



Aprendiendo en lengua ajena: Una evaluación cuasi-experimental de los efectos académicos de la instrucción en castellano

Informe final

Proyecto mediano

Código de control: 429

Lima, 1 de Octubre, 2014

Ramiro Burga Villanueva

Luís Ricardo Maertens Odria

ÍNDICE

Resumen.....	2
1. Introducción	3
2. Marco Teórico.....	8
3. Datos.....	14
3.1 Análisis de Cohortes.....	17
3.2 Estadísticos descriptivos.....	18
4. Estrategia empírica.....	22
4.1 Resultados.....	25
5. Conclusiones y recomendaciones de política.....	27
6. Plan de incidencia	30
Bibliografía	32
Apéndice	36

Resumen

En el Perú, una gran cantidad de niños que tienen al quechua como lengua materna aprenden sólo en castellano, mientras que otros aprenden principalmente en su lengua materna. Este desajuste lingüístico entre la lengua materna de los alumnos y la lengua de instrucción por parte de los maestros ha sido posible debido a una histórica e inexistente política de lenguas en el sistema educativo peruano ¿Cuál es el efecto de aprender en una lengua distinta a la materna sobre las tasas de deserción y repetición escolar? Usando nuevos datos recopilados por el Ministerio de Educación sobre las lenguas de instrucción en las escuelas del Perú estimamos que la enseñanza en castellano (una lengua ajena), en nuestra población, está asociada a un aumento en las tasas de deserción escolar en tercero de primaria y repetición de grado en cuarto de primaria de 0.92 y 1.23 puntos porcentuales, respectivamente. Dichas cifras representan el 24% y 12% de las probabilidades incondicionales de deserción y repetición escolar en dichos grados.

Abstract

Among the Peruvian children that have Quechua as their mother tongue there is a large proportion that is schooled completely in Spanish, while the rest is schooled mainly in their mother tongue. This language mismatch between the mother tongue of the students and the language used for teaching in class has persisted due to the lack of language policy in the Peruvian educational system. What is the effect of learning in a language different to one's mother tongue on dropout and grade-repetition rates? Using newly collected data by the Ministry of Education on the languages used in class for teaching we estimate that learning in a Spanish (a "foreign" language), among our relevant population, is associated with an increase of 0.92 percent in the dropout rate in third grade and an increase of 1.23 percent in the grade-repetition rate in fourth grade. These numbers constitute 24% and 12% of the unconditional dropout and grade-repetition rates, respectively.

1. Introducción

El objetivo principal de esta investigación es evaluar el impacto académico, entre niños que tienen una lengua originaria como lengua materna, de aprender sólo en castellano vis-à-vis aprender en lengua originaria (y aprender el castellano como segunda lengua). Los datos de los censos escolares muestran que, entre las escuelas que atienden a una gran mayoría de niños que tienen una lengua originaria como lengua materna, existen algunas que enseñan sólo en castellano y otras que enseñan mayoritariamente en lengua originaria. Por ejemplo, según el censo escolar 2011, de las escuelas primarias donde el 100% de alumnos tiene una lengua originaria como lengua materna, el 33% ofrece clases solo en castellano.¹ A su vez, un diagnóstico realizado en el 2011 por la Defensoría del Pueblo sobre el estado de la Educación Intercultural Bilingüe da cuenta del mismo fenómeno. ¿Cuáles son los efectos de aprender en una lengua distinta a la materna?

La elección de la(s) lengua(s) de instrucción en las escuelas tiene potenciales consecuencias sobre: (i) la adquisición de habilidades lingüísticas en dichas lenguas, (ii) la adquisición de conocimientos curriculares en el aula y (iii) las tasas de repetición y deserción escolar. En la presente investigación nos concentramos en el último ítem.

Los efectos de la instrucción bilingüe, en sus distintas modalidades, han sido estudiados sólo recientemente en las literaturas de neurociencia educacional y economía. En la primera, la agenda se ha centrado en evaluar los efectos de la instrucción bilingüe sobre las habilidades lingüísticas de los alumnos, es decir, el ítem (i) arriba mencionado.² En un estudio con niños de 7 a 9 años, Kovelman et al. (2008) encuentran que niños monolingües en castellano o inglés, que asisten a escuelas bilingües, tienen mejores habilidades para leer, capacidades de reconocimiento fonético y competencias lingüísticas que sus pares monolingües que asisten a escuelas monolingües. Asimismo, la edad a la que un niño es expuesto por primera vez a una segunda lengua es un determinante crucial de sus habilidades lingüísticas en ambas lenguas. El desarrollo lingüístico en dos lenguas es óptimo cuando el niño es expuesto a una segunda lengua a una temprana edad –antes de los 5 años. (Ver Petitto y Dunbar, 2009 para una discusión exhaustiva) En cuanto a los efectos de la

¹ Dicha cifra es aproximadamente de 42% y 46% según los censos del 2010 y 2009, respectivamente.

² Revisar Kim et al. (1997), Kovelman y Petitto (2003), Petitto et al. (2003), Petitto y Dunbar (2009) y Wartenburger et al. (2003).

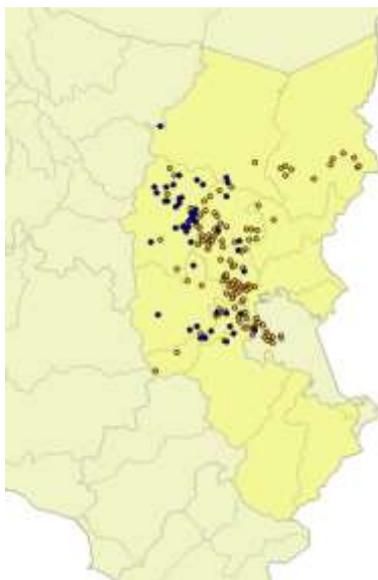
educación bilingüe sobre habilidades extra-lingüísticas, Baker et al. (2003) encuentran que los niños bilingües desde el nacimiento tienen una ventaja cognitiva respecto de sus pares monolingües. Esta literatura se relaciona con la pregunta de investigación en este documento en tanto que aprender en una lengua ajena es una manera de transitar al bilingüismo: bilingüismo por inmersión. Adicionalmente la evidencia empírica de estos estudios proviene, en su mayoría, de aulas escolares en las que se manipula la lengua de instrucción y el momento en el que se introduce una segunda lengua.

Desde la economía, la medición de los efectos de una educación bilingüe ha recibido gran atención; sin embargo, en su gran mayoría, los estudios se han centrado en programas destinados a poblaciones de inmigrantes. Matsudaira (2005) explota una discontinuidad en la asignación de niños inmigrantes a programas bilingües o a clases regulares, para evaluar el efecto de los programas bilingües sobre el desempeño académico. El autor utiliza datos para un distrito escolar urbano en los EE.UU. y encuentra que el efecto de una educación bilingüe es cercano a cero. En línea con este resultado, Chin et al. (2012) encuentran que la provisión de programas bilingües a niños de entornos inmigrantes en Texas tiene un efecto nulo sobre las notas de los niños inmigrantes. No obstante, la provisión de dichos programas, al cambiar la composición de las clases, aumenta significativamente las notas del resto de los alumnos –no inmigrantes.

La evidencia sobre los efectos de una educación bilingüe en poblaciones no inmigrantes es bastante reducida. Angrist et al. (2008) utilizan un cambio en la lengua de instrucción en las escuelas en Puerto Rico en 1949 para evaluar el efecto de la enseñanza en inglés sobre el aprendizaje del idioma inglés. Los autores encuentran que la enseñanza en inglés no aumenta significativamente el conocimiento del idioma. Anghel et al. (2012) utilizan un programa de educación bilingüe en Madrid para evaluar el impacto de aprender ciertas materias en inglés (una lengua extranjera) sobre el rendimiento académico. Los autores concluyen que el efecto de aprender una materia en inglés sobre el desempeño en dicha materia es negativo para los niños cuyos padres no acabaron la secundaria; y que no hay efectos significativos para nadie sobre el rendimiento en las materias que se enseñan en castellano. Ramachandran (2013) analiza el efecto de la introducción de un cambio en las lenguas de instrucción en las escuelas primarias en Etiopía sobre la deserción escolar. Dicho cambio, permitió que en el año 1994, diversos grupos étnicos obtuvieran acceso a la instrucción en lengua materna. Ramachandran realiza un análisis de dobles-diferencias, comparando las

diferencias en tasas de deserción antes y después de 1994 de un grupo étnico que siempre tuvo acceso a la instrucción en lengua materna –los Amharic—con las de otro grupo étnico que sólo luego de 1994 recibió instrucción en lengua materna –los Omoro. El autor encuentra que la instrucción en lengua materna aumentó en un 12% la proporción de niños que completaron 6 o más años de educación primaria. Jain (2013) analiza un experimento natural en India para evaluar el efecto de la lengua de instrucción en la escuela primaria sobre la deserción escolar. Dicho experimento es la reorganización de 1956 de los estados indios, a través de la cual se reagruparon distintos distritos en estados en función a las lenguas habladas en dichos distritos. Los nuevos estados implementaron una política de educación en la lengua mayoritaria del estado, lo que permite comparar las diferencias –antes y después de 1956-- de los resultados de un mismo grupo étnico en estados que enseñan en su lengua materna y en los que no. Jain encuentra que aprender en lengua materna aumenta la tasa de completar la primaria en 24.4%.

Gráfico 1: Escuelas rurales que atienden a poblaciones 100% quechua-hablantes en Puno según lengua de instrucción



Nota: Los puntos azules indican escuelas en las que la instrucción se realiza en lengua originaria. Los puntos naranja indican escuelas en las que la instrucción se realiza en castellano.

Fuente ESCALE-MINEDU/ Elaboración propia

Al igual que Ramachandran (2013) y Jain (2013), este estudio analiza el efecto de la lengua de instrucción sobre la deserción escolar. Adicionalmente, y a diferencia de Ramachandran (2013) y Jain (2013), nosotros investigamos el efecto de la lengua

de instrucción sobre la tasa de retención escolar. Nuestra identificación econométrica presenta una ventaja frente a las de Ramachandran y Jain: en nuestra base de datos tenemos variación en la lengua de instrucción al interior de un mismo estado/provincia/unidad de gestión educativa local (UGEL) y de un mismo grupo étnico. Dicha variación es evidente en el gráfico 1, donde se observan escuelas que enseñan tanto en castellano como lengua originaria, y que atienden a poblaciones 100 por ciento quechua-hablantes, al interior de una misma provincia.³ Esto nos permite ahorrar en supuestos acerca de las posibles diferencias (en niveles o tendencias) en la deserción escolar en distintos distritos y a través de grupos étnicos.⁴

Sin embargo, para identificar el efecto causal de la lengua de instrucción sobre resultados académicos es necesario que la heterogeneidad en la lengua de instrucción no esté correlacionada con determinantes no observables de dichos resultados, una vez que controlamos por determinantes observables. En este estudio, controlamos por una serie de características de las escuelas y maestros; y explotamos la falta de una política relacionada a la lengua de instrucción en escuelas primarias, tanto a nivel del gobierno central como de gobiernos locales, con el propósito de identificar el efecto causal de la lengua de instrucción. Un informe de la Defensoría del Pueblo (2011) describe el contexto legal y administrativo que permitió el desajuste lingüístico entre las lenguas de los niños indígenas y la lengua de instrucción utilizada por los maestros. Antes del año 2010, en el que se emitió la Directiva 012-2010, no había virtualmente ninguna ley, norma, o directiva que requiriese a un potencial maestro dominar la lengua materna de la población a la que atendería de ser contratado. Más aun, a pesar de que dicho vacío legal dejó de existir en el año 2010, el informe de la Defensoría del Pueblo indica que los comités de contratación regionales de maestros, en su gran mayoría, no evaluaban el adecuado dominio de la lengua local por parte de los maestros candidatos. A esta situación, se suma una disminución significativa en el número de Institutos Superiores Pedagógicos que ofrecen la especialidad de Educación Intercultural Bilingüe: “Los datos no dejan dudas: a la fecha solo cinco ISP en todo el Perú ofrecen las carreras de EIB. Para una población que bordea alrededor

³ No presentamos el mismo gráfico para todo el Perú debido a que la base de datos de locales educativos georeferenciados del MINEDU tan sólo cubre alrededor del 50% del total de locales educativos registrados en el Padrón. Se eligió reportar el gráfico para la región Puno debido a que es una de las regiones con mayor número de escuelas rurales georeferenciadas que atienden a poblaciones 100% quechua-hablantes.

