

**PONTIFICIA UNIVERSIDAD
CATÓLICA DEL PERÚ**

FACULTAD DE CIENCIAS SOCIALES



El efecto del tiempo de desplazamiento de la vivienda al centro educativo sobre
el rendimiento escolar en el área rural del Perú

Tesis para obtener el título profesional de Licenciado en Economía presentado
por:

Villanueva Hinojosa, Anibal Daniel

Asesor:

García Núñez, Luis Javier

Lima, 2023

Informe de Similitud

Yo, García Núñez, Luis Javier, docente de la Facultad de Ciencias Sociales de la Pontificia Universidad Católica del Perú, asesor(a) de la tesis/el trabajo de investigación titulado El efecto del tiempo de desplazamiento de la vivienda al centro educativo sobre el rendimiento escolar en el área rural del Perú del/de la autor (a)/ de los(as) autores(as) Villanueva Hinojosa, Anibal Daniel dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de 13%. Así lo consigna el reporte de similitud emitido por el software *Turnitin* el 29/11/2023.
- He revisado con detalle dicho reporte y la Tesis o Trabajo de Suficiencia Profesional, y no se advierte indicios de plagio.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lugar y fecha: Lima, 04 de diciembre del 2023

Apellidos y nombres del asesor / de la asesora: <u>García Núñez, Luis Javier</u>	
DNI: 09672621	Firma 
ORCID: 0000-0001-7674-5461	

Agradecimientos

Agradezco a Dios por tanto bien recibido a lo largo de mi vida y en mi etapa universitaria. Asimismo, agradezco a mi mamá, Deycy Hinojosa, por todo su amor y apoyo en todo momento de mi vida. Además, quiero agradecer a mi tía abuela Emma Hinojosa por su amor, consejos y apoyo en mi etapa universitaria. Considero importante mencionar sus nombres y apellidos, porque mis logros los comparto con ellas. También quiero agradecer a mi tío Raúl Hinojosa Ramos por su gran afecto y cariño hacia mi persona. Asimismo, agradezco al apoyo de toda mi familia por su amor, afecto y apoyo conmigo.

Quiero también agradecer a mis amistades más cercanas, especialmente, aquellas con las que entablé una amistad en la etapa universitaria, su cariño y afecto son un pilar fundamental. También, quiero agradecer al Padre Juan Bytton S.J., quien me ha guiado con afecto y amistad en esta etapa universitaria. Asimismo, quiero agradecer al Profesor Juan León y a mi asesor, Profesor Luis García, por su apoyo en mi tesis.

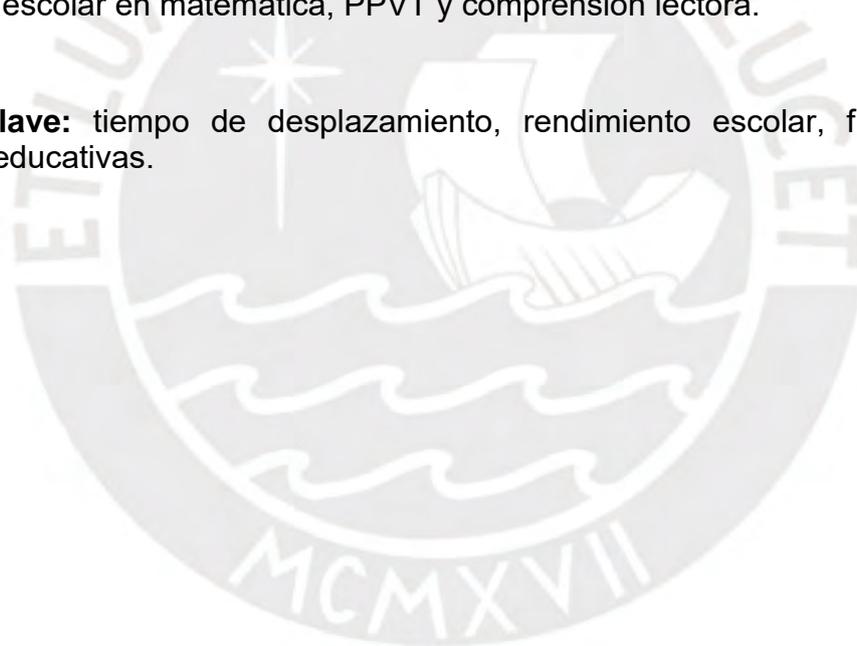
Agradezco al Centro de Asesoría Pastoral Universitaria (CAPU) por ser un espacio de encuentro, servicio y fe, pilares fundamentales en mi vida. También quiero agradecer a la PUCP por ser un espacio de encuentro, diálogo, debate y cultura.



