoya. (Ver figura 13).

Figura 13
Ubicación de las escuelas unidocentes (color azul del mapa), 2016.



Elaboración propia con información del MEP y PEE, 2016

En estos 12 cantones existe una clara predominancia de escuelas de menos de 200 alumnos, especialmente escuelas unidocentes de menos de 30 alumnos.

Nótese que ningún cantón de la lista anterior corresponde a la Gran Área Metropolitana, lo cual muestra que dichas escuelas están ubicadas principalmente en zonas rurales (76%).

El 41% de las escuelas Unidocentes del país se concentran en 8 cantones, los cuales tiene más de 50 escuelas de este tipo (Buenos Aires (92), Pérez Zeledón (74), Turrialba (68), Puntarenas (68), San Carlos (67), Osa (58), Golfito (58) y Nicoya (51)). Mientras que en el Gran Área Metropolitana (GAM) solo se ubican 58 escuelas de este tipo, representando el 4% de las mismas.

La creación de escuelas unidocentes en Costa Rica parece estar asociadas a zonas muy aisladas y con poca población, que se mantienen como pequeñas porque la forma de crecimiento poblacional en estas regiones fue mediante la creación de nuevas áreas agrícolas con población igualmente dispersa (y alejada de la escuela original), lo cual a su vez requiere de otra escuela pequeña.

ii) Problemas asociados a las escuelas unidocentes

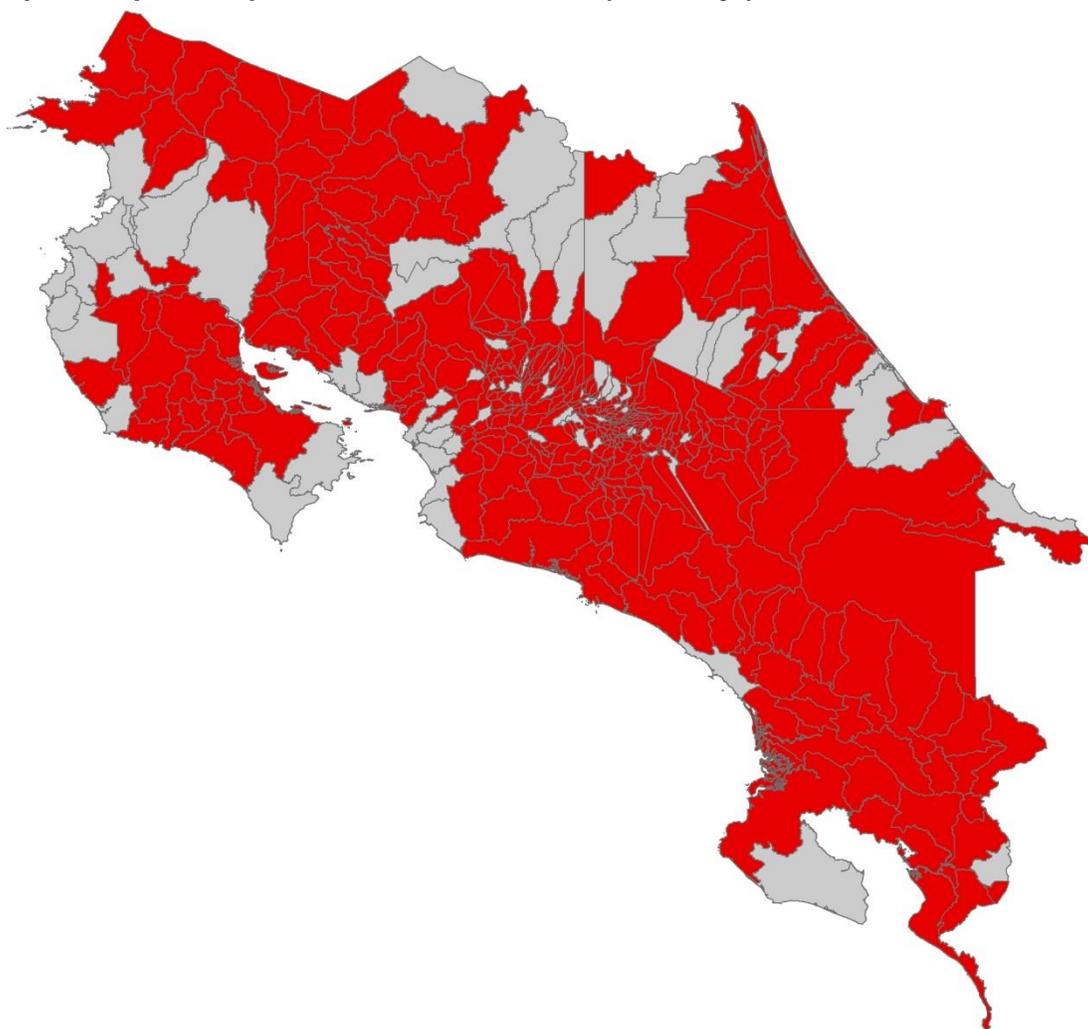
La creación de centros muy pequeños, sin embargo, plantea graves problemas en términos de gestión del sistema. Primeramente, para crear una escuela es necesaria una inversión inicial mínima independientemente de la cantidad de estudiantes, que incluye la construcción de aulas, adquisición de terreno, infraestructura auxiliar (baños, pasillos, cercas perimetrales, etc.) Esto significa que el costo por estudiante de hacer centros educativos pequeños es mayor que en centros de mayor tamaño. Cuando la diferencia en este costo es mucho mayor que los costos de transporte, tiene sentido considerar el traslado de estudiantes como estrategia de promoción de eficiencia.

El segundo gran problema de los centros educativos muy pequeños es la ausencia de economías de escala en la operación de los centros educativos. Así como existen costos fijos iniciales mayores por estudiante – de infraestructura – también hay costos mayores por estudiante si se proveyera a los centros más pequeños de docentes en materias especiales (música, deportes, artes), y a los centros de apoyo administrativo. En la práctica, como no es posible afrontar estos costos, el resultado es que los centros más pequeños carecen de este tipo de servicios, lo cual reduce la calidad de la experiencia educativa en ellos. Así, en una escuela unidocente, el docente tiene como recargo las funciones administrativas propias del director en un centro más grande y típicamente no cuenta con aula de música, artes, laboratorios y profesores especializados de ciencias, bibliotecas, etc.

Aunado a lo anterior, los resultados mostrados en la sección anterior arrojan que el tamaño del centro educativo presenta una relación directa con la repitencia. Conforme la escuela es más pequeña mayor probabilidad existe de presentar repitencia. De igual forma, en 2015 el promedio de deserción en escuelas unidocentes (5,8%) fue mayor al resto de escuelas (3,1%) y al promedio nacional (4,1%).

Por último, y probablemente de los aspectos más importantes, está el cambio en la estructura demográfica que ha venido experimentando el país. El grupo de población entre 6 y 12 años ha experimentado en los últimos años una disminución importante en el país, los datos de los dos últimos censos de población muestran que entre el año 2000 y el 2011 este grupo de población disminuyó en un 14%. Desde una perspectiva espacial, de los 472 distritos del país un total de 390 distritos perdieron población de 6 a 12 años, es decir un 82%. Muchos de ellos donde están localizadas las escuelas unidocentes (ver figura 14).

Figura 14
Distritos que han perdido población de 6 a 12 años (color rojo) entre el 2000-2011



Elaboración propia con información del INEC.

iii) Patrones de localización de escuelas unidocentes: Un análisis de proximidad

A continuación, se realizan dos análisis detallados para caracterizar la problemática mencionada anteriormente. Por un lado, se elabora un análisis con sistemas de información geográfica para estimar las distancias entre escuelas y determinar los patrones de distribución espacial de proximidad. Por otro lado, se estiman áreas de cobertura para cada escuela para diferentes distancias con el fin de identificar la cantidad de niños de 6 a 12 años de edad en cada área de cobertura.

En el marco de las políticas generales de creación de centros educativos, algunas suposiciones y aspiraciones sobre el funcionamiento de las escuelas contribuyen a explicar su proliferación. Las escuelas tradicionalmente han sido concebidas como más propias de la comunidad que los colegios. Se espera que los estudiantes viajen a ellas a pie. Bajo estas dos premisas, tiene mucho sentido crear centros lo más cerca posible de la población en edad de asistir, independientemente de la cercanía con otros centros (que "perteneceerían a otras comunidades", en este esquema).

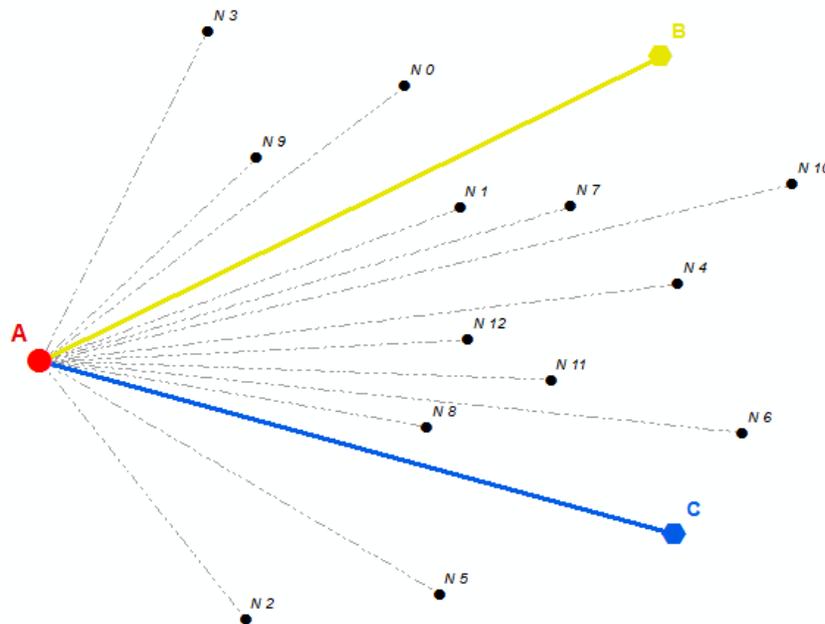
La ausencia de una visión del sistema ha resultado en numerosos casos de escuelas muy cercanas entre sí, con los problemas señalados en párrafos anteriores. El funcionamiento de las escuelas más pequeñas es críticamente dependiente de la calidad del docente responsable y su capacidad de involucrar a la comunidad en el apoyo al centro, el carácter extraordinario de quienes puedan enfrentar este reto introduce una dimensión adicional a la escasez de recursos y la necesidad de mayor eficiencia en el sistema.

Existen, además, serias dudas sobre el rendimiento académico que los estudiantes de escuelas unidocentes puedan tener en el colegio – aunque sea imposible una evaluación objetiva de esta dimensión si no existen datos de los estudiantes para la calibración de modelos que controlen por numerosos factores individuales.

Para determinar la distancia entre dos o más escuelas se utiliza la megabase de escuelas geo-referenciadas actualizada al 2016. Las mediciones de distancia son más precisas cuando los datos de entrada estén en un sistema de coordenadas proyectadas en equidistancia (UTM), sobre todo para reducir errores en el cálculo.

La metodología utilizada se muestra en diagrama 2. El diagrama ejemplifica como se estima la distancia de la escuela A con las escuelas B, C y N (en este ejemplo cada letra representa un shapefile); este proceso se lo puede llevar a cabo gracias a la herramienta Point distance (Distancia de punto) localizada en las herramientas de proximidad del ArcToolbox del programa ArcGIS utilizado.

Diagrama 2 Metodología para estimar distancias entre escuelas



Fuente: ArcGIS.