⁴ El análisis de Ramachandran tiene sólo variación en lengua de instrucción al interior de un estado/provincia, por lo que el grupo de control es otro grupo étnico en el mismo estado. El análisis de Jain tiene sólo variación en lengua de instrucción al interior de un mismo grupo étnico, por lo que el grupo de control está compuesto de estudiantes del mismo grupo étnico pero que viven en otro estado.

de 4 millones de habitantes, contar únicamente con cinco instituciones de formación docente es incomprensible e injustificable.” (Defensoría del Pueblo, 2011, p. 2040). Adicionalmente, las respuestas a una breve encuesta a las Direcciones Regionales de Educación (DREs), indica gran aleatoriedad en la asignación de maestros (cualquiera sea las lenguas que domina) a escuelas –reproducimos las preguntas y un ejemplo de respuesta en el apéndice A.⁵

En el Perú, existen dos estudios que abordan la relación entre rendimiento académico y lengua. Cueto y Secada (2003) evalúan los efectos de los programas de Educación Intercultural Bilingüe (EIB) sobre el logro académico y deserción escolar, en una muestra de escuelas en Puno. Dicho estudio tiene dos problemas principales. Primero, como es mencionado por los mismos autores, las escuelas que son tomadas como EIB (así consideradas por el Ministerio de Educación) generalmente no tienen maestros que sepan leer y escribir en lengua originaria, lo que pone en cuestión el carácter bilingüe de dicha educación. Segundo, los resultados de los niños que atienden a colegios EIB son comparados con los resultados de niños que atienden colegios monolingües en castellano. Dado que las características de los niños que atienden a colegios monolingües en castellano pueden ser fundamentalmente distintas a las características de los niños que atienden escuelas EIB, es cuestionable que los primeros sean un contrafactual adecuado para los segundos. Cueto et al. (2009) evalúan los efectos sobre el desempeño escolar de tener al castellano o a una lengua originaria como lengua materna. Este estudio es fundamentalmente distinto al nuestro, en tanto que nosotros buscamos evaluar el efecto de la lengua de instrucción (no necesariamente la lengua materna) sobre las tasas de repetición y deserción. Nuestra estrategia empírica utiliza nuevos datos recolectados por el Ministerio de Educación, sobre la lengua de instrucción en las escuelas, y métodos cuasi-experimentales para identificar dicho efecto causal. Debe quedar claro que esta investigación no pretende evaluar los efectos del programa de EIB, ya que como ha sido documentado por la Defensoría del Pueblo (2011), que un colegio sea denominado como de EIB no implica necesariamente que la enseñanza se realice en la lengua materna de los estudiantes.

Adicionalmente, nuestra investigación tiene como objetivo específico contribuir a la formulación de políticas educativas que fomenten la equidad, lo que es particularmente relevante si se toma en cuenta que la mayoría de hogares indígenas

⁵ Se intentó contactar telefónicamente a las 8 regiones con la mayor proporción de estudiantes que no son instruidos en su lengua materna. Se mantuvieron conversaciones telefónicas con 4 de ellas, y se les envió un breve cuestionario referido a los temas conversados. A la fecha, 2 regiones han respondido los cuestionarios.

viven en condiciones de pobreza (Trivelli, 2005). Si, efectivamente, una educación monolingüe en castellano contribuye al peor rendimiento (o mayor tasa deserción o repetición) de los niños que tienen una lengua originaria como lengua materna, la necesidad de una educación propiamente bilingüe sería manifiesta. Ello, no sólo por consideraciones socio-lingüísticas, sino por consideraciones educativas que, finalmente, son importantes determinantes del bienestar económico.

2. Marco Teórico

En esta sección desarrollamos un modelo de aprendizaje con el propósito de entender los efectos de recibir una educación escolar primaria en una lengua distinta a la materna. El modelo se sitúa dentro de la literatura de atención limitada (Sims, 1998, 2003) y parte del principio que los individuos tienen acceso a una gran cantidad de información y que, sin embargo, no pueden prestarle atención a toda ella debido a que la atención es un recurso escaso: los individuos optan racionalmente por prestar más atención a aquello que les interesa que a aquello que no.⁶ En el contexto de la enseñanza en la escuela primaria, que empieza típicamente a los 6 años, el supuesto de atención limitada tiene soporte empírico tanto para el desarrollo lingüístico (Petitto y Dunbar, 2009) como para el desarrollo de conocimientos extra-lingüísticos (Baker et al. 2003). Petitto y Dubar, resumiendo resultados de análisis de cerebros usando resonancias magnéticas funcionales (fMRI por sus siglas en inglés) concluyen: “Los [niños] bilingües ‘*tardíos*’ exhiben mayor activación *bilateral*, requieren de más uso de tejido en el lóbulo frontal (incluyendo memoria de trabajo y áreas inhibitorias) y frecuentemente exhiben mayor esfuerzo cognitivo... en el desarrollo de pruebas de lingüísticas.” (Petitto y Dunbar, 2009, p. 191 – traducción propia, cursiva original)

Asumimos que existe un aula escolar con un único alumno cuya lengua materna es una lengua originaria (o) –por definición distinta del castellano. Dicho alumno puede ser instruído en una de dos lenguas $a = c, o$: castellano (c) o su lengua materna (o). En el aula, el alumno puede aprender dos cosas: el contenido que imparte el profesor y la lengua en la que recibe instrucción. Formalmente, asumimos que existen cuatro variables aleatorias $A_a: \Omega_{A,a} \rightarrow R$ y $L_a: \Omega_{L,a} \rightarrow R$, $a = c, o$, que representan el aprendizaje por parte del alumno del contenido de la clase y de la

⁶ Sims (1998) propone un modelo macroeconómico de atención limitada para explicar la rigidez de precios y salarios. Sims (2003) discute las implicancias de la teoría de atención limitada para la macroeconomía. Nuestro modelo es, hasta donde sabemos, el primero en aplicar la teoría de atención limitada a un problema de aprendizaje en contextos bilingües.

lengua de instrucción, respectivamente, cuando es instruido en castellano (c) y lengua originaria (o). $\Omega_{A,a}$ y $\Omega_{L,a}$, representan toda la información relevante que tiene el estudiante sobre el contenido de la clase y la lengua de instrucción, respectivamente, cuando es instruido en castellano y lengua originaria $a = c, o$. Por ejemplo, A_c puede representar la respuesta del alumno a la pregunta “¿Cuánto es 3 por 0?” cuando es instruido en castellano. Por otro lado L_o puede representar el tono con el que el alumno pronuncia la segunda letra “a” en “papá” cuando aprende en lengua originaria. Claramente, en una clase en el mundo real los alumnos aprenden sobre varios contenidos y diversos aspectos de la lengua de instrucción, entre otras cosas. Modelar el extenso conjunto de conocimientos que se manejan en el aula aumentaría la dimensionalidad del modelo exponencialmente y escapa al propósito de este trabajo.

El alumno no conoce con certeza el contenido de la clase ni el correcto uso de la lengua de instrucción, sin embargo tiene creencias acerca de las mismas. Asumimos que dichas creencias son normales y con media cero. El último supuesto es una normalización y no altera los resultados del modelo. Formalmente,

$$\begin{bmatrix} A_a \\ L_a \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{A,a}^2 & 0 \\ 0 & \sigma_{L,a}^2 \end{bmatrix} \right)$$

donde $\sigma_{A,a}^2$ representa la incertidumbre del alumno respecto del contenido de la clase cuando es instruido en la lengua $a = c, o$ y donde $\sigma_{L,a}^2$ representa la incertidumbre del alumno respecto de la lengua de instrucción cuando es instruido en la lengua $a = c, o$. En este modelo asumimos que la incertidumbre a priori del alumno (antes de recibir instrucción) respecto del contenido de la clase es la misma si es instruido en castellano o lengua originaria. Sin embargo, la incertidumbre a priori sobre la lengua de instrucción es menor si dicha lengua coincide con su lengua materna. Ello implica que $\sigma_{A,c}^2 = \sigma_{A,o}^2 = 1$ (la última igualdad no es necesaria, pero facilita el cálculo) y que $\sigma_{L,o}^2 < \sigma_{L,c}^2$.

El profesor, que conoce los verdaderos valores de A y L los comunica al alumno. Sin embargo este, que como todo el resto de personas tiene una atención limitada, no puede aprender con certeza lo comunicado por el profesor. Formalmente decimos que el alumno recibe una señal ruidosa sobre el verdadero mensaje

$$\begin{bmatrix} S_{A,a} \\ S_{L,a} \end{bmatrix} = \begin{bmatrix} A \\ L \end{bmatrix} + \begin{bmatrix} \varepsilon_{A,a} \\ \varepsilon_{L,a} \end{bmatrix},$$

$$\begin{bmatrix} \varepsilon_{A,a} \\ \varepsilon_{L,a} \end{bmatrix} \sim iid N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \eta_{A,a}^2 & 0 \\ 0 & \eta_{L,a}^2 \end{bmatrix} \right)$$

donde $\eta_{A,a}^2$ y $\eta_{L,a}^2$ miden el nivel de ruido en la señal sobre el contenido de la clase y la lengua de instrucción recibida por el alumno, respectivamente, cuando es instruido en la lengua $a = c, o$. Optamos por asumir que las señales son bivariadas para resaltar el hecho que es difícil, sino imposible, comunicar el contenido de la clase sin usar el lenguaje y que, por lo tanto, toda señal sobre contenido está acompañada por una señal sobre la lengua de instrucción.

Gracias a la señal, el alumno puede actualizar sus creencias acerca de la distribución del contenido de la clase y la lengua de instrucción. En el modelo, asumimos que el alumno es racional y que actualiza sus creencias usando la regla de Bayes. Por lo tanto, la creencia a posteriori sobre las distribuciones de A_a y L_a , cuando es instruido en la lengua $a = c, o$ son

$$\begin{bmatrix} A_a \\ L_a \end{bmatrix} \sim N \left(\begin{bmatrix} \frac{1}{1 + \eta_{A,a}^2} S_{A,a} \\ \frac{\sigma_{L,a}^2}{\sigma_{L,a}^2 + \eta_{A,a}^2} S_{L,a} \end{bmatrix}, \begin{bmatrix} \frac{\eta_{A,a}^2}{1 + \eta_{A,a}^2} & 0 \\ 0 & \frac{\sigma_{L,a}^2 \eta_{L,a}^2}{\sigma_{L,a}^2 + \eta_{L,a}^2} \end{bmatrix} \right).$$

Por cuestiones de cálculo, será conveniente reescribir la distribución de A_a y L_a en función a la precisión y no la varianza del ruido de las señales. Así definimos la precisión de la señal que recibe el alumno acerca de las variables A_a y L_a como $p_{A,a} = \frac{1}{\eta_{A,a}^2}$ y $p_{L,a} = \frac{1}{\eta_{L,a}^2}$, respectivamente.

$$\begin{bmatrix} A_a \\ L_a \end{bmatrix} \sim N \left(\begin{bmatrix} \frac{1}{1 + \eta_{A,a}^2} S_{A,a} \\ \frac{\sigma_{L,a}^2}{\sigma_{L,a}^2 + \eta_{A,a}^2} S_{L,a} \end{bmatrix}, \begin{bmatrix} (p_{A,a} + 1)^{-1} & 0 \\ 0 & \left(p_{L,a} + \frac{1}{\sigma_{L,a}^2} \right)^{-1} \end{bmatrix} \right)$$

Asumimos que el alumno tiene la misma función de pérdida, sin importar si es instruido en castellano o lengua originaria, que depende negativamente de la varianza (o imprecisión) a posteriori que tiene acerca del conocimiento impartido en clase y la

lengua de instrucción. Esto simplemente quiere decir que al alumno le interesa reducir la imprecisión de sus conocimientos sobre lo enseñado en clase y la lengua en la que es instruido.

$$EU_a = \lambda(p_{A,a} + 1)^{-1} + (1 - \lambda) \left(p_{L,a} + \frac{1}{\sigma_{L,a}^2} \right)^{-1}$$

La importancia relativa del primer término viene dada por $\lambda \in (0,1)$. Notamos que, si el alumno pudiesen elegir el valor de $p_{A,a}$ y $p_{L,a}$, es decir, la precisión con la que reciben los mensajes del maestro, elegiría $p_{A,a} = \infty$ y $p_{L,a} = \infty$, que hacen que la pérdida esperada sea cero. Ello implicaría que el alumno entiende perfectamente el mensaje (bivariado) del maestro. Sin embargo, siguiendo la literatura de atención limitada, hacemos 3 supuestos estándar: (1) el alumno puede elegir a qué información prestar más atención (el contenido de la clase o la lengua de instrucción) y por ende reducir más su varianza a posteriori acerca de la misma; (2) el alumno tiene una atención limitada y por tanto existe un límite a la capacidad que tienen para reducir la “incertidumbre total” acerca de su conocimiento del contenido del aula y la lengua de instrucción⁷; (3) el alumno no pueden “desaprender” lo que ya sabe.