Resumen

La presente investigación aporta a la discusión del efecto del tiempo de desplazamiento de la vivienda al centro educativo sobre el rendimiento escolar en el área rural del Perú. En ciertas zonas del área rural, las viviendas se encuentran dispersas, por lo que algunos escolares tienen que desplazarse largas distancias desde sus viviendas al centro educativo (ida y vuelta) para recibir una educación en el nivel de primaria o secundaria. En ese sentido, el tiempo que emplean para desplazarse al centro educativo puede disminuir el tiempo que los escolares destinan a sus horas de estudio, sea en el centro educativo y/o en la vivienda. Sin embargo, puede ocurrir que algunos escolares emplean un mayor tiempo de desplazamiento para asistir a un centro educativo lejano a su comunidad a cambio de recibir una educación de buena calidad, lo cual incide favorablemente en su rendimiento escolar. A partir de ello, surge la pregunta de esta investigación: ¿Cuál es el efecto del tiempo de desplazamiento de la vivienda al centro educativo sobre el rendimiento escolar en el área rural del Perú? La hipótesis plantea que, si aumenta el tiempo de desplazamiento de vivienda al centro educativo (ida y vuelta), disminuye el rendimiento escolar. Para responder a esta interrogante, se usa la base de datos Niños del Milenio, y se emplea un modelo de Efectos Fijos. Según los resultados obtenidos, no hay evidencia de un efecto del tiempo de desplazamiento sobre el rendimiento escolar en matemática, PPVT y comprensión lectora.

Palabras clave: tiempo de desplazamiento, rendimiento escolar, funciones de producción educativas.



Índice de contenidos

1.	Introducción	1
2.	Marco teórico.....	3
2.1.	Distribución del tiempo.....	3
2.2.	Funciones de producción educativas	5
2.2.1.	Especificación de insumos contemporáneos	6
2.2.2.	Especificación de insumos acumulados	7
2.2.3.	Especificación de valor añadido.....	8
3.	Revisión de literatura empírica	11
4.	Hechos estilizados.....	16
4.1.	Tiempo de desplazamiento	18
4.2.	Puntaje en pruebas escolares	21
4.2.1.	Puntaje en matemática	22
4.2.2.	Puntaje en PPVT	24
4.2.3.	Puntaje en comprensión lectora.....	26
4.3.	Gráficos de dispersión	27
5.	Hipótesis	31
6.	Metodología.....	32
6.1.	Variables a analizar.....	32
6.2.	Descripción estadística de las principales variables de análisis.....	34
6.3.	Análisis estadístico	38
6.4.	Modelo de estimación	39
7.	Estimaciones: Principales resultados.....	44
7.1.	Resultados en matemática.....	44
7.2.	Resultados en PPVT.....	47
7.3.	Resultados en comprensión lectora	49
7.4.	Balance de los resultados	51
8.	Limitaciones.....	53
9.	Conclusiones	54
10.	Referencias bibliográficas	56
11.	Anexos.....	60