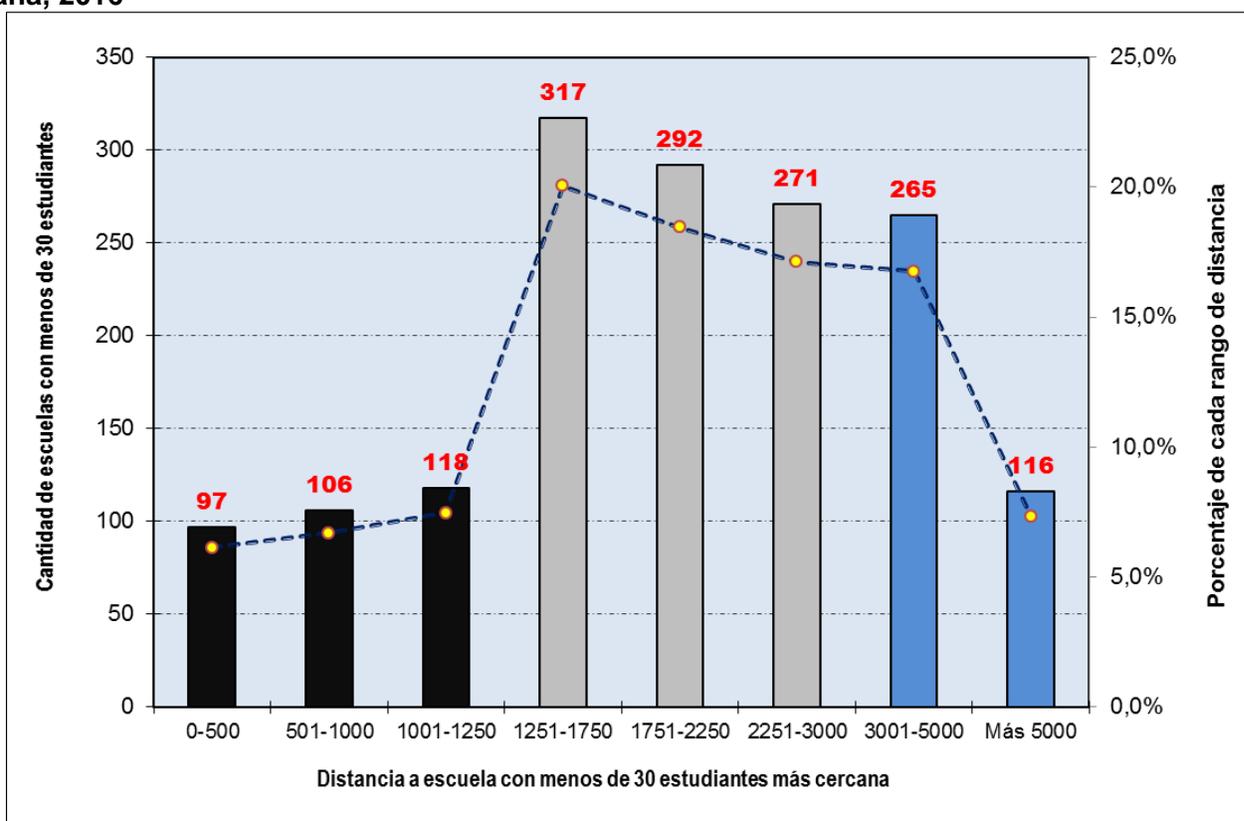
El procedimiento para determinar la distancia entre A-B, A-C y A-N se realiza de igual forma, los campos requeridos en Point Distance son los siguientes:

- Input Features, es el punto desde el cual se determinará la distancia (punto A).
- Near Features, es el punto donde se calculará la distancia desde el punto de entrada (puntos B y C contienen un solo punto, punto N contiene varios puntos).
- Output Table, requiere un directorio o geodatabase para almacenar la tabla con los resultados.
- Search Radius (optional), parámetro opcional que restringe el cálculo de distancia a un radio determinado. En este caso solo interesaban escuelas a 2km o menos de distancia.
- El último campo permite establecer la unidad de medida (metros, kilómetros, pies, yardas, etc).
- La distancia entre dos escuelas se estimó por carreteras para evitar sesgos de topografía.

Los resultados derivados del análisis se muestran en el gráfico 1 donde es posible deducir los siguientes aspectos:

- a) El 6,1% de las escuelas unidocentes tienen otra escuela con su mismo tamaño (menores de 30 estudiantes) a una distancia menor a 500 metros. Esto representa 97 centros educativos de primaria.
- b) Un total de 106 escuelas unidocentes tienen otra escuela unidocente a una distancia por carretera entre los 500 y 1.000 metros. Estos centros educativos agrupan el 6,7% de dichas escuelas del país.
- c) El 7,5% (118) de los centros educativos de primaria unidocente están ubicados a una distancia de 1.000 a 1.250 metros de otro centro unidocente. Otro 20% (317) están ubicadas a una distancia entre 1250 metros y 1750 metros.
- d) En resumen, se tiene que el 20,3% (321) de las escuelas unidocentes está a menos de 1.250 metros de distancia de otra escuela con características similares de tamaño, si este valor se amplía a 1.750 metros el porcentaje de las escuelas aumenta al 40% (638).

Grafico 1
Frecuencia de escuelas unidocentes según distancia a la escuela unidocente más cercana, 2016



Elaboración propia.

iv) Patrones de cobertura (población de 6 a 12 años) de las escuelas unidocentes en Costa Rica.

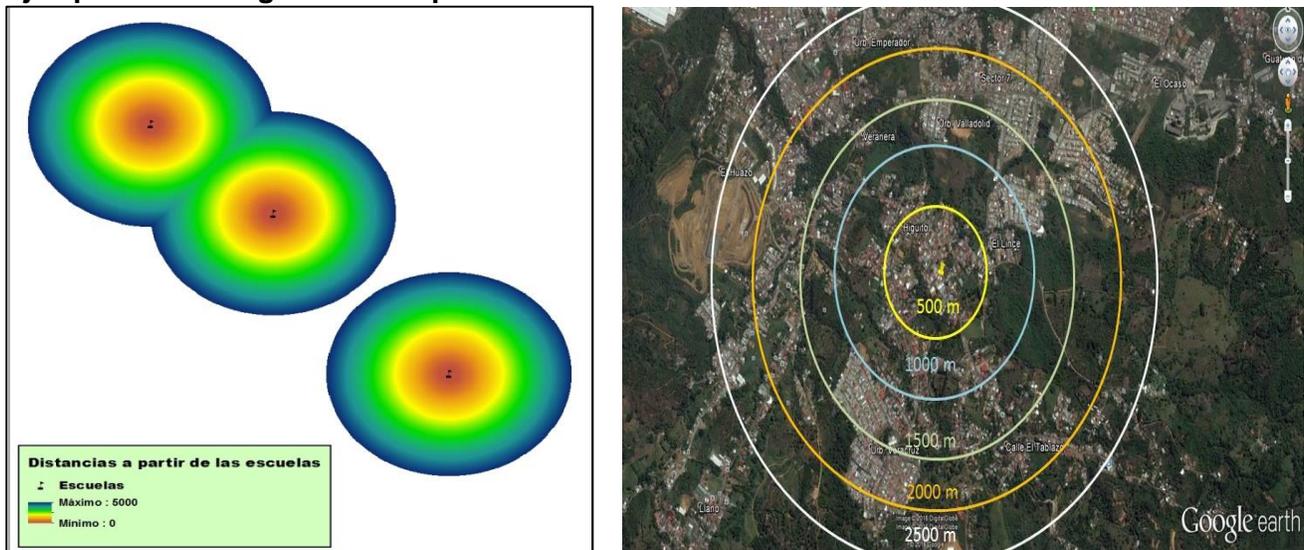
Como se mencionó anteriormente el país experimenta desde hace años un cambio en su estructura demográfica. La importancia relativa dentro de la pirámide poblacional de los menores de 12 años cada vez es menor. Bajo este panorama es importante estimar la demanda en términos de población que las escuelas del país tienen actualmente (2011) y a partir de dichos resultados pensar en el panorama de los próximos 10 a 20 años.

Para efectos de este análisis se estimó para cada escuela (aunque solo se muestra los resultados para centros educativos unidocentes) la cantidad de niños entre 6 y 12 años que tienen a diferentes radios de distancia (De 0 a 5.000 metros). Para ello, se utilizó la megabase de datos georreferenciada de escuelas y se integró con la cartografía censal del 2011 (unidades geoestadísticas mínimas (UGM)) lo cual permite calcular áreas de cobertura.

Las zonas de influencia de cada escuela se estimaron en ArcGis. Para efectos del análisis se crearon anillos cada 500 metros hasta llegar a los 5.000 metros. Se utilizó la herramienta zona de influencia en anillos múltiples para clasificar las áreas de alrededor de cada escuela en clases de distancia cercana (menos de 1000 metros), distancia moderada (entre 1000 y 1500 metros) y distancia larga para un análisis (De 1.500 a 5.000 metros).

Como se muestra en el diagrama 3 las zonas de influencia en anillos múltiples crean entidades de área a una distancia específica (o varias distancias especificadas) alrededor de escuelas.

Diagrama 3
Ejemplo metodológica utilizada para las áreas de cobertura

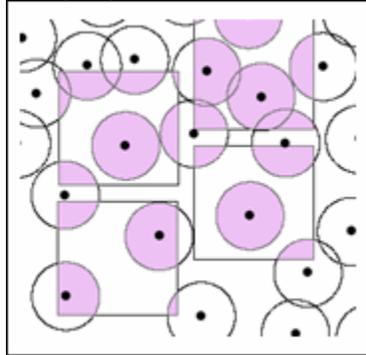


Elaboración propia.

Las zonas de influencia se pueden utilizar para seleccionar entidades en otra clase de entidad, o se pueden combinar con otras entidades mediante una herramienta de superposición, para buscar las partes de las entidades que caen en las zonas de influencia. En este caso escuelas y UGMs. (Ver diagrama 4).

Diagrama 4

Ejemplo metodología utilizada para superposición con UGMs



Fuente: ArcGIS.

Los resultados derivados del análisis se muestran en el cuadro 2 donde es posible deducir los siguientes aspectos:

- Para facilitar la interpretación del cuadro 13 hay que aclarar ciertos aspectos. En primero lugar las columnas muestran la cantidad de población de 6 a 12 años en rangos, por ejemplo, el primer rango hace referencia a menos de 10 niños entre 6 y 12 años, el segundo a una cantidad de 11 a 20 niños y así sucesivamente. Las filas muestran las áreas de influencia de las escuelas estimadas cada 500 metros (hasta los 5.000 metros). Y cada celda muestra la cantidad de escuelas.
- De esta manera, la primera celda del cuadro “314” nos indica que existen 314 escuelas unidocentes que tienen menos de 10 niños de 6 a 12 años en un radio de 500 metros o menos. Es importante aclarar que cada radio contiene al anterior, por ejemplo, el radio de 1000 metros o menos contiene las escuelas ubicadas en el radio de 500 metros o menos.
- El análisis muestra que 933 escuelas unidocentes (casi el 60% del total) tienen menos de 10 niños en edad escolar en un radio de 1 km. Una tercera parte de estas escuelas tienen menos de 10 niños a un radio de 500 metros.
- Si el área de influencia se amplía a 1,5 km se tienen un total de 777 escuelas unidocentes que tienen menos de 10 niños en dicho rango, este valor es de 895 escuelas cuando el radio se amplía 2km y de 943 escuelas cuando se pasa a 2,5km.

- Al utilizar un área de influencia de 5km se tiene que 1005 escuelas unidocentes (es decir el 65%) tiene menos de 10 niños en dicho radio.
- El 64% de las escuelas unidocentes tiene menos de 30 niños en edad escolar en un radio de 1km.
- Solo el 3% de las escuelas unidocentes tiene más de 30 niños en edad escolar a 1km o menos de distancia. Y solo el 0,7% tiene más de 50 niños en dicha distancia.
- Si el área de influencia de cada escuela unidocente se extiende a 5km (lo cual es 5 veces lo recomendado) solo el 10% de dichos centros tendría una cantidad de niños mayor a 30.
- Es importante mencionar que el análisis se hace escuela por escuela, por lo que algunas escuelas podrían estar compartiendo áreas de influencias con los cual la cantidad de niños en edad escolar para cada escuela es aún mucho menor.

Cuadro 2

Áreas de cobertura de las escuelas unidocentes: Cantidad de niños de 6 a 12 años de edad según radios de influencia.

Distancias (áreas de cobertura)	Población entre 6 y 12 años					
	10 o menos	11 - 20	21 - 30	31 - 40	41 - 50	51 o más
500 m o menos	314	99	38	12	5	2
1000 m o menos	619	271	101	37	12	9
1500 m o menos	777	432	149	55	23	19
2000 m o menos	895	517	182	69	33	34
2500 m o menos	943	560	201	75	35	36
3000 m o menos	973	595	221	80	35	38
3500 m o menos	989	603	229	83	36	40
4000 m o menos	998	610	234	84	36	42
4500 m o menos	1000	611	238	87	36	43
5000 m o menos	1005	614	241	89	36	43

Elaboración propia.

Es evidente que muchas escuelas pequeñas son inevitables, en el sentido de que están demasiado aisladas y evidentemente la población cercana es muy pequeña

de modo que la escuela no va a crecer, pero tampoco puede eliminarse porque dejaría a esta población sin acceso a educación primaria. En general, las escuelas pequeñas a más de 5 Km. de una escuela grande.