El problema que enfrenta el alumno que es instruido en la lengua $a = c, o$ es:

$$\underset{p_{A,a}, p_{L,a}}{MIN} EU = \lambda(p_{A,a} + 1)^{-1} + (1 - \lambda) \left(p_{L,a} + \frac{1}{\sigma_{L,a}^2} \right)^{-1}$$

sujeto a:

$$\ln(|\Sigma_{apri,a}|) - \ln(|\Sigma_{apost,a}|) \leq k \quad (\text{capacidad limitada})$$

y

$$diag(\Sigma_{apri,a}) > diag(\Sigma_{apost,a}) \quad (\text{no desaprender})$$

donde $\Sigma_{apri,a}$ y $\Sigma_{apost,a}$ son las matrices de varianzas y covarianzas a priori y a posteriori (antes y después de la señal), respectivamente, del alumno sobre A_a y L_a . k es un número real positivo que indica la capacidad del alumno de reducir su

⁷ Formalmente, hablamos de reducción de entropía, que es un término usado inicialmente en la literatura de termodinámica y que constituye una medida inversa de grado de orden. Ver Veldkamp (2011) para una introducción a las medidas de entropía.

incertidumbre total. En nuestro modelo la capacidad del alumno es la misma si es instruido en lengua originaria o castellano. El alumno debe elegir la precisión en sus señales ($p_{A,a}$ y $p_{L,a}$): elegir una mayor precisión de una señal es equivalente a prestar más atención al contenido de dicha señal. Cuánta atención presta el alumno a una señal dependerá de sus preferencias (λ), de su capacidad cognitiva (k), y de cuánto ya sabe acerca de A_a y L_a . Para ver este último punto, basta notar que si el alumno es instruido en la lengua a y conoce a la perfección el contenido de la clase (A_a) esto significa que $\sigma_{A,a}^2 = 0$. Dado que un alumno no puede desaprender lo que ya conoce, su incertidumbre a posteriori sobre A_a también será 0, independientemente de si presta atención a la señal que envía el maestro o no. Por lo tanto, no prestará atención al contenido y sólo prestará atención a la lengua de instrucción.

A continuación nos concentramos en la solución interior al problema descrito. En este caso, la restricción de capacidad limitada es limitante y se cumple con igualdad. La solución al problema es:

$$p_{A,a}^* = \left[\left(\frac{1-\lambda}{\lambda} \right) \frac{e^k}{\sigma_{L,a}^2} \right]^{0.5} - 1$$

$$p_{L,a}^* = \left[\left(\frac{\lambda}{1-\lambda} \right) \frac{e^k}{\sigma_{L,a}^2} \right]^{0.5} - \frac{1}{\sigma_{L,a}^2}$$

Una solución interior está garantizada siempre y cuando,

$$\left(\frac{1-\lambda}{\lambda} \right) \frac{1}{e^k} < \sigma_{L,a}^2 < \left(\frac{1-\lambda}{\lambda} \right) e^k.$$

Por lo tanto, asumiremos, como lo habíamos mencionado antes, que el alumno tiene menos incertidumbre acerca de la lengua de instrucción, cuando aprende en su lengua materna. Adicionalmente, asumimos que los niveles de incertidumbre acerca de la lengua de instrucción, tanto si el alumno es instruido en castellano o lengua originaria, son “moderados”: ni tan grandes que siempre prefieran sólo prestar atención a la lengua de instrucción, ni tan pequeños que sólo quieran prestar atención al contenido de la clase,

$$\left(\frac{1-\lambda}{\lambda} \right) \frac{1}{e^k} < \sigma_{L,o}^2 < \sigma_{L,c}^2 < \left(\frac{1-\lambda}{\lambda} \right) e^k.$$

En una solución interior, dada la normalidad de todas las distribuciones, la restricción de no desaprender se satisface trivialmente.

Con los elementos arriba descritos, ahora podemos responder dos preguntas cruciales acerca del aprendizaje del contenido por parte del alumno: (1) ¿El alumno presta más atención al contenido de la clase cuando aprende en lengua originaria o castellano?; y, asociada a la pregunta anterior (2) ¿El alumno entiende mejor el contenido de la clase cuando aprende en lengua originaria o castellano?

Las respuestas a estas preguntas se desprenden directamente de las soluciones a $p_{A,a}^*$ para $a = c, o$. La diferencia en niveles de atención prestada al contenido en clase cuando el alumno aprende en castellano y lengua originaria viene dada por

$$p_{A,c}^* - p_{A,o}^* = \left[\left(\frac{1-\lambda}{\lambda} \right) e^k \right]^{0.5} \left[\frac{1}{\sigma_{L,c}^2} - \frac{1}{\sigma_{L,o}^2} \right].$$

Notar que el primer término es siempre positivo y que, por supuesto, el segundo término es negativo, por lo que el alumno que aprende en una lengua distinta a la materna – en este caso el castellano- siempre presta menos atención al contenido de la clase que cuando aprende en su lengua materna. La razón es simple. El niño que tiene una lengua originaria como lengua materna maneja menos el castellano que su propia lengua. Sin embargo, tanto si es instruido en castellano como en lengua originaria, al niño le interesa de igual manera poder dominar la lengua de instrucción. Por lo tanto, cuando el niño es instruido en una lengua distinta la suya, tendrá que desviar atención del contenido de la clase y dirigirla a aprender castellano. Si la atención no fuese un recurso escaso, el niño podría aprender todo perfectamente, sin importar en que lengua es instruido. No obstante, hemos asumido, realísticamente, que este no es el caso.

De este último resultado se sigue que el alumno acabará teniendo mayor incertidumbre acerca de los conocimientos del aula si es instruido en castellano, en comparación con el caso en que es instruido en su propia lengua. Esto es claro al analizar la varianza a posteriori de A_a , que es una función estrictamente decreciente de $p_{A,a}^*$.

En la medida que las decisiones de retención (por parte de la escuela) y deserción (por parte del alumno) escolar estén negativamente relacionadas a la comprensión de los conocimientos del aula, es de esperarse que los niños que sufren un desajuste lingüístico en la escuela (entre su lengua materna y la de instrucción) sean retenidos más en algún grado y que opten más frecuentemente por desertar la escuela.

3. Datos

La fuente de datos principal para este estudio son los Censos Escolares de los años 2008, 2009, 2010, 2011 y 2012. Estas bases de datos, como su nombre lo indican, son censos que se recogen a lo largo de todas las instituciones educativas a nivel nacional. Si bien los censos se recogen desde el año 2004, se decidió trabajar solo con aquellos posteriores al 2008 puesto que a partir de este año permiten identificar la lengua hablada por los estudiantes y la de instrucción. Es preciso mencionar que la mayor limitación de los censos es que no cuentan con información sobre las características de los alumnos.

La muestra relevante para este trabajo está conformada por aquellos colegios rurales donde el 100% de los alumnos tiene al quechua como lengua materna. Además, sólo se trabaja con las escuelas que pertenecen a los departamentos de Ancash, Apurímac, Ayacucho, Cusco, Huancavelica, Huánuco y Puno; puesto que en cada uno de ellos se encuentra un número importante de escuelas donde se habla mayoritariamente quechua (ver cuadro 2). Además, los porcentajes de escuelas que enseñan solo en castellano y las que no, se encuentran relativamente bien balanceados (ver cuadro 1). Como resultado, la muestra consta de aproximadamente 3000 instituciones educativas⁸.

⁸ Se verá más adelante que la muestra se restringirá aún más para el análisis estadístico.

Cuadro 1: Número de escuelas rurales según tratamiento

Región	2009			2011		
	Quechua	Solo castellano	Total	Quechua	Solo castellano	Total
Ancash	158 30%	371 70%	529 100%	354 65%	193 35%	547 100%
Apurímac	235 57%	180 43%	415 100%	361 83%	73 17%	434 100%
Ayacucho	342 55%	285 45%	627 100%	547 97%	19 3%	566 100%
Cusco	483 64%	273 36%	756 100%	638 86%	107 14%	745 100%
Huancavelica	201 57%	151 43%	352 100%	305 71%	123 29%	428 100%
Huánuco	26 12%	185 88%	211 100%	55 52%	50 48%	105 100%
Puno	155 29%	373 71%	528 100%	215 44%	275 56%	490 100%
Total	1,600 47%	1,818 53%	3,418 100%	2,475 75%	840 25%	3,315 100%

Fuente: Elaboración propia en base a Censos Escolares

Cuadro 2: Número de escuelas rurales según porcentaje de niños quechua-hablantes matriculados en el año 2009⁹

Región	Resto	[70%; 80[[80%; 90%[[90%; 100%[100%	Total
Amazonas	601 100%	0 0%	0 0%	0 0%	0 0%	601 100%
Ancash	574 49%	8 1%	20 2%	47 4%	529 45%	1178 100%
Apurímac	139 22%	10 2%	29 5%	50 8%	416 65%	644 100%
Arequipa	239 82%	3 1%	4 1%	1 0%	43 15%	290 100%
Ayacucho	236 26%	9 1%	9 1%	21 2%	628 70%	903 100%
Cajamarca	1043 100%	0 0%	0 0%	0 0%	1 0%	1044 100%
Cusco	478 37%	9 1%	10 1%	48 4%	756 58%	1301 100%
Huancavelica	548 59%	1 0%	4 0%	28 3%	352 38%	933 100%
Huánuco	822 78%	0 0%	4 0%	12 1%	211 20%	1049 100%
Ica	169 99%	0 0%	0 0%	0 0%	2 1%	171 100%
Junín	1265 99%	1 0%	1 0%	3 0%	5 0%	1275 100%
La Libertad	705 100%	0 0%	0 0%	0 0%	0 0%	705 100%
Lambaqueye	368 90%	0 0%	1 0%	4 1%	38 9%	411 100%
Lima	373 100%	0 0%	0 0%	0 0%	0 0%	373 100%
Loreto	1878 99%	0 0%	0 0%	0 0%	12 1%	1890 100%
Madre de Dios	123 99%	0 0%	1 1%	0 0%	0 0%	124 100%
Moquegua	65 68%	2 2%	3 3%	2 2%	24 25%	96 100%
Pasco	425 100%	0 0%	0 0%	0 0%	0 0%	425 100%
Piura	1397 100%	0 0%	0 0%	0 0%	0 0%	1397 100%
Puno	751 57%	5 0%	9 1%	20 2%	528 40%	1313 100%
San Martín	951 100%	0 0%	0 0%	0 0%	4 0%	955 100%
Tacna	83 100%	0 0%	0 0%	0 0%	0 0%	83 100%
Tumbes	70 100%	0 0%	0 0%	0 0%	0 0%	70 100%
Ucayali	634 100%	0 0%	0 0%	0 0%	0 0%	634 100%
Total	13937 78%	48 0%	95 1%	236 1%	3549 20%	17865 100%

Fuente: Elaboración propia en base a Censos Escolares

⁹ El Apéndice B contiene la misma información para el año 2011.

3.1 Análisis de Cohortes

Nuestra población de interés son las escuelas donde el 100% de los estudiantes tiene al quechua como lengua materna. Dentro de esta población definimos el tratamiento como recibir instrucción escolar en primero y segundo de primaria únicamente en castellano. Para medir los efectos del tratamiento sobre el resultado escolar, se ha procedido a trabajar con tres cohortes de estudiantes (indexadas por el año en el que los alumnos cursaban primero de primaria): 2008, 2009 y 2010. Para identificar la lengua de instrucción de una cohorte en primero y segundo de primaria utilizamos las respuestas de los directores en los censos escolares de los años en que dichas cohortes cursaban dichos grados. Por ejemplo, en la cohorte del 2008, los colegios tratados son aquellos que reciben instrucción en castellano en el 2008 y en el 2009 (de acuerdo a la información brindada por los censos).

Posteriormente, se formaron tres pool de datos que contienen los resultados académicos de tres grados: tercero, cuarto y quinto. El primer pool agrupa los resultados académicos de la cohorte del 2008, 2009 y 2010, cuando estos alumnos completan el tercer grado en los años 2010, 2011, y 2012, respectivamente. Similarmente, el segundo pool agrupa los estudiantes de las cohortes 2008 y 2009, cuando los estudiantes culminan el cuarto grado en los años 2011 y 2012. El último pool de datos solo contiene los resultados académicos de la cohorte 2008 cuando los estudiantes llegan a quinto grado en el 2012¹⁰. El cuadro 3 resume lo descrito:

Cuadro 3: Análisis de cohortes

Año de resultado	Pool de datos según grado:		
	3ero	4to	5to
Resultados académicos año 2012	Tratamiento: Enseñanza en castellano en el 2010 y 2011	Tratamiento: Enseñanza en castellano en el 2009 y 2010	Tratamiento: Enseñanza en castellano en el 2008 y 2009
Resultados académicos año 2011	Tratamiento: Enseñanza en castellano en el 2009 y 2010	Tratamiento: Enseñanza en castellano en el 2008 y 2009	
Resultados académicos año 2010	Tratamiento: Enseñanza en castellano en el 2008 y 2009		

¹⁰ A la fecha no se ha incorporado resultados académicos del año 2013 debido a que aún no se encuentran públicamente disponibles.