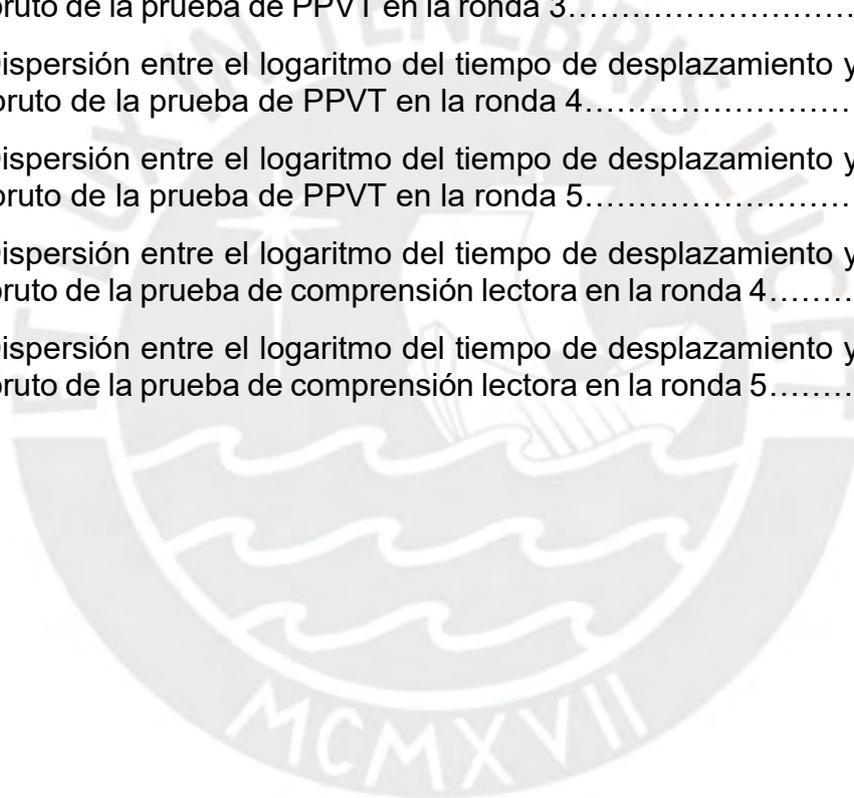
Índice de tablas

Tabla 1. Variables a analizar.....	32
Tabla 2. Estadística de las variables.....	35
Tabla 3. Variables asociadas al escolar.....	37
Tabla 4. Matriz de correlación de las variables en la ronda 3.....	38
Tabla 5. Matriz de correlación de las variables en la ronda 4.....	38
Tabla 6. Matriz de correlación de las variables en la ronda 5.....	39
Tabla 7. Resultados en matemática.....	45
Tabla 8. Test de Hausman para estimaciones en matemática.....	45
Tabla 9. Resultados en PPVT.....	48
Tabla 10. Test de Hausman para estimaciones en PPVT.....	48
Tabla 11. Resultados en comprensión lectora.....	50
Tabla 12. Test de Hausman para estimaciones en comprensión lectora.....	50
Tabla 13. Test de Breusch-Pagan para estimaciones en matemática.....	66
Tabla 14. Test de Hausman para estimaciones en matemática.....	66
Tabla 15. Test de Breusch-Pagan para estimaciones en PPVT.....	67
Tabla 16. Test de Hausman para estimaciones en PPVT.....	67
Tabla 17. Test de Breusch-Pagan para estimaciones en Comprensión lectora.....	68
Tabla 18. Test de Hausman para estimaciones en Comprensión lectora.....	68
Tabla 19. Estimaciones con dummies del tiempo de desplazamiento en matemática.....	69
Tabla 20. Estimaciones con dummies del tiempo de desplazamiento en PPVT.....	70
Tabla 21. Estimaciones con dummies del tiempo de desplazamiento en comprensión lectora.....	71

Índice de figuras

Figura 1. Tiempo de desplazamiento en el área rural en la ronda 3.....	19
Figura 2. Tiempo de desplazamiento en el área rural en la ronda 4.....	19
Figura 3. Tiempo de desplazamiento en el área rural en la ronda 5.....	19
Figura 4. Tipo de desplazamiento en la ronda 3.....	20
Figura 5. Tipo de desplazamiento en la ronda 4.....	21
Figura 6. Tipo de desplazamiento en la ronda 5.....	21
Figura 7. Puntaje en matemática en la ronda 3.....	23
Figura 8. Puntaje en matemática en la ronda 4.....	23
Figura 9. Puntaje en matemática en la ronda 5.....	24
Figura 10. Puntaje en PPVT en la ronda 3.....	25
Figura 11. Puntaje en PPVT en la ronda 4.....	25
Figura 12. Puntaje en PPVT en la ronda 5.....	25
Figura 13. Puntaje en comprensión lectora en la ronda 4.....	26
Figura 14. Puntaje en comprensión lectora en la ronda 5.....	27
Figura 15. Dispersión entre el tiempo de desplazamiento y el puntaje en matemática en la ronda 3.....	28
Figura 16. Dispersión entre el tiempo de desplazamiento y el puntaje en matemática en la ronda 4.....	28
Figura 17. Dispersión entre el tiempo de desplazamiento y el puntaje en matemática en la ronda 5.....	28
Figura 18. Dispersión entre el tiempo de desplazamiento y el puntaje en PPVT en la ronda 3.....	29
Figura 19. Dispersión entre el tiempo de desplazamiento y el puntaje en PPVT en la ronda 4.....	29
Figura 20. Dispersión entre el tiempo de desplazamiento y el puntaje en PPVT en la ronda 5.....	29
Figura 21. Dispersión entre el tiempo de desplazamiento y el puntaje en comprensión lectora en la ronda 4.....	30
Figura 22. Dispersión entre el tiempo de desplazamiento y el puntaje en comprensión lectora en la ronda 5.....	30
Figura 23. Puntaje estandarizado en matemática en el área rural en la ronda 3.....	60
Figura 24. Puntaje estandarizado en matemática en el área rural en la ronda 4.....	60
Figura 25. Puntaje estandarizado en matemática en el área rural en la ronda 5.....	60