Las escuelas ubicadas a entre 1,2 km y 5 Km de la escuela más cercana podrían estar lo suficientemente cerca de otra como para pensar en unir la escuela pequeña a la más grande, según la densidad de la red vial, rutas de transporte público y el relieve.

Sin embargo, no se trata de argumentar en este documento por una política general de consolidación de escuelas sino de identificar oportunidades de consolidación, tanto geográficas como funcionales, que permitan el diseño de estrategias óptimas para la administración del sistema.

En este sentido, si la racionalización de la ubicación de centros educativos va a ser una necesidad a mediano plazo, las escuelas que estén a menos de 1Km. de otros centros son candidatos naturales para un programa de consolidación (unir escuelas pequeñas entre sí o unir escuelas pequeñas a centros más grandes existentes).

Serán necesarios análisis económicos y financieros de modo que pueda evaluarse la sostenibilidad en el tiempo del sistema. En esta dimensión, el objetivo debe ser promover la eficiencia y aprovechar economías de escala para mejorar la calidad de la educación.

No se trata solo de reducir el costo por estudiante al nivel del sistema general: incluso podría aceptarse un costo por estudiante mayor a las escuelas no unidocentes (pero sí menor que el sistema unidocente actual, pues de lo contrario no resultaría sostenible). Pero este costo debe implicar mejoras reales en la oferta educativa y la operación del sistema. Por ejemplo, deberían reforzarse las "materias especiales" (artes plásticas, deportes, música) y el funcionamiento administrativo de las escuelas.

Finalmente, debe optimizarse la ubicación del centro consolidado. Esto implica seleccionar como sede física de la escuela, de las escuelas candidatas a consolidación: la que tenga infraestructura más amplia, y/o mejores perspectivas de ampliación (un lote más grande y plano, un edificio al cual pueda añadirse una segunda planta), acceso a servicios básicos y accesibilidad al sistema de transportes y en núcleos consolidados de población.

Se trata de optimizar la localización y de crear sistemas de transportes de escolares, si fuera necesario, para facilitar la llegada de los estudiantes a quienes se les ha aumentado excesivamente la distancia de viaje (tómese en cuenta que este es un costo adicional que no existía antes, y en este sentido las dimensiones señaladas son complementarias: deben ser evaluadas en conjunto, con el objetivo de encontrar combinaciones de acciones que produzcan mejoras en todas ellas a la vez).

Este conjunto de esfuerzos puede concebirse como una variedad de "experimentos naturales" de política pública que, como mínimo, tendrán la misma

calidad de educación el sistema precedente y que además permitirán obtener lecciones sobre acciones efectivas y no efectivas en la consolidación de centros, que permitan guiar políticas más generales posteriormente.

IV. CONSIDERACIONES FINALES

– El objetivo central de esta ponencia, es incrementar el conocimiento de los factores críticos que afectan la calidad de la educación primaria en Costa Rica, para facilitar el diseño de políticas orientadas a mejorar el desempeño escolar.

– El estudio presenta una serie de factores que tienen mayor incidencia sobre el desempeño escolar de las escuelas públicas del país. Aunque tales modelos pretenden ser generales, en el sentido de que podrían aplicarse en diferentes zonas del país, es importante plantear dos advertencias fundamentales. En primer lugar, que los diferentes factores no deben verse como independientes el uno del otro, sino que, por el contrario, éstos se condicionan y retroalimentan mutuamente. En segundo lugar, que la combinación de características, y la forma como interactúan estos factores dependen de la ubicación geográfica de la escuela que a su vez refleja diferentes contextos (institucionales, culturales, sociales y políticos) en que operan. En este sentido el estudio aporta insumos importantes diferenciando en que cantones del país los factores identificados tienen una mayor o menor importancia para explicar el fenómeno del desempeño escolar.

– Aunque muchos trabajadores del sector de la educación en Costa Rica argumentan que los problemas y soluciones relacionados al desempeño de la educación son bien conocidos, y señalan que existen demasiados estudios sobre el tema, la evidencia empírica en esta ponencia muestra lo contrario y que aún es limitada. A pesar de que las características y variables de los modelos empleados en este estudio y otros realizados en el país en esta línea ha aumentado sustancialmente, la cantidad de variables acerca de los procesos de inputs y de la organización de la escuela incluida en las regresiones es aún limitada; la mayoría de las variables operacionalizan los aspectos cuantitativos de los inputs más que los cualitativos.

– En línea con el punto anterior, una futura agenda de investigación necesita incluir características hasta ahora excluidas de los análisis precedentes en el país, como son el liderazgo del docente y del director y otras actitudes como la autoestima y la impotencia; la estabilidad del personal, la articulación organización del currículum, el compromiso de los padres en la actividades escolares y la actitud hacia la escolaridad; la naturaleza de la capacitación en servicio; y otras características del manejo de la escuela y la organización de la clase. También se necesita continuar incluyendo características como salud y nutrición e intervenciones sanitarias. La experimentación cuidadosa de modelos innovadores bien diseñados y la evaluación sistemática de los resultados son necesarios para apoyar políticas específicas.

– Si bien es cierto una mayor titulación de los docentes no garantiza una mejor calidad de enseñanza, los resultados encontrados dan indicios de que existen relaciones positivas con el desempeño escolar. En este sentido, el Ministerio de Educación y los centros de formación deben seguir trabajando en la necesidad de proporcionar a los docentes una sólida formación en su disciplina. La formación no solo se debe centrar en el plano pedagógico como hacen la mayoría de programas de formación universitaria sino también llevarlos a comprender las formas de aprender del alumno. En muchos centros educativos las dificultades en el aprendizaje del alumno no se abordan de forma específica durante la preparación básica.

– En línea con lo anterior, en el nivel de formación continua y permanente que reciben los docentes se repiten las mismas tendencias que en el nivel de la formación inicial universitaria. En este sentido, se debe promover una sensibilización ante la problemática de la expulsión de los niños y jóvenes del sistema educativo y sus posibles soluciones. En la actualidad, es normal encontrar cursos de formación continua y permanente de los docentes estrictamente enfocados a una actualización de los conocimientos en su disciplina y a algunas jornadas de estudio e intercambio de experiencias, a pesar de que el docente es un factor elemental en la lucha por conseguir el éxito escolar del máximo número de alumnos. No solo se trata de escolarizar a los niños sino también conocer la calidad del docente con que se está escolarizando. En resumen, el docente debe también sentirse responsable del fracaso escolar de sus alumnos, y no solo de su buen rendimiento.

– Otro de los factores importante identificados en el estudio para explicar el desempeño educativo es la infraestructura. Si bien es cierto, no se cuenta con información del estado o uso que se le da, el solo hecho de tener bibliotecas, laboratorios, computadoras, etc presenta una relación estrecha con escuelas de mejor desempeño escolar. En este sentido, para mejorar las condiciones del sistema educativo es importante crear políticas sociales que atiendan estas condiciones de infraestructura, y no únicamente acciones educativas. En esta línea, es importante destinar ingresos al mantenimiento y conservación de los edificios escolares, además de una supervisión del uso que se les da a los distintos espacios de los que dispone cada escuela para aprovecharlos de la mejor manera, dando prioridad al equipamiento y conservación de aulas escolares, así como a los espacios físicos que apoyan el aprendizaje de los alumnos como son bibliotecas, salones de cómputo y laboratorios de ciencias.

– Las inversiones en infraestructura escolar son verdaderamente trascendentales y su impacto será tangible en el mediano plazo. Es importante que las inversiones en educación se focalicen bien y busquen resolver las grandes deficiencias existentes en escuelas rurales, especialmente aquellas relacionadas con necesidades básicas. Para ello, debemos exigir y abogar por la recopilación de información completa sobre las características físicas de las escuelas y

promover más estudios que analicen el vínculo entre factores asociados, como en este caso la infraestructura escolar y los aprendizajes de niños y niñas.

– La reforma educativa de 1888 resultó en una gran expansión del sistema de educación primaria. Durante los siguientes 65 años, esta expansión se centraría en zonas rurales (i.e. fuera de los cantones que actualmente componen la GAM): dos de cada tres escuelas fueron creadas en ellos. Este proceso de apertura de escuelas se estructuró como una doble expansión: en zonas de población dispersa (de reciente colonización) y, paralelamente, de reforzamiento en las zonas centrales. La dispersión poblacional inherente al proceso de colonización en la frontera agrícola ha resultado en una gran cantidad de escuelas que aún hoy son muy pequeñas. La tendencia demográfica que justificó este proceso ha cambiado con el agotamiento de la frontera agrícola y el surgimiento de mayores oportunidades en las ciudades del país, lo cual posiblemente resultará en todavía menos habitantes en las zonas rurales servidas por escuelas unidocentes. En este sentido, el análisis estadístico-geográfico elaborado en esta ponencia permitió identificar concentraciones de escuelas pequeñas: al sur de la Península de Nicoya, en el Pacífico Central y al sur de la Región Brunca y en algunas zonas específicas de la Región Huetar Norte y en las montañas de Talamanca.

– Con respecto al punto anterior, las aglomeraciones de escuelas pequeñas son zonas servidas casi exclusivamente por escuelas pequeñas. Esta condición es poco eficiente pues, dada su cercanía, sería posible servir varias de estas poblaciones con una escuela de mayor tamaño centralmente localizada. Esto permitiría aprovechar economías de escala para mejorar la gestión y calidad educativa del sistema. Existen áreas, sobre todo al sur de la Gran Área Metropolitana, donde las escuelas pequeñas están rodeadas por escuelas con mayor matrícula: pueden haberse constituido en una solución a problemas de sobrepoblación en las escuelas vecinas, en cuyo caso serían síntoma de un problema más grave. Sin embargo, esta zona requiere de análisis más profundos.

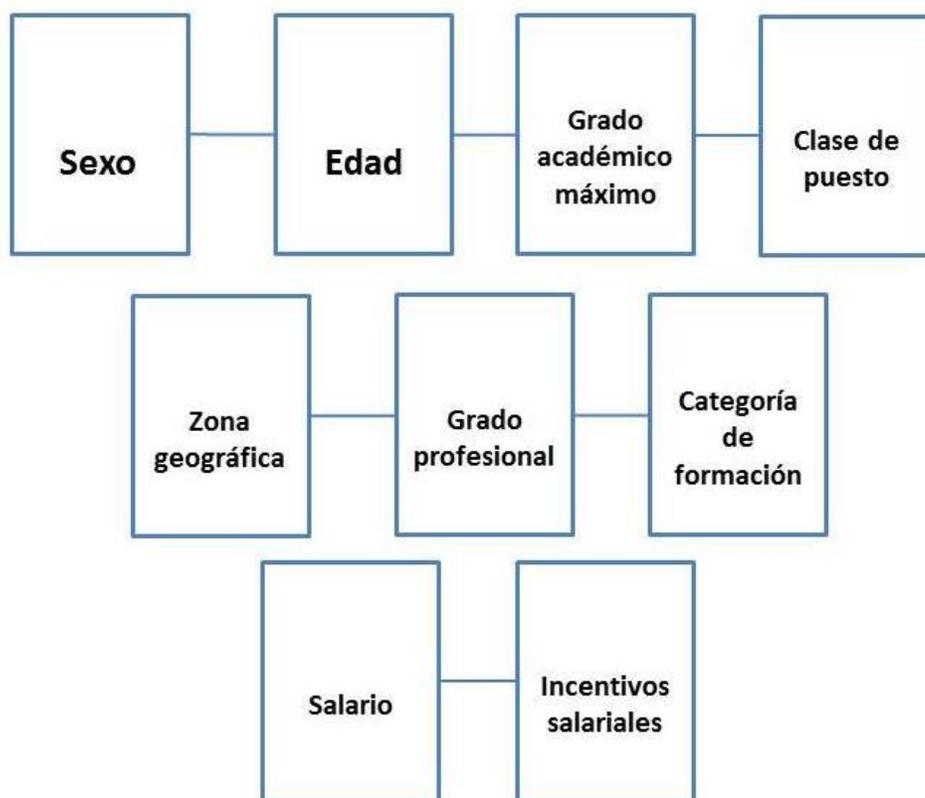
ANEXOS

ANEXO 1: Características de las bases de datos utilizadas.

Para el estudio se utilizó la base de datos suministrada por el Programa Estado de la Educación sobre docentes de primaria en escuelas públicas para el año 2015 del Ministerio de Educación Pública (MEP).

Las variables utilizadas de la base de datos se muestran en el esquema siguiente:

Figura 1: Variables utilizadas para determinar el perfil del docente en primaria, 2015.



Elaboración propia con información del MEP y PEE, 2016.