3.2 Estadísticos descriptivos

La falta de una política sobre las lenguas de instrucción en las escuelas ha permitido que un grupo significativo de estudiantes peruanos, que tienen al quechua como lengua materna, sea instruido en castellano. A pesar que estos grupos no están concentrados geográficamente y que las DRE reportan gran aleatoriedad en la asignación de maestros (cualquiera sea su lengua materna) a escuelas, es aún una pregunta empírica analizar el grado de aleatoriedad con el que los maestros que hablan una lengua originaria (o castellano) son asignados a diferentes escuelas que atienden a niños que tienen al quechua como lengua materna. Si la asignación fuese completamente aleatoria, las características de las escuelas a las que los profesores son asignados, deberían ser las mismas entre aquellos que enseñan en castellano y lengua originaria.

Utilizando el Censo Escolar somos capaces de determinar exactamente la composición lingüística de cada escuela. De acá en adelante nos concentramos las escuelas que cumplen las siguientes características: (i) Pertenecen a regiones que presentan porcentajes relativamente equilibrados entre tratados y no tratados, (ii) son rurales, (iii) el 100% de los alumnos tienen al quechua como lengua materna¹¹ y (iv) pertenecen a UGELs que presentan variación intracluster en la variable de tratamiento (es decir, donde hay colegios que instruyen en castellanos y otros que no lo hacen).

¹¹ En concreto se está incluyendo a los colegios que indicaron en ambos años del tratamiento (primero y segundo grado) que el 100% de alumnos eran quechua hablantes. Por ejemplo, en el caso de la cohorte del 2008 solo se incluye a las escuelas que en el 2008 y 2009 reportaron que su totalidad de estudiantes eran quechua hablantes.

Cuadro 4: Test de diferencia en medias

Variable	Controles	Tratados	Dif.	N
Número de docentes	2.635 (0.0724)	3.005 (0.106)	0.319 (0.0873)***	7963
% de docentes nombrados	0.707 (0.026)	0.774 (0.0335)	0.0147 (0.0129)	7963
% de profesores con contrato mayor a un año	0.174 (0.0226)	0.119 (0.0223)	-0.0159 (0.0102)	7963
% de docentes en nivel magisteria: I-II	0.647 (0.0296)	0.603 (0.0632)	-0.00237 (0.0146)	6598
% de docentes en nivel magisterial: III-V	0.249 (0.0276)	0.311 (0.0641)	0.00909 (0.012)	6598
Ratio estudiantes por docente	19.82 (0.593)	19.79 (0.685)	-0.0735 (0.273)	7963
Nro de grados dictados en la IE	5.298 (0.0469)	5.489 (0.0845)	0.0705 (0.0313)**	7963
Promedio de alumnos por sección	9.511 (0.383)	10.09 (0.266)	0.605 (0.272)**	7959
1 si IE es unidocente	0.316 (0.0168)	0.244 (0.0379)	-0.0264 (0.0177)	7963
1 si IE es polidocente multigrado	0.59 (0.0144)	0.623 (0.0354)	-0.0176 (0.0157)	7963
1 si IE es polidocente completa	0.0938 (0.0101)	0.133 (0.00947)	0.044 (0.0108)***	7963
Número de estudiantes	47.31 (2.167)	52.84 (1.663)	4.394 (1.649)***	7963
Nro de minutos asignados a clase	237 (2.332)	235 (1.797)	-0.802 (0.549)	7962
1 si hay apoyo alimentario en IE	0.823 (0.0218)	0.837 (0.028)	0.00404 (0.0139)	7803
1 si hay control de salud en IE	0.311 (0.0258)	0.231 (0.0378)	-0.0503 (0.0171)***	7803
1 si IE tiene paredes de plástico o madera	0.0137 (0.00694)	0.0162 (0.00549)	0.00577 (0.00377)	7701
1 si IE tiene pared de piedra, quincha o adobe	0.859 (0.0184)	0.823 (0.0162)	-0.0216 (0.0119)*	7701
1 si IE tiene paredes de concreto	0.121 (0.0158)	0.16 (0.0167)	0.0188 (0.012)	7701
1 si la IE tiene piso de tierra o madera	0.6 (0.063)	0.559 (0.0832)	-0.0147 (0.0125)	7701
1 si la IE tiene piso de cemento	0.364 (0.0634)	0.416 (0.0828)	0.0185 (0.0123)	7701
1 si la IE tiene piso de loceta o parquet	0.0297 (0.00473)	0.0226 (0.00429)	-0.00157 (0.00566)	7701

Notas: Se incluyen las cohortes de los años 2008, 2009 y 2010. Las diferencias en medias controlan por efectos fijos a nivel de UGEL y cohorte. Los errores estándar son mostrados en paréntesis y están clusterizados a nivel de UGEL. La significancia al 1%,5% y 10% está denotada por los símbolos *, ** y ***, respectivamente.

Cuadro 4: Test de diferencia en medias (Continuación)

Variable	Controles	Tratados	Dif.	N
Nro de pabellones	1.899 (0.0671)	2.095 (0.0847)	0.0625 (0.0465)	7957
1 si IE tiene conexión eléctrica de red pública	0.508 (0.039)	0.56 (0.0278)	0.0628 (0.0182)***	7700
1 si IE tiene conexión de agua de red pública	0.44 (0.0355)	0.42 (0.0551)	0.0186 (0.0197)	7700
1 si IE tiene biblioteca	0.241 (0.0189)	0.234 (0.0368)	-0.0303 (0.0143)**	7700
Nro de aulas	3.833 (0.123)	4.129 (0.118)	0.264 (0.0976)***	7705
Nro de carpetas	8.772 (1.039)	8.857 (0.555)	-0.361 (0.545)	7677
Área del terreno de la IE	3345 (184.5)	3666 (408.6)	31.45 (109.1)	7444
1 si el centro poblado (CP) tiene desagüe	0.0769 (0.0112)	0.0955 (0.0248)	0.0242 (0.0102)**	7677
1 si el CP tiene posta médica	0.207 (0.0174)	0.218 (0.0206)	0.0213 (0.0116)*	7677
1 si el CP tiene teléfono público	0.132 (0.0113)	0.121 (0.0179)	0.0152 (0.0109)	7677
1 si el CP tiene cabina de Internet	0.00931 (0.00151)	0.0137 (0.00324)	0.00836 (0.00389)**	7677
1 si el CP tiene entidad bancaria	0.000792 (0.000354)	0.00151 (0.000774)	0.00153 (0.00102)	7705
1 si el CP tiene biblioteca	0.00634 (0.0015)	0.00723 (0.00258)	0.000114 (0.00178)	7677
1 si juntos llegó al distrito	0.835 (0.0353)	0.722 (0.0482)	-0.0246 (0.0213)	7963
1 si crecer llegó al distrito	0.913 (0.0205)	0.829 (0.0275)	-0.00719 (0.0255)	7963
distancia en minutos a la municipalidad	213.7 (17.18)	166 (20.82)	-35.97 (11.74)***	7963
distancia en minutos a la ugel	323.5 (29.56)	279.5 (37.29)	-16.89 (16.81)	7963
Altitud	3548 (83.27)	3460 (93.83)	-78.9 (18.52)***	7924
Nro de alumnos que tardan más de 30 min en llegar	4.167 (0.589)	3.316 (0.344)	-0.342 (0.296)	7963
Nro de alumnos que tardan más de 30 min en llegar	0.475 (0.052)	0.417 (0.0463)	-0.0247 (0.0221)	7963

Notas: Se incluyen las cohortes de los años 2008, 2009 y 2010. Las diferencias en medias controlan por efectos fijos a nivel de UGEL y cohorte. Los errores estándar son mostrados en paréntesis y están clusterizados a nivel de UGEL. La significancia al 1%, 5% y 10% está denotada por los símbolos *, ** y ***, respectivamente.

Evaluamos si las características de las escuelas varían entre aquellas que enseñan en castellano y lengua originaria. En el cuadro 4 mostramos los test de diferencia en medias tomando todas las escuelas en nuestra muestra para tercer grado de primaria. Realizamos dicho análisis agrupando a todas las cohortes disponibles (por eso el número elevado de observaciones)¹². El cuadro compara las características iniciales (recogidas en el primer grado de primaria) entre los colegios que son instruidos en quechua y en castellano, controlando por efectos fijos a nivel de UGEL y de cohorte. Esto implica que la diferencia en medias entre tratados y no tratados proviene de un promedio de diferencias al interior de cada UGEL. En general, existen varias características que difieren entre las escuelas que enseñan en lengua originaria y las que enseñan en castellano.¹³ En promedio las escuelas en las que los maestros enseñan en castellano son más grandes, en el sentido que tienen un mayor número de maestros, dictan más grados, y poseen más alumnos por sección, que aquellas en las que los maestros enseñan en lengua originaria. Es importante mencionar, no obstante, que estas escuelas tienen también más alumnos y que, cuando observamos las diferencias entre ratios de alumnos por maestros, no encontramos diferencias entre aquellas que enseñan en castellano y aquellas que enseñan en lengua originaria. En términos de indicadores de pobreza, no encontramos diferencias significativas en la provisión de apoyo alimentario entre las escuelas que enseñan en castellano y en lengua originaria. Sin embargo, sí encontramos diferencias significativas en la provisión de servicios de salud en la escuela. Adicionalmente, los centros poblados donde se ubican las escuelas que atienden a quechua-hablantes en castellano tienen una mayor probabilidad de tener desagüe, una posta médica y cabinas de internet, y se encuentran más cerca de las municipalidades, que los centros poblados donde se ubican las escuelas que atienden a quechua-hablantes en lengua originaria.

Los resultados anteriores indican que la lengua de instrucción en las escuelas está correlacionada con una serie de características que potencialmente afectan la

¹² Hay que destacar que según el cuadro 1, donde se indica que aproximadamente son 3000 los colegios incluidos en el estudio, el cuadro 4 debería presentar alrededor de 9000 observaciones. Esto no se observa por dos motivos: (i) Para estar seguros de que se está incluyendo colegios 100% quechua hablantes, para el análisis de inferencia se consideró solo a los colegios que en ambos años del tratamiento (primero y segundo) indicaron que el 100% de sus estudiantes era quechua hablantes y, (ii) se incluyó solo aquellas UGEL que presentan variación intracluster. En el Apéndice C se muestran los test de medias considerando solo a las escuelas que en el primer año de tratamiento manifestaron tener un 100% de alumnos quechua hablantes. Se puede observar que los resultados son muy similares a los del cuadro 4.

¹³ Se ha podido comprobar que las diferencias se mantienen regularmente si se realizan los test de medias para cohortes específicas. Estos resultados no están siendo reportados.

tasa de deserción y retención escolar. Por lo mismo, la estimación de los efectos de la lengua de instrucción debe controlar por dichas características. Afortunadamente, el Censo Escolar recopila una amplia gama de información acerca de las características de los docentes y escuelas, lo que nos permite separar los efectos de la lengua de instrucción de los efectos de estas otras características que están correlacionadas con la misma. No obstante, podrían existir factores no observables que conjuntamente afectan la lengua de instrucción en una escuela y las tasas de deserción y repetición escolar. Por ejemplo si los maestros capaces de enseñar en lengua originaria son, en promedio, más hábiles o motivados que sus contrapartes castellano hablantes, una regresión de lengua de instrucción sobre las tasas de deserción o retención escolar sobre estimaría el posible efecto negativo de aprender en una lengua distinta a la materna. En la medida que la habilidad y motivación del maestro sean capturadas por su nivel en la carrera magisterial (variables que son observables para este estudio), este tipo de problemas resultan menores. No obstante, resaltamos que los resultados que presentamos, por ahora, asumen que una vez que controlamos por una serie de características de las escuelas y alumnos, la razón por la que se enseña en castellano o una lengua originaria puede ser considerada aleatoria.

4. Estrategia empírica

La ecuación a estimar, en su versión reducida, es de la siguiente forma:

$$Y_{ti} = \beta T_{ti} + \alpha X_{ti} + \gamma f_i + \tau C_i + u_{ti} \dots (1)$$

Donde i se refiere a una determinada escuela, t a un momento en el tiempo, Y representa cualquiera de las variables de resultado (tasa de deserción o tasa de repetición), X es un vector de características observables, T es la variable de tratamiento (enseñanza en castellano), f es un efecto fijo por UGEL¹⁴, C es un efecto fijo por cohorte (que es lo mismo que un efecto fijo por año) y u refleja características no observadas que influyen en Y .