Figura 26. Puntaje estandarizado en comprensión lectora en el área rural en la ronda 4.....	61
Figura 27. Puntaje estandarizado en comprensión lectora en el área rural en la ronda 5.....	61
Figura 28. Dispersión entre el logaritmo del tiempo de desplazamiento y el logaritmo del puntaje bruto de la prueba de matemática en la ronda 3.....	62
Figura 29. Dispersión entre el logaritmo del tiempo de desplazamiento y el logaritmo del puntaje bruto de la prueba de matemática en la ronda 4.....	62
Figura 30. Dispersión entre el logaritmo del tiempo de desplazamiento y el logaritmo del puntaje bruto de la prueba de matemática en la ronda 5.....	62
Figura 31. Dispersión entre el logaritmo del tiempo de desplazamiento y el logaritmo del puntaje bruto de la prueba de PPVT en la ronda 3.....	63
Figura 32. Dispersión entre el logaritmo del tiempo de desplazamiento y el logaritmo del puntaje bruto de la prueba de PPVT en la ronda 4.....	63
Figura 33. Dispersión entre el logaritmo del tiempo de desplazamiento y el logaritmo del puntaje bruto de la prueba de PPVT en la ronda 5.....	63
Figura 34. Dispersión entre el logaritmo del tiempo de desplazamiento y el logaritmo del puntaje bruto de la prueba de comprensión lectora en la ronda 4.....	64
Figura 35. Dispersión entre el logaritmo del tiempo de desplazamiento y el logaritmo del puntaje bruto de la prueba de comprensión lectora en la ronda 5.....	64



1. Introducción

La educación en el área rural del Perú presenta bajos niveles de rendimiento escolar (Ministerio de Educación [MINEDU], ca. 2023a; 2023b). En ese sentido, existen distintas variables que afectan el rendimiento escolar. Una de ellas puede ser el tiempo de desplazamiento al centro educativo, ya que el tiempo es un recurso escaso. Al respecto, en el área rural, hay escolares que se desplazan desde 10 minutos, o menos, hasta dos horas para llegar a su centro educativo, dependiendo del tipo de región, la oferta educativa estatal y la concentración o dispersión de las viviendas rurales (Arregui et al., 2004; Montero et al., 2001; Montero & Uccelli, 2023).

La literatura que aborda el efecto del tiempo de desplazamiento sobre el rendimiento escolar presenta distintos hallazgos. Jacoby et al. (1999), Benavides (2002), Tigre et al. (2017), Afoakwoah y Koomson (2021) encuentran un efecto negativo. Asimismo, Glewwe & Zhao (2010), Iqbal et al. (2020) solo encuentran un efecto negativo en ciertos rangos de distancia. Sin embargo, Thomas (2016) y Contreras et al. (2018) encuentran una asociación positiva cuando usan mínimos cuadrados ordinarios¹. Finalmente, Pov et al. (2021) no encuentran algún efecto.

A partir de la discusión de la literatura, el objetivo de esta tesis es conocer el efecto del tiempo de desplazamiento (ida y vuelta) de la vivienda al centro educativo sobre el rendimiento escolar en el área rural peruana, ya que esta área tiene algunas particularidades. Por un lado, hay un bajo rendimiento escolar (MINEDU, 2023a; 2023b). Por otro lado, hay viviendas que se ubican en asentamientos concentrados o dispersos, aunado al déficit de centros educativos, genera que algunos escolares tengan que emplear un mayor tiempo de desplazamiento al centro educativo (Arregui et al., 2004; Montero et al., 2001; Montero & Uccelli, 2023).

Ante ello, la pregunta de esta investigación es ¿Cuál es el efecto del tiempo de desplazamiento de la vivienda al centro educativo sobre el rendimiento escolar en el área rural del Perú? Ante esta interrogante, la hipótesis plantea que si hay un mayor tiempo de desplazamiento de la vivienda al centro educativo (ida y vuelta), disminuye el rendimiento escolar. Tomando en cuenta las particularidades del área rural mencionadas anteriormente, aquellos escolares que emplean mayor tiempo para

¹ Thomas (2016) encuentra una asociación positiva para el caso de matemática.

trasladarse al centro educativo pueden tener menos tiempo para estudiar en el centro educativo o en su vivienda, lo cual puede perjudicar su rendimiento escolar.