Algunos aspectos importantes de la base de datos utilizada es la siguiente:

- La base cuenta con un total de 39.547 docentes de primaria. El 81,95% (32.408) son mujeres, un 17,39% (6.877) son hombres y sin información se presentaron 262 casos (0,66%).
- El 94,54% son docentes, 4,42% administrativos y un 1,04% técnicos.
- El 59,66% se encuentran en propiedad y el restante 40,34% son interinos.

- Respecto al grado profesional, la mayoría (49,87%) cuentan con un grado de enseñanza general básica titulada grado 6 (PT6) y en menor medida con grado PT5 (10,68%) y PT4 (4,30%). Otros casos, aunque de menor importancia son los titulados con enseñanza técnica profesional grado 6 (VT6) que representan el 4,53% y los VT5 con un 2,26%.
- De acuerdo a la clase de puesto, la mayoría son profesores de enseñanza General Básica 1 (I y II ciclos) (G. de E.) los cuales agrupan el 52,62% de los docentes. En segundo lugar se ubican los docentes de Enseñanza Preescolar (G.de E) concentrando el 16,19% y en menor medida los docentes de enseñanza técnico profesional (Enseñanza Preescolar, o, I y II ciclos) (G. de E.) con el 9,85%, los docentes de idioma extranjero (I y II ciclos) (G.de E.) con el 7,28% y los docentes de enseñanza unidocentes (I y II ciclos) con el 4,62%.
- En lo que respecta a la edad del docente, el 10,98% tiene 30 años o menos, el 34,54% tiene entre 31 y 40 años, el 30,82% entre 41 y 50 años y el 23,67 más de 50 años.
- El 86,56% labora los 12 meses del año, un 3,03% 11 meses, 4,59% entre 10 y 8 meses y el restante 5,82% menos de 8 meses.
- Existen diferencias salariales según características del docente. El cuadro 3 muestra estos valores (se filtró solo para docentes que trabajan ocho meses o más) :

Cuadro 3: *Salario bruto según características del docente, 2015.*

Variable	Categoría	Salario Mensual en colones
Edad promedio del maestro	Menos de 30 años	494.757
	Entre 30 y 40 años	967.604
	Entre 41 y 50 años	1.155.446
	Más de 50 años	949.692
Titulación del maestro (PT6)	PT6	1.214.515
	Resto de PT's	743.508
Condición Laboral del maestro	Interinos	785.752
	Propiedad	1.205.503
	Aspirantes	351.221
Zona de la Escuela	Escuela Urbana	1.016.239
	Escuela Rural	1.110.796
Tamaño de la Escuela	Menos de 90 estudiantes	1.100.621
	Entre 90 y 400 estudiantes	1.068.996
	Más de 400 estudiantes	997.334

Elaboración propia con información del MEP y PEE, 2016.

ANEXO 2: Metodología para la caracterización de los docentes y su distribución espacial

La metodología empleada trata de responder a la siguiente interrogante: ¿Las características de los docentes relacionadas con su formación, experiencia, situación laboral, salarios e incentivos se distribuyen de manera homogénea en las escuelas del país? De no ser así, ¿cuál es esa distribución?, ¿presenta patrones de concentración en algunas zonas o cantones?

La metodología utilizada consiste en la evaluación de la aglomeración de las unidades geográficas de observación (escuelas) en torno a los valores exhibidos por un atributo de interés (características del docente), entendido éste como una variable que describe características específicas de estas unidades.

Normalmente los métodos estadísticos tradicionales no involucran coordenadas geográficas ni distancias físicas y asumen que no hay interacción entre las unidades de observación, es decir, que los valores de los atributos son independientes entre sí. Cuando se realiza inferencia estadística se asume que hay aleatoriedad en el comportamiento de los valores de los atributos, comúnmente dicha aleatoriedad se asocia a una distribución estadística que da origen a los datos (una distribución normal, una poisson,).

Sin embargo, es importante cuestionarse: ¿Qué supone el análisis estadístico tradicional para estudiar unidades geográficas? El análisis estadístico tradicional utilizado para analizar variables asociadas a un espacio geográfico, se caracteriza por construirse sobre supuestos en los que no hay auto-correlación entre las unidades de observación. La dependencia espacial significa que los valores de una misma variable, que son medidos en localidades que son cercanas entre sí, tienden a ser similares; es decir, a mayor cercanía geográfica corresponde una mayor similitud en los valores. Esto significa que la dependencia espacial se produce cuando el valor de la variable dependiente en una unidad espacial es parcialmente función del valor de la misma variable en unidades vecinas. La dependencia espacial hace diferente a la estadística espacial de la tradicional debido a que considera la distancia y la contigüidad existente entre unidades de observación.

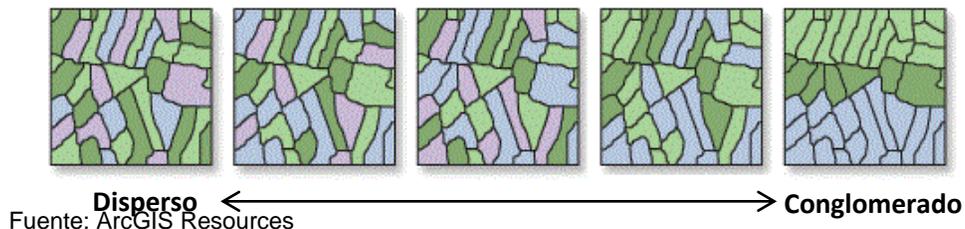
Dado lo anterior, la metodología propuesta, considera la dependencia espacial entre las unidades de observación. Con ello, las características geográficas asociadas a las características del docente en cada escuela y su distribución espacial pueden ser analizadas a través de técnicas que consideren la dependencia espacial y la auto-correlación espacial.

La idea es analizar si las características de los docentes de primaria se distribuyen de forma sistemática en el espacio geográfico, en cuyo caso se probaría la hipótesis de existencia de auto-correlación espacial de la variable. El efecto que

podemos encontrar al evaluar la auto-correlación espacial puede ser de signo positivo o negativo, así como nulo.

- a) Será **positivo** cuando se presentan asociaciones de valores similares entre características de docentes y localizaciones cercanas (escuelas); es decir, cuando, en el espacio geográfico, los valores altos de una variable están rodeados por valores altos y viceversa. Éste sería el caso, del llamado efecto contagio o desbordamiento (“*spillover*”) que se produce en muchos fenómenos socioeconómicos, en general, en los que su presencia en una región es causa de su extensión a regiones vecinas, favoreciendo la concentración del fenómeno en la zona (ver recuadro de la derecha en la figura 1).
- b) Se presenta auto-correlación espacial **negativa** en un espacio cuando los valores altos de una variable (características del docente) se encuentran rodeados por valores bajos de la misma, y viceversa. Esta configuración, en la que se produce una mayor disimilitud entre unidades geográficas cercanas que entre las lejanas, por ejemplo, la que se produciría en fenómenos de jerarquías espaciales del tipo centro-periferia (ver recuadro de la izquierda en la figura 15).

Figura 15: Posibles efectos de auto-correlación espacial
(Izquierda efecto negativo y derecha efecto positivo)



- c) Por último, se produce **ausencia** de auto-correlación espacial en una variable geográfica cuando ésta se distribuye de manera aleatoria sobre el espacio

Contrastes de auto-correlación espacial global

La perspectiva global del fenómeno de auto-correlación espacial tiene por objeto el contraste de la presencia de tendencias o estructuras espaciales generales en la distribución de una variable (características de los docentes en primaria) sobre un espacio geográfico completo (Costa Rica). En otras palabras, se trata de contrastar la hipótesis de que ciertas características de los docentes en primaria pública se encuentren distribuidos de forma totalmente aleatoria en todo el territorio costarricense, o si, por el contrario, existe algún tipo de asociación significativa de valores similares entre escuelas vecinas. Para efectos del estudio, se estimará el “test” I de Moran (1948), del cual se pasa a detallar su formulación matemática:

Siguiendo a Cliff y Ord (1973 y 1981) y Anselin (1995), el “test” I de Moran fue inicialmente formulado como función de una variable (Y), considerada en los puntos del espacio (i,j), en desviaciones a la media, y los elementos de la matriz binaria de interacciones espaciales [δ_{ij}]. Esta expresión inicial de Moran podría ser generalizada, sustituyendo la matriz de interacciones por la más general matriz de pesos espaciales, W_{ij} de la siguiente forma:

$$I = \frac{N}{S_0} \frac{\sum_{(2)} w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

Donde

W_{ij} : son los elementos de la matriz de pesos espaciales correspondientes al par (i, j)

$S_0 = \sum_i \sum_j w_{ij} = \sum_{(2)} w_{ij}$ es la suma de los pesos espaciales.

\bar{y} : es el valor medio o esperado de la variable y.

N: número de observaciones o tamaño muestral.

Al utilizar una matriz de interacciones espaciales estandarizada por filas, que es la situación óptima de aplicación de este “test”, el término $S_0 = N$, dado que la suma de los valores de cada fila es igual a la unidad. De este modo, el estadístico I queda reducido al cociente del producto espacial cruzado de los valores de la variable partido por la varianza:

$$I = \frac{\sum_{(2)} w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

En este caso I está basado en los productos cruzados de las desviaciones de Y_i respecto de \bar{y} . También es evidente que el “test” I de Moran es similar al coeficiente de auto-correlación temporal: el término del numerador es una medida de la covarianza entre valores de Y en dos localizaciones distintas (i,j) y el denominador expresa la varianza de Y en el punto i. Sin embargo, aunque parecido, el test I no es equivalente al clásico coeficiente de correlación, fundamentalmente porque no se encuentra centrado en el valor cero. De hecho, la media teórica de la I de Moran es el cociente $\frac{-1}{N-1}$.

Dado lo anterior, el valor esperado de I es negativo y es función únicamente del tamaño de la muestra (N), aunque esta media tiende a cero a medida que el tamaño de la muestra aumenta. En cuanto a la varianza teórica del coeficiente I, se verá más adelante que depende de determinados supuestos estocásticos. Un coeficiente I de Moran mayor que su valor esperado sería indicativo de auto-correlación espacial positiva, mientras que un valor de I inferior a la media pondría de manifiesto la existencia de auto-correlación espacial negativa. Este estadístico

estará muy afectado por aquellos puntos vecinos sensiblemente distintos de la media de la variable en estudio.

Según Cliff y Ord (1981) respecto a la distribución del contraste I , cuando el tamaño muestral es suficientemente amplio, la expresión estandarizada del “test” I se distribuye como una normal tipificada, $N(0,1)$, como el “test” de recuento de vínculos de Moran. Por eso, también suele considerarse el estadístico inicial I , el proceso inferencial suele utilizar los valores estandarizados (z) de cada uno de ellos, obtenidos a través del cociente entre la diferencia del valor inicial y la media teórica, y la desviación típica teórica, de la manera siguiente:

$$z_I = \frac{I - E[I]}{SD[I]}$$

Siendo $E[I]$: La media teórica del estadístico I y $SD[I]$ la desviación típica del estadístico I . La interpretación de los valores estadísticamente significativos de la variable tipificada Z_I sería la siguiente:

- I. -Valores no significativos del test I estandarizado, Z_I , correspondiente a una variable Y , conducirían a aceptar la hipótesis nula de no auto-correlación espacial o inexistencia de patrones de comportamiento de dicha variable sobre el espacio.
- II. Valores significativos de $Z_I > 0$ serían indicativos de auto-correlación espacial positiva, es decir, que es posible encontrar valores parecidos (altos o bajos) de la variable Y , espacialmente agrupados, en mayor medida de como estarían por casualidad.
- III. Valores significativos de $Z_I < 0$ serían indicativos de auto-correlación espacial negativa, es decir, que se produce una no-agrupación de valores similares (altos o bajos) de la variable Y superior a lo normal en un patrón espacial aleatorio. Se trata de un concepto algo más difícil de captar.

Cuadro 4: Interpretación de los valores estandarizados de los estadísticos de auto-correlación espacial global

Test	Hipótesis nula (z no significativo)	Hipótesis alternativa (z significativo) $Z > 0$	Hipótesis alternativa (z significativo) $Z < 0$
I Moran	No auto-correlación espacial	Auto-correlación espacial POSITIVA	Auto-correlación espacial NEGATIVA

Fuente: Arias y Sánchez (2014)

Resultados del contraste de auto-correlación espacial global

Como se mencionó anteriormente la herramienta Autocorrelación espacial (I de Moran global) mide la autocorrelación espacial basada en las ubicaciones y los valores de las entidades simultáneamente, en este caso de las escuelas públicas.