La identificación del efecto causal implica comparar —al interior de cada UGEL— los resultados de los colegios con 100% de niños quechua-hablantes que reciben clases en castellano con los resultados que hubieran ocurrido si en esos

¹⁴ Otras opciones de efectos fijos fueron municipalidades o provincias, no obstante, se consideró adecuado incluir los efectos fijos a nivel de UGEL ya que es la mínima unidad administrativa que agrupa un conjunto de colegios.

mismos colegios las clases hubieran sido en quechua. Naturalmente, como ambos escenarios no son observables al mismo tiempo, se necesita identificar un aproximado del escenario contrafactual.

Para poder aislar la influencia de factores externos y poder realizar la evaluación se debe cumplir el supuesto de independencia condicional. Este supuesto establece que el beneficio obtenido por los individuos será independiente a su decisión de participar o no del programa (Rosenbaum y Rubin, 1983):

$$Y_0 \perp T \rightarrow E[Y_{0i}|T_i = 1] = E[Y_{0i}|T_i = 0] \dots (2)$$

Si la asignación del tratamiento (que en nuestro caso es recibir en educación en castellano) ocurre de manera aleatoria, la hipótesis de independencia condicional se satisface y el grupo contrafactual estaría conformado por los colegios que recibieron educación en lengua originaria (no tratados). En este caso, la identificación del efecto promedio del tratamiento en los tratados (ATT) estaría dado por la comparación de promedios de las variables de resultado entre los colegios tratados y no tratados:

$$\begin{aligned} ATT &= E[Y_{1i} - Y_{0i}|T_i = 1] \\ ATT &= E[Y_{1i}|T_i = 1] - E[Y_{0i}|T_i = 1] \\ ATT &= E[Y_{1i}|T_i = 1] - E[Y_{0i}|T_i = 0] \dots (3) \end{aligned}$$

Lo que no es otra cosa que el cálculo del estimador β en la ecuación 1.

No obstante, como se mostró en el cuadro 4, hay características observables en las cuales los dos tipos de escuelas difieren (selección en observable). Por lo tanto, existen razones para suponer que la asignación del tratamiento en el presente estudio no ha ocurrido de manera aleatoria y que la hipótesis de independencia condicional no se cumpliría por lo que la comparación simple de ambos tipos de colegio introduciría un sesgo en el cálculo del efecto causal:

$$\begin{aligned} \beta &= E[Y_{1i}|T_i = 1] - E[Y_{0i}|T_i = 0] \\ \beta &= E[Y_{1i} - Y_{0i}|T_i = 1] + \{E[Y_{0i}|T_i = 1] - E[Y_{0i}|T_i = 0]\} \dots (4) \end{aligned}$$

El sesgo estaría dado por el término $E[Y_{0i}|T_i = 1] - E[Y_{0i}|T_i = 0]$ en la ecuación 4.

En consecuencia, la identificación del efecto causal implica la construcción de grupos similares de tratados y no tratados por métodos cuasi experimentales. En el presente estudio se proponen realizar un método de identificación cuasi experimental por Métodos de *matching*.

Este método consiste en calcular la probabilidad condicional (en características observables) de recibir el tratamiento para luego emparejar las unidades (que en nuestro caso son colegios) tratadas y no tratadas que tienen probabilidades similares. Bajo este enfoque, el efecto causal está dado por la diferencia promedio de resultados entre los colegios emparejados. La validez de este método exige dos requerimientos: (i) independencia condicional en observables y (ii) soporte común.

El concepto de independencia condicional en observables se refiere a que la asignación del tratamiento se ha dado estrictamente por factores observables e implica:

$$Y_0 \perp T|X \rightarrow ATT = E[Y_{0i}|T_i = 1, X] = E[Y_{0i}|T_i = 0, X] \dots (5)$$

No obstante, las características en las cuales pueden diferir los grupos de control y de tratamiento pueden ser muy diversas, lo que puede originar un problema de sobredimensionalidad. Respecto de esta complicación, Dehehnia y Wahba (1998) señalan que la probabilidad de recibir el tratamiento condicional en las características observables que lo determinan ($P(T = 1|X)$), es un indicador que resume de manera adecuada todas las variable observables. Por lo tanto, si $P(T = 1|X) = P(X)$, se tendría:

$$Y_0 \perp T|P(X) \rightarrow ATT = E[Y_{0i}|T_i = 1, P(X)] = E[Y_{0i}|T_i = 0, P(X)] \dots (6)$$

Por su parte, el concepto de soporte común se refiere a la existencia de un grupo de tratados y no tratados que tengan probabilidades similares de recibir el tratamiento para que el emparejamiento pueda ser realizado.

En este estudio, las características observables que se utilizarán para calcular la probabilidad de ser un colegio que recibe educación en castellano estarán a nivel de infraestructura, de acceso a servicios públicos, de características de los profesores y

de factores invariantes de localización (altitud, distancia a municipalidad, distancia a la UGEL, etc.).

4.1 Resultados

El cuadro 5 presenta los estimados del efecto causal de la instrucción en castellano sobre las tasas de deserción y repetición escolar en tercero, cuarto y quinto de primaria. Los resultados controlan por una serie de características de las escuelas, incluyendo efectos fijos a nivel de cohorte y de UGEL. Entre las características de las escuelas, se han incluido tanto las variables balanceadas como las no balanceadas, de acuerdo al test de medias realizado en el cuadro 4. Por otro lado, los efectos fijos a nivel de UGEL capturan las diferencias en las medias de las tasas de repetición y deserción a través de las diferentes UGELs que son constantes en el tiempo (entre diferentes cohortes de un mismo grado). Adicionalmente, también se incluye una especificación que contiene la interacción entre los efectos fijos de UGEL y cohorte, lo que estaría permitiendo que el efecto fijo a nivel de UGEL varíe dependiendo de la cohorte o año¹⁵. Es necesario precisar que la última especificación no aplica para las regresiones de quinto grado (ver cuadro 5), ya que en este grado solo hay una cohorte.

Como se puede observar en el cuadro 5, encontramos que la enseñanza en castellano, entre escuelas rurales que atienden a poblaciones donde el 100% de alumnos tiene al quechua como lengua materna, aumentan la tasas de deserción escolar en tercero de primaria y repetición de grado en cuarto de primaria en 0.92 y 1.23 puntos porcentuales, respectivamente. Dichas cifras representan el 24% y 12% de las probabilidades incondicionales de deserción y repetición escolar en dichos grados¹⁶. No encontramos efectos significativos para sobre la deserción escolar en cuarto ni en quinto de primaria; tampoco encontramos efectos significativos sobre la repetición de grado en tercero ni en quinto de primaria. La falta de resultados para quinto de primaria se debe, en parte, a un problema natural de selección en los estudios de deserción y repetición escolar. El grupo de alumnos que abandona la escuela o repite el año no es un grupo aleatorio sino que, probablemente esté compuesto de los alumnos menos hábiles o motivados. Una vez que estos desertan o

¹⁵ Esta especificación sería de la siguiente forma: $Y_{ti} = \beta T_{ti} + \alpha X_{ti} + \gamma f_i + \tau C_i + \theta f_i C_i + u_{ti}$. Especificación (2) en el cuadro 5.

¹⁶ Las probabilidades incondicionales de deserción en tercero y repetición en cuarto son de 3.7% y 10.25%.

repiten, el grupo restante estará compuesto por alumnos más hábiles y motivados cuyos resultados dependen menos de la lengua de instrucción y más de características personales.¹⁷ Hay que destacar que el cuadro 5 muestra que los resultados son muy similares bajo la segunda especificación, ya que en términos porcentuales, los nuevos resultados indican que la instrucción en castellano aumenta las tasas de deserción escolar en tercero y repetición de grado en cuarto en 22% y 13%, respectivamente.

Cuadro 5: Efecto de enseñanza en castellano sobre tasas de deserción y repetición por grado escolar – MCO con efectos fijos

Variable	Tercero		Cuarto		Quinto	
	(1)	(2)	(1)	(2)	(1)	(2)
Dependiente: Tasa de deserción	0.0092 (0.004)**	0.0082 (0.004)*	-0.0033 (0.003)	-0.0034 (0.003)	-0.0032 (0.006)	-
Dependiente: Tasa de repetición	0.0004 (0.006)	-0.0014 (0.001)	0.0123 (0.008)***	0.0135 (0.005)***	0.0020 (0.088)	-
<i>Número de observaciones</i>	6640		4469		1966	
Variables binarias regionales (A)	Sí		Sí		Sí	
Características de los profesores (B)	Sí		Sí		Sí	
Características administrativas de la IE (C)	Sí		Sí		Sí	
Características físicas y de equipamiento de la IE (D)	Sí		Sí		Sí	
Acceso a servicios en el centro Poblado(E)	Sí		Sí		Sí	
Variables fijas de localización de la IE (F)	Sí		Sí		Sí	

Nota: **A:** Número de docentes, % de docentes nombrados en 1ero, % de profesores con contrato mayor a un año en 1ero, % de docentes nombrados en 4to, % de docentes nombrados en 4to, % de profesores con contrato mayor a un año en 4to, % de docentes en nivel magisterial: I-II, % de docentes en nivel magisterial: III-V y Ratio estudiantes por docente; **B:** Apurímac, Ayacucho, Cusco, Huancavelica, Huánuco y Puno; **C:** Nro de grados dictados en la IE, Promedio de alumnos por sección, 1 si IE es polidocente multigrado, 1 si IE es polidocente completa, Nro de minutos asignados a clase, 1 si hay apoyo alimentario en IE en 1ero, 1 si hay apoyo alimentario en IE en 4to, 1 si hay control de salud en IE en 1ero, y 1 si hay control de salud en IE en 4to; **D:** 1 si pared de IE es de piedra, quincha o adobe, 1 si la IE tiene paredes de concreto, 1 si la IE tiene piso de tierra o madera, Nro de pabellones, 1 si IE tiene conexión eléctrica de red pública, 1 si IE tiene conexión de agua de red pública, 1 si IE tiene biblioteca en 1ero, 1 si IE tiene biblioteca en 4to, Área del terreno de la IE, nro de aulas, Nro de carpetas en 1ero y Nro de carpetas en 4to; **E:** 1 si el centro poblado (CP) tiene desagüe, 1 si el CP tiene posta médica, 1 si el CP tiene teléfono público, 1 si el CP tiene cabina de Internet, 1 si el CP tiene cabina de Internet, 1 si el CP tiene entidad bancaria, 1 si CP tiene biblioteca, 1 si Juntos llegó al distrito y 1 si Crecer llegó al distrito y **F** distancia en minutos a la municipalidad, Altitud, Nro de alumnos que tardan más de 30 min en llegar, Nro de CP de donde vienen alumnos que demoran más de 30 min.

(1) y (2) se refieren a las especificaciones que contienen los efectos fijos a nivel de UGEL y cohorte, sin interactuar e interactuados, respectivamente. Debido a que en quinto solo hay una cohorte, para este grado solo aplica la especificación (1).

IE denota Institución educativa.

Errores estándar en paréntesis. La significancia al 1%, 5% y 10% está denotada por los símbolos *, ** y ***, respectivamente.

En el cuadro 6 se presentan los resultados estimados por métodos de *matching*. Vale decir que para el cálculo de las probabilidades de recibir instrucción en castellano, se utilizó un modelo logit condicional que toma en cuenta los efectos fijos a nivel de UGEL¹⁸ (en los apéndices D y E se encuentran los resultados de los modelos de probabilidad así como así como histogramas que permiten visualizar el soporte común). Posteriormente el emparejamiento se realizó mediante la utilización del comando *psmatch2* (Leuven y Sianesi, 2003). Se presentan dos técnicas de emparejamiento: La del vecino más cercano (NN por sus siglas en inglés) y la de Kernel.

¹⁷ Ver Ciccone y García-Fontes (2014) para una discusión teórica al respecto.

¹⁸ Se utilizó el comando *clogit* del software estadístico Stata 12.

Como se puede apreciar, los resultados obtenidos por técnicas de *matching* resultan muy similares a los obtenidos en la identificación por MCO con efectos fijos. En lo que se refiere a la tasa de deserción en tercer grado, los efectos obtenidos usando el emparejamiento del vecino más cercano son similares a los obtenidos por MCO, mientras que lo reportados por el método de Kernel son de mayor magnitud. En cualquier caso, ya sea con uno o con otro método, los efectos resultan significativos en ambas identificaciones. Algo similar ocurre con los efectos en tasa de repetición en cuarto grado. Si bien el efecto es bastante menor (y solo es significativo al 10%) con la técnica del vecino más cercano en comparación a lo obtenido por MCO, cuando se hace uso del emparejamiento por Kernel se observan efectos de mayor magnitud con una mayor significancia. En conclusión, podemos decir que los efectos sobre deserción en tercero y retención en cuarto son medianamente robustos a las especificaciones utilizadas.