Dado un conjunto de escuelas y un atributo asociado (perfil del docente), evalúa si el patrón expresado está agrupado, disperso o es aleatorio. La herramienta calcula el valor del Índice I de Moran y una puntuación z y un valor P para evaluar la significancia de ese índice. Los valores P son aproximaciones numéricas del área debajo de la curva de una distribución conocida, limitada por la estadística de prueba.

La herramienta calcula el valor medio y la varianza para cada característica del docente que se está evaluando. Seguidamente, resta el valor medio en cada valor de la escuela, lo que crea una desviación del valor medio. Los valores de desviación para todas las escuelas vecinas (por ejemplo, las escuelas dentro de la banda de distancia especificada) se multiplican de forma conjunta para crear un producto cruzado. El numerador para la estadística I de Moran global suma estos productos cruzados.

Para entender el indicador supongamos que las escuelas A y B son vecinas y que el valor medio para todos los valores de las escuelas en deserción es 10%.

Valores de las escuelas		Desviaciones		Productos cruzados
A=50	B=40	40	30	1200
A= 8	B=6	-2	-4	8
A=20	B=2	10	-8	-80

Cuando los valores para las escuelas vecinas son mayores que el valor medio o menores que el valor medio, el producto cruzado será positivo. Cuando un valor es menor que el valor medio y el otro es mayor que el valor medio, el producto cruzado será negativo. En todos los casos, mientras mayor sea la desviación del valor medio, mayor será el resultado del producto cruzado.

Si los valores en la base de datos tienden a agruparse espacialmente (los valores altos se agrupan cerca de otros valores altos; los valores bajos se agrupan cerca de otros valores bajos), el Índice de Moran será positivo. Cuando los valores altos rechazan otros valores altos y tienden a estar cerca de valores bajos, el Índice será negativo. Si los valores positivos de los productos cruzados equilibran los valores negativos de los productos cruzados, el Índice será cercano a cero. El numerador está normalizado por la varianza de modo que los valores del Índice oscilan entre -1,0 y +1,0.

Después de calcular el valor del Índice, la herramienta Autocorrelación espacial (I de Moran global) calcula el valor del Índice esperado. Después se comparan los valores del Índice esperado y del Índice observado. Dado el número de entidades en la base de datos y la varianza del total de los valores de datos, la herramienta calcula una puntuación z y un valor P que indican si esta diferencia es estadísticamente significativa. Los valores del índice no se pueden interpretar directamente; sólo se pueden interpretar dentro del contexto de la hipótesis nula.

El cuadro siguiente muestra los resultados de la I de Moran para las 20 variables analizadas. Se utilizaron diferentes variantes para las estimaciones (Inverso de la distancia, inverso de la distancia al cuadrado y fixed distance band).

Cuadro 5: Estimación de los estadísticos de auto-correlación espacial global

Características del docente	I de Moran					
	Inverso distancia		Inverso distancia cuadrado		Fixed Distance Band	
	Índice	Valor P	Índice	Valor P	Índice	Valor P
Salario bruto	0,032	0,000	0,032	0,000	0,028	0,000
Salario base	0,053	0,000	0,053	0,000	0,012	0,000
Incentivos	0,047	0,000	0,047	0,000	0,030	0,000
Docentes menores de 30 años	0,000	0,900	0,000	0,900	0,000	0,965
Docentes de 30 a 40 años	0,030	0,000	0,030	0,000	0,012	0,000
Docentes de 40 a 50 años	0,073	0,000	0,073	0,000	0,022	0,000
Docentes mayores de 50 años	0,069	0,000	0,069	0,000	0,016	0,000
Docentes con grado profesional VT6	0,110	0,000	0,110	0,000	0,159	0,000
Docentes con grado profesional PT6	0,030	0,000	0,030	0,000	0,014	0,000
Docentes con grado profesional KT3	0,070	0,000	0,070	0,000	0,021	0,000
Docentes con grado profesional PT5	-0,002	0,794	-0,002	0,790	0,013	0,000
Docentes con grado profesional PT4	-0,004	0,464	-0,004	0,462	0,314	0,000
Docentes con grado profesional PT3	-0,004	0,480	-0,004	0,479	0,004	0,001
Docentes con grado profesional KT2	0,000	0,928	0,000	0,931	0,011	0,000
Docentes Interinos	0,445	0,000	0,446	0,000	0,075	0,000
Docentes Aspirante	-0,313	0,000	-0,313	0,000	-0,046	0,000
Salario pagado por IDS	0,066	0,000	0,068	0,000	0,042	0,000
Salario pagado por zonaje	0,101	0,000	0,115	0,000	0,024	0,000
Docentes con IDS	0,037	0,093	0,036	0,086	0,199	0,091
Docentes con Zonaje	0,124	0,069	0,123	0,073	0,348	0,078

Elaboración propia con información del MEP y PEE, 2016.

La I de Moran es una estadística deductiva, lo que significa que los resultados del análisis siempre se interpretan dentro del contexto de la hipótesis nula. En este caso la hipótesis nula establece que las diferentes características del docente de primaria evaluadas están distribuidas en forma aleatoria entre las escuelas del país; es decir, los procesos espaciales que promueven el patrón de valores observado constituyen una opción aleatoria. Para entender mejor el punto anterior supongamos que se pudiera elegir a todos los docentes y arrojarlos sobre las escuelas y dejar que cada profesor caiga donde caiga. Este proceso (elegir y arrojar los docentes) es un ejemplo de un proceso espacial de opción aleatoria.

Cuando el valor P del cuadro anterior es estadísticamente significativo, puede rechazar la hipótesis nula. El siguiente cuadro resume la interpretación de los resultados:

Cuadro 6: Interpretación de los estadísticos de auto-correlación espacial global

criterio	Hipótesis	Interpretación	Variables Analizadas
El valor P no es estadísticamente significativo.	Las características analizadas de los docentes están distribuidas en forma aleatoria entre las escuelas públicas del país.	No puede rechazar la hipótesis nula. La distribución espacial de los docentes con grado profesional PT5, PT4, PT3 y KT2, docentes menores de 30 años y los recargos salariales que reciben por Zonaje e IDS (en ambos casos es la tenencia o no del recargo y no el monto en dinero) en las escuelas públicas sea el resultado de procesos espaciales aleatorios. Es decir, no se logró determinar conglomerados a nivel global.	Docentes con título PT5, PT4, PT3 y KT2
			Docentes menores de 30 años
			Recargo por Zonaje
			Recargo por IDS
El valor P es estadísticamente significativo y la puntuación z es positiva.	Las características analizadas de los docentes están distribuidas en forma aleatoria entre las escuelas públicas del país.	Puede rechazar la hipótesis nula. La distribución espacial de los valores altos y los valores bajos en la base de datos sobre la edad de los docentes (exceptuando los menores de 30 años), los docentes con grado profesional de PT , VT y KT3 así como los interinos están más agrupados espacialmente en ciertas escuelas públicas de lo que se esperaría si los procesos espaciales subyacentes fueran aleatorios. Mismo patrón sucede para el monto pagado en el salario base como bruto, así como los monton de dinero pagados por zonaje e IDS. Se puede deducir que existen patrones de concentración (conglomerados) de escuelas públicas para los atributos de los docentes mencionados.	Docentes con título PT6, VT6 y KT3
			Docentes de 30 a 40 años, de 40 a 50 años y mayores de 50 años
			Docentes interinos
			El salario base y bruto, monto pagado por IDS y Zonaje
El valor P es estadísticamente significativo y la puntuación z es negativa.	Las características analizadas de los docentes están distribuidas en forma aleatoria entre las escuelas públicas del país.	Puede rechazar la hipótesis nula. La distribución espacial de los valores altos y los valores bajos en la base de datos sobre los docentes aspirantes está más dispersa espacialmente de lo que se esperaría si los procesos espaciales subyacentes fueran aleatorios. Este patrón espacial disperso puede estar reflejando las particularidades de los docentes interinos que principalmente se concentran en distritos de Talamanca y Buenos Aires con un peso marginal muy bajo en otras zonas del país.	Docentes Aspirantes

Elaboración propia con información del MEP y PEE, 2016.

3.3 Contrastes de auto-correlación espacial Local

De acuerdo con varios autores (Getis y Ord (1992); Openshaw (1993); Anselin (1993) y (1995); Ord y Getis (1995) y (2001); Vayá y Suriñach (1996); Tiefelsdorf y Boots (1997) y Sokal (1998)) los estadísticos de auto-correlación global, centrados en el análisis de dependencia general propia de todas las unidades de un espacio geográfico, no son capaces de detectar la inestabilidad o deriva espacial de ciertas estructuras locales de asociación o inestabilidades locales que pueden estar, a su vez, presentes o no en una estructura global de dependencia. Para Vayá y Suriñach (1996) el problema de la dependencia espacial local puede plantearse desde dos puntos de vista

- Existe la posibilidad de que, en un espacio dado, no se detecte la presencia de auto-correlación espacial global en la distribución de una variable, aunque, de hecho, existan pequeños “clusters” espaciales en los que dicha variable experimenta una concentración (o escasez) importante.
- Existe también la posibilidad de que, habiéndose detectado dependencia a nivel global en una variable, no todas las escuelas del espacio considerado contribuyan con igual peso en el indicador global, es decir, que coexistan unas zonas en las que la variable se distribuya de forma aleatoria junto a otras con una importante contribución a la dependencia existente.

Para analizar lo anterior en las escuelas de Costa Rica, se definió un contraste de asociación local que indica hasta qué punto una escuela se encuentra rodeada por otras con valores altos o bajos de las características del maestro que se evaluó, para ello se utilizó uno de los indicadores locales de asociación espacial (LISA) (Estadístico Local de Moran), propuestos por Anselin (1995).

El estadístico Local de Moran se puede definir como:

$$I_i = z_i \sum_{j=1}^{J_i} w_{ij} z_j$$

donde z_i , z_j son variables y_i estandarizadas \sum_j es la sumatoria que unánimemente incluye los valores vecinos a i : $j \in J_i$

Para lograr una mejor interpretación de este estadístico, la matriz de pesos w_{ij} debería estar estandarizada por filas (aunque no es necesario), siendo, por convenio, cada elemento $w_{ii} = 0$. También, en este caso, es fácil comprobar que la suma de estadísticos locales I_i es el “test” I de Moran:

$$\sum_i I_i = \sum_i z_i \sum_j w_{ij} z_j = \sum_i \sum_j w_{ij} z_i z_j$$

La ecuación anterior se puede expresar como:

$$I = \frac{N}{S_0} \frac{\sum_{(2)} w_{ij} z_i z_j}{\sum_{i=1}^n z_i^2} \Rightarrow I = \frac{\sum_i I_i}{S_0 \sum_i \frac{z_i^2}{N}}$$

Donde $\sum_{(2)} = \sum_i \sum_j$

De este modo, se cumple que el sumatorio de valores del estadístico local I_i es equivalente al test global I , siendo el factor de proporcionalidad $\gamma = S_0 m_2$:

$$\sum_i I_i = \frac{1}{S_0 m_2} I = \gamma \cdot I$$

donde: $m_2 = \sum_i \frac{z_i^2}{N}$, momento de 2º orden de la variable z_i .

En el caso de que la matriz W esté estandarizada por filas, $S_0 = N$, de manera tal que el factor $\gamma = \sum_i z_i^2$ y, en el caso de trabajar con variables estandarizadas, $m_2 = 1$, lo que daría que dicho factor $\gamma = S_0$. De esta manera el “test” I local de Moran podría también expresarse del modo siguiente:

$$I_i = \frac{z_i}{m_2} \sum_j w_{ij} z_j$$

Anselin (1995), plantea que es posible calcular los momentos de I_i , bajo la hipótesis nula de ausencia de asociación espacial, para el supuesto inferencial de aleatoriedad o muestreo aleatorio. Cualquier contraste de significación de asociación espacial local puede basarse en estos momentos, aunque la distribución exacta de un estadístico de este tipo aún se desconoce. Para el autor, una forma de facilitar su interpretación es utilizando una distribución normal. De acuerdo con Anselin (1995) el método anterior puede ofrecer, mediante su representación cartográfica y el diagrama de dispersión de Moran, información sobre conglomerados y puntos atípicos (outliers) de unidades con presencia alta de miembros de un grupo.

Utilizando el mapa de la significación de los indicadores locales asociado al diagrama de dispersión podemos identificar zonas con presencia alta de miembros de un grupo rodeadas de zonas con presencia también alta (situación High-High en el diagrama de dispersión de Moran), o bien zonas con presencia alta rodeadas de unidades con presencia baja (situación “High-Low” en el diagrama de dispersión de Moran). De igual forma permite la detección de zonas con presencia baja rodeadas de unidades también con presencia baja (situación “Low-Low”), o bien zonas de presencia baja rodeadas de unidades con presencia alta de la variable analizada (situación “Low-High”). Finalmente, se pueden detectar zonas sin asociación espacial significativa.

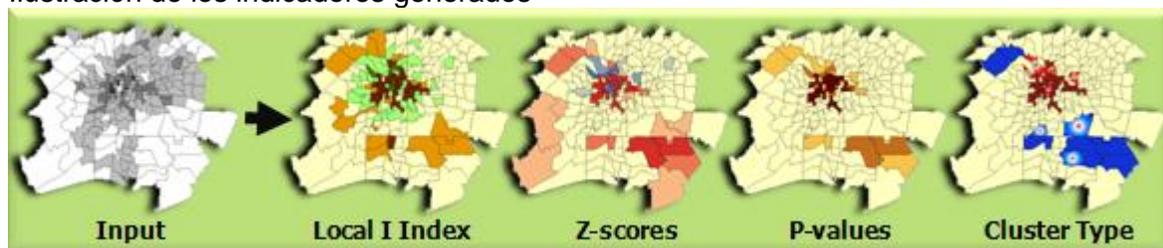
El análisis de este estudio se centró en aquellas escuelas donde hay presencia significativa de alguna característica del docente, que se encuentran rodeadas de escuelas con igual situación o bien, una escuela rodeada por otra escuela con situaciones diferentes; en ambos casos nos referiremos a ellas como zonas cluster (conglomerado).

Resultados del contraste de auto-correlación espacial local

Como se mencionó anteriormente, en este estudio se analiza un conjunto de entidades (escuelas públicas) y un campo de análisis (características del docente) mediante la técnica geo-estadística de Análisis de cluster y de valor atípico (I Anselin local de Moran). Esta herramienta de análisis de cluster y de valor atípico identifica clusters espaciales de las entidades con valores altos o bajos. La herramienta también identifica los valores atípicos espaciales. Para realizar esto, la herramienta calcula un valor I de Moran local, una puntuación z, un valor P y un código que representa la entidad estadísticamente significativa. Las puntuaciones z y los valores P representan la significancia estadística de los valores de índice calculados (ver formulación matemática en párrafos anteriores).

Figura 16

Ilustración de los indicadores generados



Fuente: ArcGIS Resources

Las puntuaciones z y los valores p son medidas de significancia estadística que indican si se rechazará la hipótesis nula, escuela por escuela. En efecto, indican si la aparente similitud (un clustering espacial de valores altos o bajos) o la falta de similitud (un valor atípico espacial) es más marcada de lo que se espera en una distribución aleatoria. La puntuación z está basada en el cálculo de la hipótesis nula de aleatorización.

Una puntuación z positiva alta para una escuela (en una característica del docente analizada) indica que escuelas circundantes tienen valores similares (ya sea valores altos o bajos). El campo COType en la Clase de entidad de salida escuela (en una característica del docente analizada) será HH (Alto/Alto) para un cluster de valores altos estadísticamente significativo y LL (Bajo/Bajo) para un cluster de valores bajos estadísticamente significativo.

Una puntuación z negativa baja (por ejemplo, inferior a -3,96) para una escuela indica un valor atípico de datos espacial estadísticamente significativo. El campo COType en la Clase de entidad de salida indicará si la escuela tiene un valor alto y está rodeada por escuelas con valores bajos HL (Alto/bajo) o si la escuela tiene un valor bajo y está rodeada por escuelas con valores altos LH (Bajo/alto).

El campo COType indicará siempre unos clusters y valores atípicos estadísticamente significativos para un nivel de confianza del 95 por ciento.

Solamente las escuelas estadísticamente significativas tienen valores para el campo COType.

Resultados del contraste de auto-correlación espacial local para profesores en condición de interinato

Se identificaron 10 conglomerados de valores altos de docentes en condición de interinato en escuelas públicas del país y 7 conglomerados de valores bajos.

Cuadro 7: *Ubicación a nivel cantonal de los clústeres de escuelas según condición de interinato del docente en 2015.*

Conglomerado	Conglomerados de escuelas (profesores interinos)	
	Valores altos	Valores bajos
1	Aguirre; Garabito, Parrita	Cantones de la Región Chorotega (menos Abangares)
2	Puntarenas	Cantones de la GAM
3	Osa	Pérez Zeledón
4	Golfo	Coto Brus
5	Buenos Aires	Upala
6	Limón y Talamanca	San Carlos
7	Turrialba	Sarapiquí
8	Abangares	
9	Los Chiles	
10	Guatuso	

Elaboración propia con información del MEP y PEE, 2016

Resultados del contraste de auto-correlación espacial local para docentes con grado profesional PT6

Se identificaron 11 conglomerados de valores altos de docentes con grado profesional PT6 en escuelas públicas del país y 6 conglomerados de valores bajos.

Cuadro 8

Ubicación a nivel cantonal de los clústeres de escuelas según grado profesional PT6 del docente en 2015.

Conglomerado	Conglomerados de escuelas (Docentes con grado profesional PT6)	
	Valores altos	Valores bajos
1	Nandayure, Hojancha y Puntarenas (Península de Nicoya)	Carrillo, San Cruz, Nicoya y Bagaces
2	Puntarenas, Montes de Oro y Abangares	Sarapiquí y Pococí
3	Parrita	Limón y Talamanca
4	Garabito, San Mateo, Orotina y Turubares	Coto Brus (escuelas de menos de 90 alumnos)
5	Pérez Zeledón y Aguirre	Sur de San Carlos, Alfaro Ruiz y Valverde Vega
6	Osa	Cantones en el Norte Este y Sur Este de la GAM
7	Golfito y Corredores	
8	Coto Brus (Escuelas de 91 a 400 alumnos)	
9	Upala	
10	Guatuso y San Carlos	
11	Cantones de la GAM	

Elaboración propia con información del MEP y PEE, 2016

Resultados del contraste de auto-correlación espacial local para docentes según edad.

Se identificaron 7 conglomerados de valores altos y 9 conglomerados de valores bajos.

Cuadro 9

Ubicación a nivel cantonal de los clústeres de escuelas según edad del docente en 2015.

Conglomerado	Conglomerados de escuelas (Docentes según edad)	
	Valores altos	Valores bajos
1	Cantones de la región Chorotega	Los Chiles, San Carlos y Guatuso
2	Zona Occidentes región Central (Palmares, Atenas, Grecia, San Ramón, Naranjo, Valverde Vega y Alfaro Ruiz.	Sarapiquí y Pococí
3	Cantones de la GAM	Osa
4	Turrialba (zona urbana)	Golfito
5	Tarrazú, Acosta y Dota	Buenos Aires y Coto Brus
6	Puntarenas, Esparza, San Mateo y Montes de Oro	Limón y Talamanca
7	La Cruz y Upala	Turrialba (zona rural)
8		Aguirre
9		Garabito

Elaboración propia con información del MEP y PEE, 2016

Resultados del contraste de auto-correlación espacial local para docentes aspirantes.

Se identificaron 3 conglomerados de valores altos de docentes en condición de aspirantes en escuelas públicas del país y 6 conglomerados de valores bajos.

Cuadro 10

Ubicación a nivel cantonal de los clústeres de escuelas según el número de docentes en condición de aspirantes en 2015.

Conglomerado	Conglomerados de escuelas (Docentes aspirantes)	
	Valores altos	Valores bajos
1	Turrialba	Upala
2	Talamanca, Matina y Limón	Bagaces
3	Buenos Aires	Pococí
4		Turrialba
5		San Carlos, Alfaro, Ruiz, Valverde Vega, Naranjo, Grecia, Poás,
6		Cantones de la GAM

Elaboración propia con información del MEP y PEE, 2016

Resultados del contraste de auto-correlación espacial local para los salarios de los docentes.

El cuadro 9 muestra los conglomerados identificados por cantón.

Cuadro 11

Ubicación cantonal de los clústeres de escuelas según el salario promedio pagado a los docentes de acuerdo al tamaño del centro educativo en 2015.

Conglomerado	Conglomerados de escuelas según salario					
	Escuelas de menos de 90 alumnos		Escuelas entre 90 y 400 alumnos		Escuelas de más de 400 alumnos	
	Valores altos	Valores bajos	Valores altos	Valores bajos	Valores altos	Valores bajos
1	Carrillo, Cañas y Liberia	Pococi, Sarapiquí y Limón	San Ramón, Palmares y Atenas	Nicoya	Pérez Zeledón	Pérez Zeledón (zonas rurales)
2	Upala, Guatuso y San Carlos	Pérez Zeledón, Osa, Golfito y Corredores (zonas rurales)	Cantones Área Metropolitana de San José, Cartago y Heredia	Garabito, Orotina y Turrubares	Aguirre	Nicoya
3	Puntarenas y Garabito (zona urbana)	Alajuela, Grecia y Poás	Turrialba, Pococi y Siquirres	Pérez Zeledón (zona rurales)	Cantones del Área Metropolitana de Cartago	Carrillo y Bagaces
4	Pérez Zeledón, Osa, Golfito y Corredores (zonas urbanas)	Garabito y Aguirre		Los Chiles, San Carlos y Guatuso (zonas rurales)	Pococi	San Carlos (zonas rurales)
5	Pococi, Sarapiquí y Limón	Tilaran		La Cruz y Upala		
6	Cantones Área Metropolitana de San José y Heredia	Los Chiles				

Elaboración propia con información del MEP y PEE, 2016

Resultados del contraste de auto-correlación espacial local para el monto salarial pagado por zona y recargos a los docentes.

El cuadro 10 muestra los conglomerados identificados por cantón.

Cuadro 12

Ubicación a nivel cantonal de los clústeres de escuelas según salario promedio pagado a los docentes por concepto de zonaje y recargos en 2015

Conglomerado	Conglomerados de escuelas (zonaje)		Conglomerados de escuelas (recargos)	
	Valores altos	Valores bajos	Valores altos	Valores bajos
1	Cantón de Puntarenas y Nicoya (Península)	Cantones de la Región Chorotega	La zona urbana de San Carlos y el cantón de Sarapiquí	Cantones de la región Chorotega menos Hojanca y Nandayure.
2	Cantones región Central (menos cantones de la periferia sur)	Cantones de la Región Huetar Norte menos Sarapiquí.	Los cantones de la región Central	Cantones de la región Huetar Norte menos Sarapiquí.
3	San Carlos	Garabito, San Mateo y Orotina	Algunos distritos de los cantones de Buenos Aires, Coro Brus y Corredores.	Garabito, San Mateo y Orotina
4	Cantones región Atlántica	Acosta, Tarrazú, León Cortes, Dota y Pursical		Acosta, Tarrazú, León Cortes, Dota y Pursical
5	Coto Brus	Pérez Zeledón y Aguirre		Pérez Zeledón y Aguirre

Elaboración propia con información del MEP y PEE, 2016

ANEXO 3: Metodología para estimar los efectos de las características de los profesores, del centro educativo y del entorno geográfico en el cual se localizan las escuelas sobre la deserción, aprobación y repitencia.

La metodología busca responder a las siguientes preguntas:

- ¿Qué implicaciones tiene sobre los porcentajes de deserción, aprobación y repitencia en escuelas el hecho de que no exista una distribución homogénea de las características de los profesores?
- ¿Existe relación entre los salarios, recargos e incentivos asignados a los docentes y los porcentajes de repitencia, aprobación y deserción que presentan?
- ¿Existe relación entre las características de la escuela (tamaño, tipo, infraestructura, etc) y los porcentajes de repitencia, aprobación y deserción que presentan?
- ¿Existe alguna relación entre altos porcentajes de repitencia en la escuela y altos niveles en deserción?
- ¿Influyen las características de localización de la escuela (entorno) sobre los porcentajes de repitencia, aprobación y deserción?

El análisis se realiza solo para las escuelas públicas. Se desarrollan modelos econométricos con Mínimos Cuadrados Generalizados para estimar las relaciones. El análisis no consiste en abordar una función de producción educativa, sino, tratar de identificar relaciones estadísticas entre las características del docente descritas en las secciones anteriores y los valores promedios de repitencia, aprobación y deserción, incorporando su entorno espacial. A pesar de ello, se toman en consideración variables relacionadas con el docente, con el centro educativo e indirectamente con las características del hogar al incluir las zonas de pobreza (bajo clima educativo y bajos ingresos), así como la distancia del centro educativo al principal centro urbano que introduce el tema urbano-rural, accesibilidad y acceso a bienes y servicios. Se analizaron 25 variables introduciendo en los modelos solo aquellas que resultaron significativas y tenían una fuerte relación con la teoría.