Cuadro 6: Efecto de enseñanza en castellano sobre tasas de deserción y repetición por grado escolar – Matching

Variable	Tercero		Cuarto		Quinto	
	NN	Kernel	NN	Kernel	NN	Kernel
Dependiente: Tasa de deserción	0.0098 (0.004)**	0.0127 (0.003)***	-0.0016 (0.004)	-0.0018 (0.003)	-0.0069 (0.006)	-0.0023 (0.004)
Dependiente: Tasa de repetición	-0.0060 (0.007)	-0.0008 (0.005)	0.0093 (0.007)*	0.0151 (0.005)***	-0.0026 (0.009)	0.0049 (0.007)
Número de observaciones	6640	6640	4469	4467	1966	1965
Variables binarias regionales (A)		Sí		Sí		Sí
Características de los profesores (B)		Sí		Sí		Sí
Características administrativas de la IE (C)		Sí		Sí		Sí
Características físicas y de equipamiento de la IE (D)		Sí		Sí		Sí
Acceso a servicios en el centro Poblado(E)		Sí		Sí		Sí
Variables fijas de localización de la IE (F)		Sí		Sí		Sí

Nota: A: Número de docentes, % de docentes nombrados en 1ero, % de profesores con contrato mayor a un año en 1ero, % de docentes nombrados en 4to, % de docentes nombrados en 4to, % de profesores con contrato mayor a un año en 4to, % de docentes en nivel magisterial: I-II, % de docentes en nivel magisterial: III-V y Ratio estudiantes por docente; B: Apurímac, Ayacucho, Cusco, Huancavelica, Huánuco y Puno; C: Nro de grados dictados en la IE, Promedio de alumnos por sección, 1 si IE es polidocente multigrado, 1 si IE es polidocente completa, Nro de minutos asignados a clase, 1 si hay apoyo alimentario en IE en 1ero, 1 si hay apoyo alimentario en IE en 4to, 1 si hay control de salud en IE en 1ero, y 1 si hay control de salud en IE en 4to; D: pared de IE es de piedra, quincha o adobe, 1 si la IE tiene paredes de concreto, 1 si la IE tiene piso de tierra o madera, Nro de pabellones, 1 si IE tiene conexión eléctrica de red pública, 1 si IE tiene conexión de agua de red pública, 1 si IE tiene biblioteca en 1ero, 1 si IE tiene biblioteca en 4to, Área del terreno de la IE, nro de aulas, Nro de carpetas en 1ero y Nro de carpetas en 4to; E: 1 si el centro poblado (CP) tiene desagüe, 1 si el CP tiene posta médica, 1 si el CP tiene teléfono público, 1 si el CP tiene cabina de Internet, 1 si el CP tiene cabina de Internet, 1 si el CP tiene entidad bancaria, 1 si CP tiene biblioteca, 1 si Juntos llegó al distrito y 1 si Crecer llegó al distrito y F: distancia en minutos a la municipalidad, Altitud, Nro de alumnos que tardan más de 30 min en llegar, Nro de CP de donde vienen alumnos que demoran más de 30 min.

IE denota Institución educativa, NN denota vecino más próximo por sus siglas en inglés

Errores estándar en paréntesis. La significancia al 1%, 5% y 10% está denotada por los símbolos *, ** y ***, respectivamente.

5. Conclusiones y recomendaciones de política

En este trabajo hemos investigado el efecto de aprender en una lengua distinta a la materna en los primeros años de la escuela primaria sobre la retención y

deserción escolar. Para ello, explotamos la falta de una política¹⁹ sobre las lenguas de instrucción en las escuelas en el Perú, lo que generó un desajuste entre las lenguas maternas de los alumnos y las de instrucción de los maestros.

Los resultados indican que, para poblaciones cuya lengua materna es el quechua, recibir instrucción únicamente en castellano durante los dos primeros años de la escuela primaria aumenta la probabilidad de que un alumno deserte la escuela en tercer grado en 0.92 puntos porcentuales y de que un alumno desaprobe el año en cuarto de primaria en 1.22 puntos porcentuales. Estas cifras representan el 24% y 12% de las probabilidades incondicionales de desertar en tercero de primaria y repetir el grado en cuarto de primaria, respectivamente. Ambos resultados indican que el desajuste lingüístico está asociado con efectos grandes, negativos y significativos sobre la probabilidad que un alumno quechua-hablante decida quedarse en la escuela y que pueda terminarla a tiempo. En el Perú, la educación primaria es un derecho y obligación, por lo que la deserción escolar supone un problema grave. Nuestro estudio sugiere que el desajuste lingüístico generado al interior de la administración educativa del Estado Peruano contribuye a dicho problema. A su vez, el estudio sugiere naturalmente una manera de solucionarlo: promover el ajuste lingüístico.

La promoción de dicho ajuste lingüístico requiere de: (i) una norma que lo exija, (ii) unidades ejecutoras que cumplan con la misma y (iii) la existencia de una oferta adecuada de maestros capaces de enseñar en lengua originaria. En el Perú, ya se han tomado medidas en estas tres direcciones.

Desde el 2010, con la promoción de la Directiva 012-2010, la contratación de maestros requiere que los candidatos manejen adecuadamente la lengua materna de sus potenciales alumnos (numeral 6.3, inciso e.5). Sin embargo, la directiva no se pronuncia acerca de los maestros ya contratados y que desconocen la lengua materna de sus alumnos. Sería prudente evaluar la re-asignación de maestros al interior de las regiones del Perú de manera que se pueda favorecer el ajuste lingüístico con los maestros ya contratados. Adicionalmente, se recomienda promover la capacitación en el conocimiento de lenguas originarias entre los maestros que desconocen la lengua materna de la población a la que atienden. La promoción de estas políticas recae sobre el MINEDU. Asimismo, la ejecución de la última requerirá de un esfuerzo

¹⁹ De acuerdo a la Defensoría del pueblo (2011), por lo menos hasta antes del 2011 no había una política clara.

conjunto de las DREs y los Institutos Pedagógicos Superiores con especialidades en educación bilingüe, quienes cuentan con el capital humano para su implementación.

A pesar que desde el 2010 la norma que rige la contratación de maestros requiere que éstos dominen la lengua de sus futuros alumnos, en la práctica los comités de contratación no han logrado acatar la norma adecuadamente (Defensoría del Pueblo, 2011). Resulta fundamental concientizar a las DREs sobre las consecuencias negativas que tiene el desajuste lingüístico sobre el resultado escolar. El presente trabajo porta evidencia en este sentido que puede y debe ser utilizada para la promoción del ajuste lingüístico. Del mismo modo, se recomienda la creación de mecanismos fiscalizadores, dependientes del MINEDU, que velen por el cumplimiento de las normas de contratación de maestros.

En cuanto a la oferta adecuada de servicios de educación bilingüe, saludamos los esfuerzos del MINEDU²⁰, en los últimos tres años, que se han traducido en la sistematización de la oferta de profesores y colegios EIB, la producción de materiales educativos en una mayor cantidad de lenguas originarias y la apertura de un mayor número de especializaciones e institutos pedagógicos para formar y capacitar a profesores EIB (Burga, 2013). Estos esfuerzos se deben seguir llevando adelante sin perder de vista que lo óptimo es una educación bilingüe que permita que los estudiantes adquieran sin mayor dificultad conocimientos en su lengua originaria sin dejar de lado el aprendizaje del castellano como segunda lengua.

En Perú, el castellano es la lengua franca en la capital, en la administración del estado, y en los principales negocios. En ese sentido, el conocimiento del castellano podría tener un retorno beneficioso en el mercado laboral. En la medida que los niños quechua-hablantes que son instruidos sólo en castellano tienen un mejor manejo de dicha lengua que sus contrapartes quechua-hablantes que aprenden principalmente en quechua, los primeros podrían tener salarios más elevados: asociados al dominio de la lengua de poder. Angrist y Lavy (1997) encuentran que el cambio en la lengua de instrucción en las escuelas en Marruecos, de francés a árabe, redujo el dominio escrito del francés y los salarios de los alumnos sujetos a este cambio de política. Por lo tanto, si bien una política orientada al ajuste lingüístico en las escuelas rurales puede fomentar la escolaridad, es importante tomar en cuenta los potenciales efectos que ello pueda tener en el mercado laboral. Lamentablemente, la falta de datos que

²⁰ A través de la Dirección General de Educación Intercultural Bilingüe y Rural.

permitan seguir a los niños peruanos desde la escuela al mercado laboral imposibilita analizar estas consideraciones.

En este estudio la unidad de observación ha sido la escuela. Utilizando el Censo Escolar hemos podido recopilar una gran cantidad de información sobre las mismas y los maestros que trabajan en ellas. Sin embargo, la información sobre los niños en las escuelas es escasa. Evidentemente, los ingresos familiares, tiempo que los padres le dedican a sus hijos, estado de salud de padres y los niños, son factores importantes de resultados escolares. Un estudio que tome en cuenta estos factores sólo es posible si se tiene información a nivel de alumnos. La construcción bases con esta característica no es una necesidad específica a este estudio sino virtualmente de todos los estudios que se interesen por resultados educativos. Adicionalmente, dicha base de datos debería tener una estructura de panel que permita seguir a los estudiantes en el tiempo. Ello no sólo facilitaría investigaciones académicas, sino que permitiría al Ministerio de Educación y las Direcciones Regionales de Educación observar qué niños están abandonando las escuelas y en qué condiciones: identificar a estos niños permitiría diseñar políticas específicas a sus necesidades que fomenten la escolaridad.

6. Plan de incidencia

Nuestro plan de incidencia se basa en los tres puntos señalados en la sección anterior sobre la promoción del ajuste lingüístico. Los puntos (i) y (iii) señalan la necesidad de la promoción de una norma que exija dicho ajuste y de una oferta adecuada de maestros bilingües, respectivamente. La ejecución de ambas recae sobre el MINEDU. Como se mencionó anteriormente, el MINEDU ya ha tomado acciones en estas direcciones, promulgando la Directiva 012-2010 y ampliando la oferta de instituciones formativas con especialidad en educación bilingüe. Nuestra investigación provee soporte científico para la promoción del ajuste lingüístico y, en ese sentido, respalda la reciente labor del MINEDU. Con el propósito de consolidar las políticas existentes que favorecen el ajuste lingüístico y de promover las recomendaciones de política en la sección anterior, difundiremos nuestra investigación entre el personal del MINEDU, a través de la DIGEBIR.

El punto (ii) sobre la promoción del ajuste lingüístico resalta la necesidad de unidades ejecutoras que cumplan con las normas –como la contratación de maestros bilingües. En el contexto peruano, son las direcciones regionales de educación las

responsables de la contratación de maestros y de garantizar que estos dominen la lengua materna mayoritaria en las poblaciones a las que atenderían. Con el propósito de llamar la atención sobre la importancia del ajuste lingüístico en el proceso de contratación de maestros, difundiremos este trabajo entre las DREs de las regiones que atienden a una cantidad significativa de alumnos que tienen como lengua materna una lengua originaria.

La difusión de este trabajo, tanto en el MINEDU como las DREs se hará, principalmente, por medios electrónicos. Adicionalmente, se mantendrán reuniones presenciales o virtuales con las autoridades que lo requieran.

Finalmente, se difundirá este trabajo entre la comunidad académica nacional a través de presentaciones y seminarios.

Cronograma:

- Difusión MINEDU: Diciembre, 2014
- Difusión DREs: Febrero, 2015
- Difusión comunidad académica: Septiembre, 2015

Bibliografía

Anghel, B., A. Cabrales, & J.M. Carro. (2012) Evaluating a bilingual education program in Spain: the impact beyond foreign language learning. WP Universidad Carlos III.

Angrist, J., A. Chin, & R. Godoy. (2008) Is Spanish-Only schooling responsible for the Puerto Rican language gap? *Journal of Development Economics*, 85, 105-128.

Angrist, J. y V. Lavy. (1997) The effect of a change in language of instruction on the returns to schooling in Morocco, *Journal of Labor Economics*, 15(1), S48-S76.

Angrist J. & Pischke (2008) *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Burga E. (2013) ¿Qué hemos avanzado en la implementación de la Educación Intercultural Bilingüe? DIGEIBIR. Disponible en:

<http://www.digeibir.gob.pe/articulos/%C2%BFqu%C3%A9-hemos-avanzado-en-la-implementaci%C3%B3n-de-la-educaci%C3%B3n-intercultural-biling%C3%BCe>

Baker, S.A., I. Kovelman, E. Bialystok, & L.A. Petitto (2003) Bilingual children's complex linguistic experience yields a cognitive advantage. *Published Abstracts of the Society for Neuroscience*, 1506.

Chin, A., N. Meltem Daysal, & S.A. Imberman (2012) Impact of bilingual education programs on limited English proficient students and their peers: Regression discontinuity evidence from Texas. NBER Working Paper 18197.

Ciccione, A. y W. García-Fontes (2014) Gender peers effects in school, a birth cohort approach. *Manuscrito*.