Fuentes de información y variables utilizadas

Para el desarrollo del modelo se incorporan cuatro bases de datos. La primera es la base de datos de funcionarios del MEP, la segunda empleada es la base de datos de distancias entre el centro educativo y el centro urbano más cercano, la tercera es la Megabase de datos de escuelas del MEP y la cuarta es la base de concentraciones de pobreza. A continuación, se esbozan las variables evaluadas:

a) Características del Docente

1. Sexo del docente
2. Edad
3. Grado Profesional
4. Tipo de Nombramiento

5. Salario
6. Incentivos y recargos

b) Variables del centro educativo y su entorno

1. Tamaño (número de estudiantes)
2. Alumnos por centro educativo
3. Infraestructura (gimnasio, comedor, biblioteca, taller y laboratorio de informática).
4. Escuelas que imparten inglés.
5. Servicios (agua, internet y electricidad)
6. Distancia de la escuela al centro urbano más próximo (en Km por carretera).
7. Ubicación del centro educativo según zona y región.
8. Ubicación del centro educativo según conglomerados de pobreza.
9. Salario Bruto y base promedio del centro educativo.
10. Incentivos y recargos promedios del centro educativo

c) Variables de rendimiento educativo

1. Aprobación
2. Repitencia
3. Deserción

Planteamiento del modelo

Como se mencionó anteriormente, se desarrollan 3 modelos econométricos para explorar la relación entre las características de los profesores y el centro educativo; respecto a los promedios de repitencia, aprobación y deserción que obtienen las escuelas del país. De esta manera, dadas las características de las variables dependientes (repitencia, aprobación y deserción) se plantean modelos econométricos con el método Mínimos Cuadrados Generalizados para las escuelas públicas.

La razón para proceder con este método, es porque en este caso, si bien los valores de la variable endógena (porcentaje de repitencia, aprobación y deserción) están acotados en el rango 0-1 (valores continuos) no es recomendable usar el método de MCO, dada la presencia de heterocedasticidad, en cuyo caso el modelo obliga a estimar por Mínimos Cuadrados Generalizados, para garantizar el cumplimiento de las propiedades de los parámetros estimados, utilizándose la inversa de la varianza de los errores como ponderación del modelo.

Dado que la variable a modelar es continua (acotada en el rango 0-1), la estimación del modelo podría realizarse mediante el procedimiento habitual utilizado para estimar regresiones lineales, para ello es necesario linealizar el modelo:

$$\ln\left(\frac{M_i}{1-M_i}\right) = \alpha + \beta_k X_{ki} + \varepsilon_i$$

donde ε_i es el valor de la perturbación aleatoria incluida en la especificación de todo modelo de regresión lineal y que cumple las hipótesis de perturbación esférica y ausencia de auto-correlación. El modelo así transformado puede estimarse por el procedimiento habitual de Mínimos Cuadrados Ordinarios (MCO). Sin embargo, y dado que el valor de M_i es desconocido y debe sustituirse por su estimación muestral P_i , el modelo a estimar quedaría como:

$$\text{Ln}\left(\frac{P_i}{1-P_i}\right) = \alpha + \beta_k X_{ki} + \varepsilon_i + \varepsilon_i'$$

donde ε_i' recoge el error cometido al utilizar la estimación muestral de la probabilidad P_i , en vez de su valor desconocido M_i . Al sustituir M_i por su estimación muestral P_i , los errores, supuestos independientes, cumplen la condición asintótica de normalidad exigida para realizar contrastaciones y construcción de intervalos de confianza, pero, dejan de cumplir la condición de homocedasticidad ya que su varianza no es constante¹⁰.

La presencia de heterocedasticidad impide la estimación a través de Mínimos Cuadrados Ordinarios, siendo necesario aplicar el método de Mínimos Cuadrados Generalizados, que sin exigir la condición de homocedasticidad de los errores, permite estimar estimadores ELIO. Este procedimiento transforma el modelo a estimar en otro, donde todas las variables quedan ponderadas por los inversos de las varianzas de los errores, y dado que se desconocen dichos valores verdaderos, éstos se sustituyen por su estimación muestral P_i , de donde:

$$s_i = \frac{1}{\widehat{\text{Var}}(\varepsilon_i')} = n_i P_i (1 - P_i)$$

quedando el modelo a estimar como:

$$s_i \text{Ln}\left(\frac{P_i}{1-P_i}\right)_i = \alpha s_i + \beta_k X_{ki} s_i + \varepsilon_i$$

Estadísticas descriptivas de las variables utilizadas en el modelo

La base de datos filtrada, con todas las escuelas con información disponible para las variables elegidas en el modelo, está compuesta por 3.712 observaciones (escuelas). Las variables incluidas se pueden agrupar en 4 categorías: las de rendimiento, las institucionales, las del profesor y las de distrito. A continuación se muestran las estadísticas descriptivas para tales variables. (Ver Cuadro 10).

¹⁰ La varianza de la perturbación aleatoria no es homocedástica ya que depende del nivel en que se encuentre la variable explicativa X , al definirse

$$\varepsilon_i' = \text{Ln}\left(\frac{P_i}{1-P_i}\right) - \text{Ln}\left(\frac{M_i}{1-M_i}\right)$$

En cuanto a las variables de rendimiento (variables dependientes en los modelos econométricos) se cuenta con: repitencia, aprobación y deserción. Respecto a repitencia, la media en las escuelas es de 5,5%, con una desviación estándar de 7,7%, mientras que en deserción el promedio es 4,1%, teniendo una desviación estándar de 7,5%. Resulta de interés notar que la variable deserción tiene una mediana de 0, es decir, más de la mitad de las escuelas no tiene deserción alguna, además de que la escuela con el máximo nivel de deserción tiene un porcentaje de 71% para tal variable. Por otro lado, la aprobación tiene una media de 90,8% con una desviación estándar del 11,9%.

Respecto a las variables institucionales, de media se tienen 15,46 estudiantes por aula, con una desviación estándar de 9,42, mientras que en promedio las escuelas cuentan con 2,48 de los 3 servicios básicos contemplados (agua, internet y electricidad) y con 1,35 de 5 edificaciones contempladas en escuelas (gimnasio, comedor, biblioteca, taller de artes industriales y laboratorio de informática). En el caso del tamaño, se usó una categorización de 0 a 10, donde 0 implica escuelas con menos de 100 estudiantes matriculado, 1 de 100 a 200 estudiantes, 2 de 200 a 300 estudiantes y así sucesivamente hasta 10 que incluye las escuelas con más de 1000 estudiantes matriculados; el tamaño promedio encontrado en las escuelas es de 1,73. Por otro lado, un 52,6% de las escuelas imparten inglés. (Ver Cuadro 9).

Las variables del profesorado incluyen la edad del profesor, salario, titulación y condición laboral. Iniciando con edad, el mayor porcentaje de escuelas de la base tiene en promedio un profesorado de entre 40 y 50 años (33,7%), seguido por entre 30 y 40 años (32,6%), mientras que los de menos de 30 años representan el menor porcentaje (11%), notándose además que la mediana en esta última categoría es de 0, por lo que más de la mitad de las escuelas tiene un profesorado con edad promedio mayor a 30 años.

El salario promedio de la media de salarios por escuela es de 1,2 millones de colones por mes, con un salario base promedio de 509 mil colones mensuales, y un promedio de incentivos salariales (sin contar IDS y zonaje) de 382 mil colones por mes. Además, en promedio 73,7% de los profesores en las escuelas cuentan con incentivo salarial de zonaje y un 67,6% cuenta con incentivos por IDS. En cuanto a titulación, el promedio es de 54,3% de maestros con un título de enseñanza general básica 6 (PT6), un 13,2% con título PT5 y un 4,2% con título PT4, mientras que un 1,9% tiene una titulación de enseñanza técnica profesional 6 (VT6) y un 12,1% de enseñanza preescolar 3 (KT3). (Ver Cuadro 11).

Por último, en cuanto a profesores, por condición laboral, el promedio de profesores interinos en las escuelas es de 47%, a la vez que un 8,3% tiene una condición de aspirante (interino sin título).

Finalmente, respecto a las variables distritales se tiene pobreza y zona rural, todas a nivel de distrito. En promedio un 39,4% de las escuelas se ubica en distrito en zona de pobreza. Asimismo, un 64,1% de las escuelas se encuentra en un distrito clasificado como zona rural por el MEP. (Ver Cuadro 11).

Cuadro 13
Estadísticas descriptivas de las variables utilizadas en los modelos

Variable	Media	Mediana	Desviación estándar	Rango	Mínimo	Máximo
<i>Repitencia</i>	0.055	0.04	0.077	1.00	0.00	1.00
<i>Deserción</i>	0.041	0.00	0.075	0.71	0.00	0.71
<i>Aprobación</i>	0.908	0.94	0.119	1.00	0.00	1.00
<i>Estudiantes por aula</i>	15.460	13.75	9.421	175.50	0	175.50
<i>Servicios</i>	2.488	3	0.735	3	0	3
<i>Infraestructura</i>	1.352	1	1.035	5	0	5
<i>Tamaño</i>	1.737	1	1.692	10	1	11
<i>Inglés</i>	0.526	1	0.499	1	0	1
<i>% Menor de 30</i>	0.110	0	0.198	1	0	1
<i>% De 30 a 40</i>	0.326	0.33	0.273	1	0	1
<i>% De 40 a 50</i>	0.337	0.32	0.296	1	0	1
<i>% Mayor de 50</i>	0.227	0.17	0.267	1	0	1
<i>% Monto Salario</i>	1.200	1.20	0.246	1.65	0.28	1.93
<i>Salario Base (en millones)</i>	0.509	0.52	0.080	0.59	0.13	0.72
<i>Incentivos (en millones)</i>	0.382	0.37	0.116	0.81	0.06	0.88
<i>Zonaje (en millones)</i>	0.737	1	0.418	1	0	1
<i>IDS (en millones)</i>	0.676	0.93	0.412	1	0	1
<i>Porcentaje PT6</i>	0.543	0.52	0.297	1	0	1
<i>Porcentaje PT5</i>	0.132	0	0.232	1	0	1
<i>Porcentaje PT4</i>	0.042	0	0.126	1	0	1
<i>Porcentaje PT3</i>	0.020	0	0.099	1	0	1
<i>Porcentaje VT6</i>	0.019	0	0.041	0.33	0	0.33
<i>Porcentaje KT3</i>	0.121	0.11	0.132	0.75	0	0.75
<i>Porcentaje KT2</i>	0.013	0	0.056	0.50	0	0.50
<i>Porcentaje Interinos</i>	0.470	0.45	0.308	1	0	1
<i>Porcentaje Aspirantes</i>	0.083	0	0.170	1	0	1
<i>Rural</i>	1.641	2	0.480	1	1	2
<i>Pobreza</i>	0.394	0.37	0.158	0.86	0.07	0.93

Elaboración propia con información del MEP y PEE, 2016

ANEXO 4: Metodología para el análisis exploratorio con modelos econométricos geográficamente ponderados a nivel de escuela.

Los modelos estimados en la sección anterior son modelos a nivel global, es decir estiman parámetros para todo el conjunto de escuelas del país y permiten identificar relaciones entre diferentes factores (determinantes) y los indicadores de rendimiento educativo. Sin embargo, el valor de los parámetros se debe interpretar como un valor promedio a nivel nacional, es decir, se parte del supuesto que un factor identificado afecta en promedio de manera similar a todos los centros educativos lo cual resulta en un supuesto muy fuerte dada las diferencias territoriales. Por ejemplo, el rendimiento (deserción, repitencia y aprobación) de una escuela ubicada en el pacífico sur del país puede ser afectado por los mismos factores que una escuela ubicada en la frontera norte del país, sin embargo el valor del parámetro puede ser muy diferente, dada la variación que se presenta en las relaciones de los factores en cada contexto.

Dado lo anterior, en esta sección se estiman modelos econométricos geográficamente ponderados con el fin de identificar como varían en el espacio los parámetros o factores estadísticamente significativos encontrados en la sección anterior. Para estimar dichos modelos se construye una ecuación distinta para cada escuela al incorporar las variables dependientes y explicativas de las escuelas que caen dentro del ancho de banda de cada entidad de destino. Para determinar la forma y la extensión del ancho de banda se utilizaron diferentes criterios (Kernel (núcleo), Ancho de banda, Distancia y Cantidad de vecinos). La aplicación de este tipo de modelos es posible gracias a la gran cantidad de datos disponibles (3.749 escuelas) y a que los mismos se encuentran georreferenciados. Los resultados se presentan en mapas permitiendo así observar para cada escuela como cambian los valores de los parámetros según la zona del país donde se ubica.

Para aplicar el método en los modelos planteados, se emplea una matriz de pesos, en los cuales estos están definidos con respecto a la distancia de un punto (llamado “punto de regresión”), de acuerdo con la ecuación:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i$$

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y$$

Donde W es una matriz de pesos que depende de la posición del dato y de la posición del punto de regresión en el espacio (u_i, v_i). Los factores de peso que componen las entradas de W varían exponencialmente conforme se alejan del punto de regresión, de acuerdo con la ecuación:

$$w_{ij} = \exp \left[- \frac{1}{2} (d_{ij}/b)^2 \right]$$

Donde d_{ij} es la distancia euclídeana entre el punto de regresión y el dato, y b es un parámetro, el ancho de banda, a partir del cual los datos más alejados ya no influyen sobre el coeficiente de regresión que se busca estimar. Estrictamente hablando, esta formulación particular corresponde a una función de decaimiento gaussiana, en la cual el factor de peso pierde valor conforme se aleja del punto de regresión; existen otras formulaciones alternas de la variación en los factores de peso. Se estimará una ecuación para cada escuela, de modo que existen coeficientes con valores diferentes para cada centro educativo.

Estimación del modelo y resultados

Para determinar si existe correlación espacial en el modelo, se siguen dos pasos. Primero se corre un modelo con las variables a utilizar. Luego nuevamente se calcula el estadístico de Moran. Al estimar el modelo inicial, se descartan una serie de variables que presentaban bajos niveles de significancia, heterocedasticidad y multicolinealidad.

El estadístico de Moran viene dado por:

$$I = \frac{[N]}{[S]} \cdot \{(e'W e) | e'e\}$$

En este caso, e es un vector de residuales del modelo. Por su parte, S es un factor de estandarización igual a la suma de todos los elementos de la matriz de distancias (W) (matriz de escuelas de 3.749 x 3.749). Después se aplican las pruebas de multiplicadores de Lagrange para definir qué modelo (de rezago o de error) es mejor para la estimación. Cuando los multiplicadores de Lagrange son significativos para ambos tipos de especificación, se recurre a otra prueba: la del multiplicador robusto. Esta prueba determina si se usa el modelo combinado o no.

El cuadro 14 resume los estadísticos de Moran y de multiplicadores de Lagrange del modelo estimado. El estadístico de Moran nuevamente es positivo y significativo. Las pruebas del multiplicador de Lagrange, tanto para el modelo de error como para el de rezago, son significativas. De acuerdo con la prueba del multiplicador robusto, el mejor modelo para estimar los determinantes es el de rezago espacial, ya que es el único significativo

Cuadro 14 Estadístico de Moran y Multiplicadores de Lagrange

Prueba	Estadístico	P-Value
Estadístico de Moran	17,326	0,000
a) Error espacial		
Multiplicador de Lagrange	246,247	0,000
Multiplicador Robusto de Lagrange	0,491	0,379
b) Rezago Espacial		
Multiplicador de Lagrange	393,901	0,000
Multiplicador Robusto de Lagrange	62,779	0,000

Elaboración propia con información del MEP y PEE, 2016

Dado lo anterior al modelo se le adiciona de acuerdo a las pruebas estadísticas, un término de rezagado o ponderado por la matriz. Los mapas siguientes muestran los resultados de las estimaciones para los tres indicadores de rendimiento evaluados; las cuales se realizan siguiendo el método de máxima verosimilitud, dado que el modelo original presentaba correlación espacial. La interpretación de los mapas se debe hacer tomando en consideración los siguientes aspectos:

- Cada punto en el mapa es una escuela.
- Cada mapa representa una variable independiente.
- Las variables mostradas en cada mapa representan solo aquellas que dieron significativas en el modelo para explicar las variables de rendimiento.
- Cada escuela (punto) tiene un color asignado, este color responde al valor del coeficiente estimado. En total existen 5 colores que responde a la agrupación que se hizo por quintiles de la siguiente forma (*Quintil 1 menos coeficiente y Quintil 5 mayor coeficiente*):

Quintil 1	
Quintil 2	
Quintil 3	
Quintil 4	
Quintil 5	

Grado de explicación de los modelos locales

Los resultados se pueden resumir de la siguiente forma:

- En el caso de la deserción el modelo planteado tiende a ser un mejor predictor en cantones como Sarapiquí, San Carlos, Pococí, Mora, Acosta, Puriscal y Parrita. Y menos explicativo en las regiones Chorotega, Brunca y la parte sur de la región Atlántica.
- En el caso de la repitencia la significancia del modelo es mayor en los cantones de las regiones Brunca y Atlántica y menos explicativo en Pacífico Central, GAM y península de Nicoya.
- Por último, la aprobación muestra los mayores valores en las regiones Chorotega, Brunca y Huetar Norte y un menor grado de explicación dentro de la GAM.

Valor de los coeficientes: Modelo local de deserción

En el caso de la variable deserción los resultados muestran los mayores y menores coeficientes en los siguientes cantones del país:

Variable independiente	Coeficientes locales	
	Mayores valores	Menores valores
Inglés	Península de Nicoya, GAM y Pacífico Central.	Región Huetar Norte, Brunca y sur de la región Atlántica
Infraestructura	Cantones de San Carlos, Pococí y Sarapiquí, Hojancha, Nicoya y Puntarenas, Santa Bárbara y Parrita.	Buenos Aires, Coto Brus, Corredores, Osa, Golfito, Limón, Talamanca y Upala.
Estudiantes por aula	Cantones de San Carlos, Los Chiles, Upala, Guatuso, cantones de la región Chorotega.	Turrialba y cantones de la región Atlántica.

Elaboración propia con información del MEP y PEE, 2016

Valor de los coeficientes: Modelo local de repitencia

En el caso de la variable repitencia los resultados muestran los mayores y menores coeficientes en los siguientes cantones del país:

Variable independiente	Coeficientes locales	
	Mayores valores	Menores valores
Aspirantes	Cantones de la región Brunca, Talamanca, Santa Cruz, Nicoya y Hojancha, Abangares y Liberia.	Sur del cantón de San Carlos, cantones del Norte de Heredia, Pococí y Sarapiquí.
Estudiantes por aula	Región Brunca y Talamanca	GAM, Turrialba, Saraipuí, Pococí y Pacífico Central.
Infraestructura	Cantones de la región Chorotega y Huetar Norte (menos San Carlos), Talamanca y Golfito.	Puntarenas, Abangares, San Ramón, Zarcero, Alvaro Ruiz, Cartago El Guarco, Aguirre, Pérez Zeledón y Limón.
Inglés	San Carlos, Sarapiquí, Pococí, Carrillo, Liberia y Bagaces.	Parrita, Puriscal, Acosta y Mora.
% profesores con titulación KT2	Cantones de la región Chorotega, Upala, Puriscal, Parrita, Acosta, Dota, Tarrazú y Aguirre.	San Carlos, Sarapiquí, Pococí y Pérez Zeledón.
% profesores menores de 30 años	Paraiso, Jiménez, Turrialba, Matina, Sarapiquí y Pococí.	Cantones del este y norte de la GAM, San Carlos, Parrita, Aguirre, Tarrazú, Puriscal y Acosta.
Zona de pobreza	Cantones Península de Nicoya, Abangares, San Ramón, San Carlos, Guatuso, Upala, Oreamuno, Turrialba, Alvarado, Paraiso, Limón, Matina, Sarapiquí y Pococí.	Oeste de la GAM, El Guarco, Tarrazú, Dota y Aguirre.
% profesores con titulación PT6	Cantones de la GAM, Garabito, San Mateo, Orotina, Esparza, Turrubares, Puriscal, Sarapiquí y Pococí	Cantones de la región Brunca y Talamanca
Rural	Oeste de la GAM, San Ramón, Zarcero, Naranjo, Atenas, San Mateo, Garabito, Esparza, cantones de la región Brunca, Turrialba, Talamanca y el sur de San Carlos.	Sarapiquí y Pococí
Tamaño	Cantones de la región Huetar Norte, Montes de Oro, Esparza, San Ramón, Zarcero, Naranjo, Valverde Vega, Talamanca, Buenos Aires y Coto Brus.	Turrialba y Siquirres.

Elaboración propia con información del MEP y PEE, 2016

Valor de los coeficientes: Modelo local de aprobación

En el caso de la variable aprobación los resultados muestran los mayores y menores coeficientes en los siguientes cantones del país:

Variable independiente	Coeficientes locales	
	Mayores valores	Menores valores
Estudiantes por aula	Región Brunca, Talamanca y Abangares	Upala, Liberia y Carrillo
Infraestructura	San Carlos, Los Chiles, Guatuso, Sarapiquí y Tilarán	Cantones de la GAM, Puriscal, Turrubares, Tarrazú, Acosta, Orotina, León Cortés, Garabito, Parrita, Pérez Zeledón, Guácimo, Pococí y Turrialba
Inglés	Cantones de la Huetar Norte, Tilarán, Grecia, Talamanca, Limón y Pérez Zeledón	San José, Alajuelita, Escazú, Aserri, Desamparados, León Cortés, Tarrazú, Aguirre, Turrubares, Puriscal, Puntarenas y Pococí
% profesores con titulación KT2	Aguirre, Parrita, Tarrazú, El Guarco, Naranjo, Grecia, Atenas, San Carlos, Sarapiquí, Pococí	Cantones de la Península de Nicoya, Tilarán, Cañas, Upala, Puntarenas, Pérez Zeledón, Talamanca y Limón
% profesores con menos de 30 años	Cantones de la GAM, Turrubares, Puriscal, Acosta, Tarrazú, León Cortés, Garabito, Parrita, Aguirre, Sarapiquí, Grecia y San Carlos,	Pérez Zeledón, Pococí, Guácimo, Siquirres, Turrialba, Alvarado, Oreamuno, Paraíso y Jiménez
Pobreza	Región Chorotega, Upala, Guatuso, Puntarenas, Montes de Oro y Esparza	Pococí, Guácimo, Siquirres, Limón, Pérez Zeledón, Turrialba, Oreamuno, Paraíso, Alvarado, Jiménez, Cartago, Vásquez de Coronado
% profesores con titulación PT6	Cantones de la Región Brunca, Limón, Talamanca, Sarapiquí, Grecia, San Carlos, Turrialba, Jiménez, Alvarado, Cartago y Oreamuno	Nicoya, Nandayure, Hojancha, Puntarenas, Esparza, Garabito, Parrita, Aguirre, Turrubares, Puriscal, Acosta, León Cortés, Aserri, Mora, Desamparados, Santa Ana y Escazú

Elaboración propia con información del MEP y PEE, 2016

Referencias bibliográficas consultadas

- Anselin, L, Florax, R.J.G.M, Rey, S. (eds). (2004). *Advances in Spatial Econometrics, Methodology, Tools and Applications*. Berlin: Springer.
- Anselin, L, Florax, R.J.G.M. (eds). (1995). *New Directions in Spatial Econometrics*. Berlin: Springer.
- Anselin, L. (1988). *L. Spatial Econometrics: Methods and Models*. Boston, MA: Kluwer Academic.
- Anselin, L. (1992). "Space Stat tutorial. A workbook for using SpaceStat in the analysis of spatial data". Technical Report S-92-1, National Center for Geographic Information and Analysis, University of California. Santa Barbara, CA.
- Anselin, L. (1995). Local indicators of spatial Association-LISA. *Geographical Analysis*, vol. 27, nº 2, p. 93-115.
- Anselin, L. (2002) "Under the hood. Issues in the specification and interpretation of spatial regression models". *Agricultural Economics* 27; pp. 247–267.
- Anselin, L. (2003). *GeoDa 0.9 User's Guide*. Spatial Analysis Laboratory, University of Illinois, Urbana-Champaign, IL.
- Anselin, L. y R. Florax (1995). *New Directions in Spatial Econometrics*. Springer-Dordrecht: Reidel. Verlag, Berlin, Alemania.
- Aragón, C.; Aranguren, M .J.; Iturrioz, C. (2002): *El sector textil: análisis para un estrategia*, Universidad de Deusto, San Sebastián.
- Cliff, A., Ord, J. (1972). Testing for spatial autocorrelation among regression residuals. *Geographical Analysis*, vol. 4, p. 267-284.
- Cliff, A., Ord, J. (1973). *Spatial autocorrelation*. London: Pion.
- Cliff, A., Ord, J. (1981). *Spatial processes, models and applications*. London: Pion.