Cueto, S. & W. Secada (2003) Eficacia escolar en escuelas bilingües en Puno, Perú. *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 1(1), 1-23.

Cueto, S., G. Guerrero, J. León, E. Seguin, & I. Muñoz (2009) Explaining and overcoming marginalization in education: a focus on ethnic/language minorities in Peru.

Documento de trabajo preparado para el Reporte de Monitoreo del 2010 de Educación para Todos.

Defensoría del pueblo (2011). Aportes para una Política Nacional de Educación Intercultural Bilingüe a favor de los pueblos indígenas del Perú. Informe defensorial Nro 152.

Defensoría del pueblo (2013) Avances y desafíos en la implementación de la política de educación intercultural bilingüe 2012-2013. Informe defensorial Nro 163.

Dehejia, R. H., and S. Wahba (1998) Causal Effects in Non-experimental Studies: Re-evaluating the Evaluation of Training Programs. NBER Working Papers 6586. Cambridge, MA, US: National Bureau of Economic Research.

Guerrero, G., J. León, E. Rosales, M. Zapata, S. Freire, V. Saldarriaga, y S. Cueto. (2012) Young Lives School Survey in Peru: Design and Initial Findings. Working Paper 92 – Young Lives.

Jain, T. (2013) Common tongue: The impact of language on educational outcomes. Working Paper – Indian School of Business.

Khandker S., Koolwal G. & Samad H. (2010) Handbook on Impact Evaluation: Quantitative Methods and practices. The World Bank.

Kim, K.H., N. R. Relkin, K-M Lee, & J. Hirsch (1997) Distinct cortical areas associated with native and second languages. *Nature*, 388, 171-174

Knudsen, E.I. (2004) Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience*, 16(8), 1412-1425.

Kovelman, I. & L.A. Petitto (2003) States of language development in bilingual children exposed to their other language at different ages. Poster presentado en la conferencia de la Sociedad de Neurociencias, Orlando, FL.

Kovelman, I., S.A. Baker, & L.A. Petitto (2008) Age of first bilingual language exposure as a new window into bilingual reading development. *Bilingualism*, 11(2), 203-223.

Leuven, E. & B. Sianesi. (2003). "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing".

<http://ideas.repec.org/c/boc/bocode/s432001.html>.

Matsudaira, J.D. (2005) Sinking or swimming? Evaluating the impact of English immersion versus Bilingual Education on Student Achievement. University of California, Berkeley, mimeo.

Petitto, L.A. & K. Dunbar (2009) Educational Neuroscience: New Discoveries from Bilingual Brains, Scientific Brains, and the Educated Mind. *Mind, Brain, and Education*, 3(4), 185-197.

Petitto, L.A., I. Kovelman, & U. Harasymowycz (2003) Bilingual language development: Does learning the new damage the old? Abstracts of the Society for Research in Child Development, Tampa, FL.

Ramachadran, R. (2013) Language use in education and primary schooling attainment: Evidence from a natural experiment in Ethiopia. Working Paper 34 – Universidad Autónoma de Barcelona.

Rosenbaum, P. y D. Rubin (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70 (1), 41-55.

Sims, C. (1998) Stickiness. *Carnegie-Rochester Conference Series on Public Policy*, 49(0), 317-356.

Sims, C. (2003) Implications of rational inattention. *Journal of Monetary Economics*, 50(3), 665-690.

Slavin, R.E., N. Madden, M. Calderón, A. Chamberlain, & M. Hennesy (2011) Reading and language outcomes of a five-year randomized evaluation of transitional bilingual education. *Educational Evaluation and Policy Analysis*, 33(1), 47-58.

Trivelli, C. (2005) Los hogares indígenas y la pobreza en el Perú: Una Mirada a partir de la información cuantitativa. Documento de trabajo IEP Nro 141.

Veldkamp, L. (2011) Information choice in macroeconomics and finance. Princeton University Press.

Wartenburger, I., H.R. Heekeren, J. Abutalebi, S.F. Cappa, A. Villringer, & D. Perani (2003) Early setting of grammatical processing in the bilingual brain. *Neuron*, 37(1), 159-170.

Apéndice

Apéndice A:

i. En la actualidad, ¿se busca contratar maestros que puedan enseñar en la lengua materna de las poblaciones que atienden? ¿Cómo ha cambiado esto en relación a los últimos 15 años? Por favor, describir brevemente el proceso de contratación pasado y actual.

Si, se busca maestros que pueden enseñar en la lengua materna, en caso de [región] en quechua chanca (pues el contexto social es parecido entre [nombran regiones]), [...]

En estos últimos 15 años, la población [región], ha tomado conciencia de la importancia de la interculturalidad, anteriormente, las autoridades educacionales, los gobiernos locales regionales, no dieron importancia, mas al contrario se oponían a las prácticas ancestrales, como costumbres, tradiciones, a la cosmovisión andina, y otros, actualmente con la insistencia, exigencia en las capacitaciones, especialmente del MED, (Dirección General de Educación Intercultural Bilingüe y Rural) y la Dirección Regional de Educación a través de los Especialistas en EIBIR, se viene desarrollando acciones educativas utilizando la interculturalidad y bilingüe en el desarrollo de las sesiones de aprendizaje, previo estudio socio y psicolingüístico y tratamiento de lenguas, pero existe una buena cantidad de Instituciones Educativas especialmente de la zona rural y urbano marginal, no están siendo atendidos.

ii. ¿Cree usted que la falta de maestros que pueden enseñar en lengua originaria es la razón principal por la que una proporción significativa de niños, que tienen una lengua originaria como lengua materna, aprenden exclusivamente en castellano?

Si, lo creo, porque es una verdad, especialmente en las zonas rurales pues [región] cuenta en un 88% aproximado zona rural, y a la fecha la comunicación preferentemente de pobladores antiguos es el quechua o runa simi y los niños viven en ese medio por lo tanto su comunicación o su primera lengua es quechua, y cuando asiste a una Institución Educativa, la enseñanza es en castellano, a este problema se agrega que lo maestros no tiene conocimiento mínimo de quechua especialmente en los primeros grados, y lastimosamente a estos niños se evalúa en castellano y el resultado lógicamente es negativo, por ,lo tanto existe una necesidad de que los

maestros que trabajan en la región de [región] deben ser quechuahablantes e hispanohablantes.

iii. (Responder esta pregunta sólo si su respuesta fue afirmativa en la pregunta ii) Dado que la cantidad de maestros capaces de enseñar en lengua originaria es menor que la que se necesita para atender a todas las poblaciones que tienen una lengua originaria como lengua materna, ¿de qué depende que un maestro capaz de enseñar en lengua originaria termine enseñando en una escuela u otra? ¿Es el maestro quien decide en qué escuela prefiere enseñar? ¿La DRE le asigna al maestro una un escuela al azar entre las que necesitan un maestro que enseñe en lengua originaria?

Depende de cada maestro que tenga vocación de servicio, pues actualmente se selecciona a maestros bilingües, la evaluación no es suficientemente riguroso, existe oposición de algunos maestros especialmente del Sindicato (SUTE), en primera instancia se le asigna a un maestro quechuahablante en una IIEE bilingüe, pues después de dos años de permanencia en ese medio el maestro solicita, su reasignación a otra IE, y es reasignado a una IE. Hispanohablante o quechuahablante, en estos de movimiento de personal no existe la evaluación del dominio de lenguas a los docentes y es así donde el maestro termina en una IE. Diferente, de igual modo para con docentes hispanohablantes con reasignación llega a una IE quechuahablantes.

Si, es verdad que el maestro decide donde trabajar, pero para ello, se selecciona a las IIEE. Bilingües a maestros quechuahablantes, esta selección es previa evaluación a través de las Unidades de Gestión Educativa Local UGELs.

[...]

Apéndice B:

Número de escuelas rurales según porcentaje de niños quechua-hablantes matriculados en el año 2011

	Resto	[70%; 80[[80%; 90[[90%; 100[100%	Total
Amazonas	642 100%	0 0%	0 0%	0 0%	0 0%	642 100%
Ancash	619 51%	5 0%	12 1%	31 3%	547 45%	1214 100%
Apurímac	146 22%	5 1%	21 3%	44 7%	434 67%	650 100%
Arequipa	239 81%	3 1%	3 1%	2 1%	48 16%	295 100%
Ayacucho	322 34%	8 1%	14 1%	29 3%	566 60%	939 100%
Cajamarca	1174 100%	0 0%	0 0%	0 0%	0 0%	1174 100%
Cusco	509 39%	8 1%	10 1%	42 3%	745 57%	1314 100%
Huancavelica	453 48%	2 0%	8 1%	47 5%	428 46%	938 100%
Huánuco	1015 90%	0 0%	3 0%	11 1%	105 9%	1134 100%
Ica	185 100%	0 0%	0 0%	0 0%	0 0%	185 100%
Junín	1337 99%	0 0%	0 0%	2 0%	14 1%	1353 100%
La Libertad	818 100%	0 0%	0 0%	0 0%	0 0%	818 100%
Lambaqueye	380 88%	0 0%	0 0%	1 0%	51 12%	432 100%
Lima	409 100%	0 0%	0 0%	0 0%	0 0%	409 100%
Loreto	1885 100%	0 0%	0 0%	0 0%	0 0%	1885 100%
Madre de Dios	140 100%	0 0%	0 0%	0 0%	0 0%	140 100%
Moquegua	74 76%	1 1%	0 0%	3 3%	19 20%	97 100%
Pasco	438 100%	0 0%	0 0%	0 0%	0 0%	438 100%
Piura	1419 100%	0 0%	0 0%	0 0%	0 0%	1419 100%
Puno	783 59%	3 0%	17 1%	25 2%	490 37%	1318 100%
San Martín	953 99%	0 0%	0 0%	0 0%	6 1%	959 100%
Tacna	82 100%	0 0%	0 0%	0 0%	0 0%	82 100%
Tumbes	64 100%	0 0%	0 0%	0 0%	0 0%	64 100%
Ucayali	624 100%	0 0%	0 0%	0 0%	0 0%	624 100%
Total	14710 79%	35 0%	88 0%	237 1%	3453 19%	18523 100%

Fuente: Elaboración propia en base a Censos escolares

Apéndice C:

Test de medias

Variable	Controles	Tratados	Dif.	N
Número de docentes	2.634 (0.0684)	2.995 (0.0981)	0.327 (0.0862)***	9332
% de docentes nombrados	0.715 (0.0255)	0.774 (0.0289)	0.0137 (0.0126)	9332
% de profesores con contrato mayor a un año	0.169 (0.0219)	0.121 (0.0203)	-0.0165 (0.0095)*	9332
% de docentes en nivel magisteria: I-II	0.646 (0.0288)	0.618 (0.0573)	-0.00145 (0.0146)	7729
% de docentes en nivel magisterial: III-V	0.253 (0.0271)	0.296 (0.0572)	0.0085 (0.0128)	7729
Ratio estudiantes por docente	19.7 (0.585)	19.84 (0.733)	-0.317 (0.267)	9332
Nro de grados dictados en la IE	5.297 (0.0469)	5.495 (0.0733)	0.0691 (0.0286)**	9332
Promedio de alumnos por sección	9.437 (0.371)	10.01 (0.291)	0.499 (0.258)*	9328
1 si IE es unidocente	0.317 (0.0165)	0.254 (0.0326)	-0.0235 (0.0166)	9332
1 si IE es polidocente multigrado	0.588 (0.0149)	0.611 (0.031)	-0.023 (0.0136)*	9332
1 si IE es polidocente completa	0.0946 (0.00969)	0.135 (0.00907)	0.0465 (0.011)***	9332
Número de estudiantes	46.94 (2.09)	52.45 (1.77)	3.782 (1.57)**	9332
Nro de minutos asignados a clase	236.9 (2.257)	235.4 (1.699)	-0.713 (0.51)	9330
1 si hay apoyo alimentario en IE	0.818 (0.0209)	0.807 (0.0277)	-0.0103 (0.0129)	9150
1 si hay control de salud en IE	0.312 (0.0246)	0.236 (0.0326)	-0.0497 (0.0169)***	9150
1 si IE tiene paredes de plástico o madera	0.0153 (0.00745)	0.0152 (0.00486)	0.00305 (0.00315)	8985
1 si IE tiene pared de piedra, quincha o adobe	0.855 (0.0186)	0.817 (0.0152)	-0.024 (0.0104)**	8985
1 si IE tiene paredes de concreto	0.123 (0.0153)	0.165 (0.0152)	0.0224 (0.0109)**	8985
1 si la IE tiene piso de tierra o madera	0.598 (0.0614)	0.548 (0.0742)	-0.0261 (0.0126)**	8985
1 si la IE tiene piso de cemento	0.366 (0.0619)	0.426 (0.0734)	0.0286 (0.0122)**	8985
1 si la IE tiene piso de loceta o parquet	0.0299 (0.00443)	0.0237 (0.00397)	-0.00164 (0.00509)	8985

Notas: Se incluyen las cohortes de los años 2008, 2009 y 2010 pertenecientes a colegios que en primer grado manifestaron tener un 100% de estudiantes quechua hablantes. Los errores estándar son mostrados en paréntesis y están clusterizados a nivel de UGEL. La significancia al 1%, 5% y 10% está denotada por los símbolos *, ** y ***, respectivamente.

Test de medias (*Continuación*)

Variable	Controles	Tratados	Dif.	N
Nro de pabellones	1.899 (0.0666)	2.084 (0.0815)	0.0692 (0.0429)	9323
1 si IE tiene conexión eléctrica de red pública	0.509 (0.0377)	0.562 (0.0287)	0.0743 (0.0171)***	8984
1 si IE tiene conexión de agua de red pública	0.44 (0.0354)	0.422 (0.0477)	0.0194 (0.0179)	8984
1 si IE tiene biblioteca	0.244 (0.0188)	0.232 (0.0318)	-0.0289 (0.014)**	8984
Nro de aulas	3.841 (0.117)	4.152 (0.115)	0.286 (0.0914)***	8989
Nro de carpetas	8.537 (0.951)	9.077 (0.506)	0.0747 (0.465)	8940
Área del terreno de la IE	3327 (184.9)	3564 (353.3)	33.37 (100.8)	8707
1 si el centro poblado (CP) tiene desagüe	0.0799 (0.011)	0.102 (0.023)	0.0253 (0.00946)***	8940
1 si el CP tiene posta médica	0.211 (0.0168)	0.225 (0.0181)	0.0229 (0.011)**	8940
1 si el CP tiene teléfono público	0.132 (0.0107)	0.124 (0.0154)	0.012 (0.00953)	8940
1 si el CP tiene cabina de Internet	0.0104 (0.00141)	0.0133 (0.00303)	0.00665 (0.0035)*	8940
1 si el CP tiene entidad bancaria	0.00108 (0.0004)	0.00146 (0.000655)	0.00118 (0.000888)	8989
1 si el CP tiene biblioteca	0.00682 (0.00144)	0.00831 (0.00304)	0.00113 (0.00261)	8940
1 si juntos llegó al distrito	0.826 (0.0365)	0.709 (0.0461)	-0.0349 (0.0197)*	9332
1 si crecer llegó al distrito	0.906 (0.0213)	0.826 (0.0263)	-0.0106 (0.0232)	9332
distancia en minutos a la municipalidad	214.9 (16.89)	173.2 (19.37)	-39.65 (12.32)***	9332
distancia en minutos a la ugel	328.7 (29.27)	300.8 (40.03)	-23.95 (17.9)	9332
Altitud	3541 (81.55)	3407 (91.79)	-106.4 (23.48)***	9286
Nro de alumnos que tardan más de 30 min en llegar	4.105 (0.562)	3.159 (0.333)	-0.441 (0.295)	9332
Nro de alumnos que tardan más de 30 min en llegar	0.467 (0.0501)	0.396 (0.0439)	-0.028 (0.0214)	9332

Notas: Se incluyen las cohortes de los años 2008, 2009 y 2010 pertenecientes a colegios que en primer grado manifestaron tener un 100% de estudiantes quechua hablantes. Las diferencias en medias controlan por efectos fijos a nivel de UGEL y cohorte. Los errores estándar son mostrados en paréntesis y están clusterizados a nivel de UGEL. La significancia al 1%, 5% y 10% está denotada por los símbolos *, ** y ***, respectivamente.

Apéndice D:

Probabilidad de que la IE instruya en castellano (Modelo *Logit* condicional que controla por efectos fijos a nivel de UGEL)

Variable	Tercero	Cuarto	Quinto
Número de docentes	0.196 ***	0.170 **	0.207 ***
	0.052	0.054	0.071
% de docentes nombrados en primero	-0.237	-0.186	-0.029
	0.152	0.214	0.296
% de profesores con contrato mayor a un año en primero	-0.324 **	-0.339 *	-0.238
	0.160	0.175	0.244
% de docentes nombrados en grado analizado	-0.075	-0.100	0.009
	0.127	0.157	0.141
% de profesores con contrato mayor a un año en grado analizado	-0.148	-0.016	0.004
	0.196	0.205	0.174
% de docentes en nivel magisteria: I-II en primero	0.026	-0.029	-0.116
	0.159	0.201	0.240
% de docentes en nivel magisterial: III-V en primero	0.027	-0.020	-0.146
	0.171	0.196	0.262
1 si la variable de nivel magisterial es valor perdido	-0.124	-0.177	-0.233
	0.188	0.190	0.228
Ratio estudiantes por docente en primero	0.013	0.015	0.020 *
	0.009	0.010	0.011
Nro de grados dictados en la IE en primero	0.001	-0.030	-0.003
	0.046	0.056	0.067
Promedio de alumnos por sección en primero	-0.028	-0.020	-0.021
	0.020	0.021	0.023
Colegio Poli. Multigrado en primero	-0.139	-0.067	-0.286
	0.132	0.151	0.203
Colegio Poli. Completo en primero	0.063	0.133	-0.292
	0.234	0.275	0.411
Número de minutos asignados a clase	-0.001	0.000	-0.003
	0.003	0.003	0.004
IE recibe apoyo alimento en primero	0.065	0.054	-0.088
	0.116	0.127	0.170
Apoyo alimentario en grado analizado	-0.284 **	0.019	0.147
	0.116	0.133	0.157
IE recibe control de salud en primero	-0.272 ***	-0.235 *	-0.197 *
	0.091	0.128	0.111
IE recibe control de salud en grado analizado	-0.029	0.003	0.038
	0.068	0.098	0.124
1 si IE tiene pared de piedra, quincha o adobe	-0.228	-0.078	0.324
	0.303	0.419	0.316
1 si IE tiene paredes de concreto	-0.194	0.019	0.471
	0.323	0.444	0.357
1 si IE tiene piso de tierra	0.012	-0.085	-0.286
	0.233	0.310	0.403
Número de pabellones	-0.020	-0.063	-0.133 ***
	0.047	0.042	0.044

Nota: Los errores estándar son mostrados en el segundo renglón y están clusterizados a nivel de UGEL. La significancia al 1%, 5% y 10% está denotada por los símbolos *, ** y ***, respectivamente.

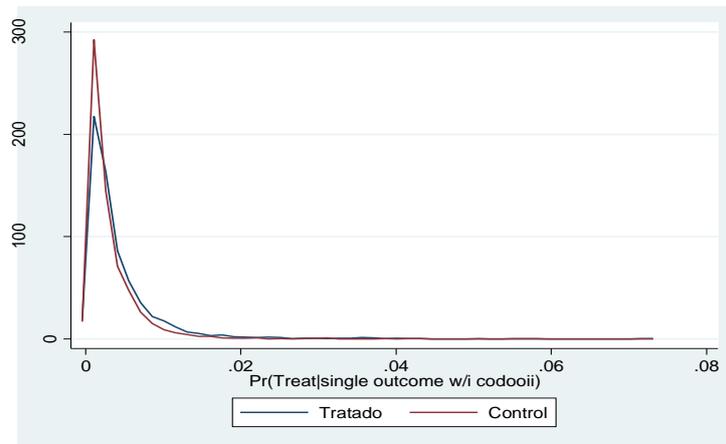
Probabilidad de que la IE instruya en castellano (Modelo *Logit* condicional que controla por efectos fijos a nivel de UGEL) (Continuación)

Variable	Tercero	Cuarto	Quinto
IE tiene conexión eléctrica en primero	0.138 0.100	0.151 0.104	0.089 0.102
IE tiene conexión de agua en primero	-0.002 0.108	-0.015 0.115	-0.011 0.140
IE tiene biblioteca en primero	-0.245 *** 0.074	-0.207 ** 0.081	-0.204 0.127
IE tiene biblioteca en grado analizado	0.016 0.100	0.158 0.106	0.067 0.150
Número de aulas en primero	-0.019 0.027	-0.038 0.028	-0.003 0.042
Número total de carpetas en primero	-0.004 *** 0.002	-0.002 0.002	0.000 0.004
Número total de carpetas en grado analizado	-0.002 0.002	-0.004 0.003	-0.002 0.003
Área del terreno	0.000 0.000	0.000 0.000	0.000 0.000
1 si el área del terreno es valor perdido	0.170 0.169	0.136 0.158	0.141 0.197
1 si el centro poblado (CP) tiene desagüe	0.181 0.136	0.310 * 0.182	0.347 0.251
1 si el CP tiene posta médica	-0.018 0.101	0.058 0.109	-0.048 0.142
1 si el CP tiene teléfono público	0.150 0.114	0.128 0.119	0.273 ** 0.138
1 si el CP tiene cabina de Internet	0.525 * 0.317	0.562 * 0.297	1.230 0.738
1 si el CP tiene entidad bancaria	1.355 ** 0.770	1.565 * 0.906	1.159 1.028
1 si el CP tiene biblioteca	0.076 0.243	-0.105 0.291	0.584 0.497
1 si juntos llegó al distrito	-0.288 * 0.172	-0.332 0.252	-0.273 0.247
1 si crecer llegó al distrito	0.189 0.315	0.095 0.372	0.047 0.348
Distancia en minutos a la municipalidad	0.000 0.000	0.000 0.000	0.000 0.000
Altitud	0.000 ** 0.000	0.000 *** 0.000	0.000 ** 0.000
Nro de alumnos que tardan más de 30 min en llegar	-0.019 ** 0.009	-0.017 *** 0.010	-0.021 * 0.012
Nro de alumnos que tardan más de 30 min en llegar	0.021 0.096	-0.009 * 0.097	0.007 0.104
Cohorte 2008	1.363 *** 0.199	0.707 *** 0.177	
Cohorte 2009	0.577 *** 0.096		
Número de observaciones	6564	4653	2451

Nota: Los errores estándar son mostrados en el segundo renglón y están clusterizados a nivel de UGEL. La significancia al 1%,5% y 10% está denotada por los símbolos *, ** y ***, respectivamente.

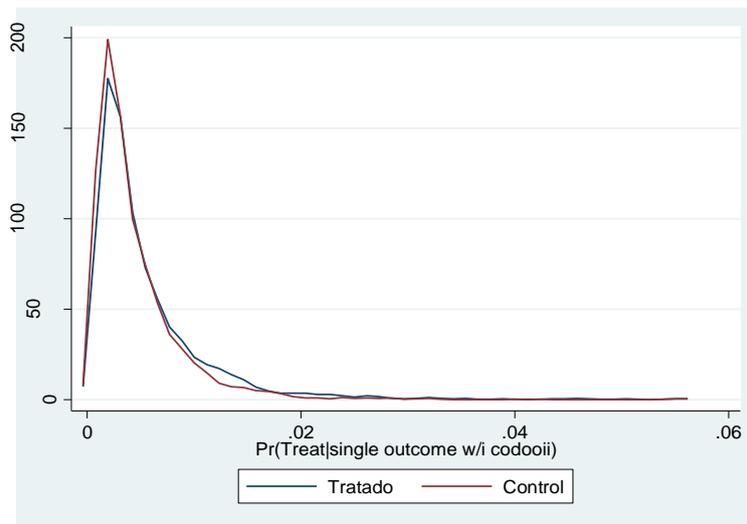
Apéndice E:

Soporte común: Tercer grado



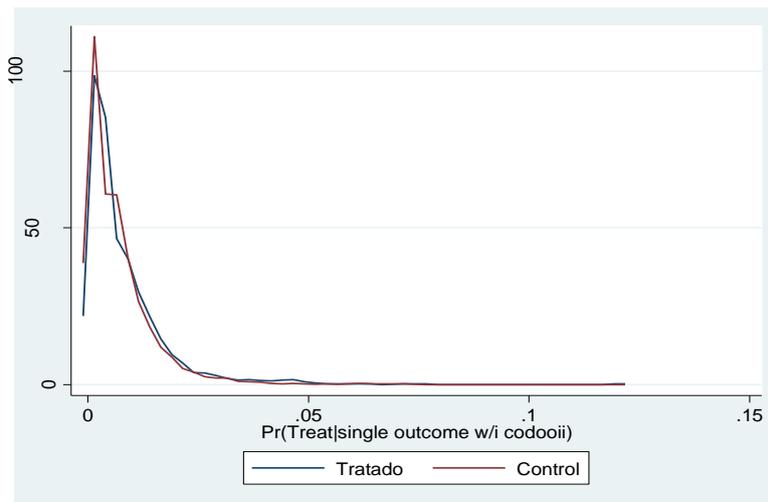
Nota: codoii es el código de UGEL

Soporte común: Cuarto Grado



Nota: codoii es el código de UGEL

Soporte común: Quinto Grado



Nota: codoii es el código de UGEL