En esa línea, se realiza un estudio econométrico con datos de panel mediante un modelo de Efectos Fijos para responder a la pregunta de investigación. Asimismo, se emplea la base de datos de niños del Milenio, la cual contiene información longitudinal de niños peruanos y de otros países. Dentro de esta base, se usa los datos de los niños de la cohorte menor (niños nacidos entre los años 2001 y 2002); específicamente, las rondas 3, 4 y 5 son las que se utilizan. Además, se usa las pruebas de matemática, comprensión lectora y Peabody Picture Vocabulary Test (PPVT). Esta última, es una prueba de adquisición de vocabulario receptivo (Leon, 2020). Se utilizan estas pruebas ya que matemáticas y lengua son las materias instrumentales que mejor predicen el rendimiento global según Pérez (1981) (citado en Adell, 2006). Asimismo, Alvaro (1990) considera que la nota es la medida más usada por los profesores y centros al momento de valorar el rendimiento de los alumnos (citado en Adell, 2006).

La estructura del presente trabajo presenta el siguiente orden. A continuación, se presenta el marco teórico y la revisión de la literatura empírica. Posteriormente, se evidencian los hechos estilizados. Estas secciones permiten plantear la hipótesis. Después de ello, se expone la metodología, estimaciones, conclusiones y anexos.

2. Marco teórico

Previamente a la introducción de los modelos y ecuaciones, es necesario definir el concepto de rendimiento escolar y tiempo de desplazamiento. Por un lado, según González (1975) “el rendimiento escolar es fruto de una verdadera constelación de factores derivados del sistema educativo, de la familia, del propio alumno en cuanto a persona en evolución” (como se citó en Alvaro et al., 1990, p. 19). Asimismo, Forteza (1975) define al rendimiento como “la productividad del sujeto, el producto final de la aplicación de su esfuerzo, matizado por sus actitudes, rasgos y la percepción más o menos correcta de los cometidos asignados” (como se citó en Alvaro et al., 1990, p. 19).

Respecto al tiempo de desplazamiento, en investigaciones anteriores, se hace referencia al tiempo que se emplea para desplazarse de la vivienda al centro educativo, usualmente. Por una parte, el tiempo de desplazamiento es medido a través de minutos u horas (Afoakwah & Koomson, 2021; Benavides, 2002; Contreras et al., 2018; Jacoby et al., 1999; Tigre et al., 2017). Por otra parte, se usa la distancia desde la vivienda al centro educativo en kilómetros como un proxy del tiempo de desplazamiento (Iqbal et al., 2020; Pov et al., 2021; Thomas, 2016; Zhao & Glewwe, 2010). En el contexto de esta investigación, el tiempo de desplazamiento se define como el tiempo que el escolar emplea para ir y volver de la vivienda al centro educativo, puesto que se prefiere usar el tiempo total que destina el escolar para desplazarse (ida y vuelta), el cual es medido en minutos.

En la presente sección, se presenta un modelo de distribución del tiempo del escolar propuesto por Levin y Tsang (1987) y considerando el aporte de Tigre et al. (2017). Posteriormente, las funciones de producción educativas propuestas por Hanushek (1972, 1979, 1992), y Todd y Wolpin (2007).

2.1. Distribución del tiempo

El tiempo dedicado al estudio es un factor importante en el proceso educativo (Becker, 1965; Hanushek, 1979). En ese sentido, el tiempo dedicado al estudio que se da en el centro educativo como en la vivienda, es relevante para el aprendizaje de

los escolares (Levin & Tsang, 1987)². Asimismo, se debe considerar el tiempo de desplazamiento a un centro educativo como un medio necesario para que el estudiante llegue a dicho centro, y obtenga un aprendizaje (Bratti & Staffolani, 2013). En esa línea, dado que el tiempo es un recurso escaso, Tigre et al. (2017) plantean un modelo teórico donde el escolar solo destina su tiempo a estudiar y a desplazarse al centro educativo como una manera de simplificación de las actividades que realiza el escolar durante el día. Si bien este modelo solo considera dos actividades del estudiante, permite explicar que un mayor tiempo empleado en el desplazamiento al centro educativo, reduce el tiempo usado para el estudio, sea en la vivienda o en el centro educativo (Tigre et al., 2017).

En ese sentido, el mecanismo de transmisión por el cual, el tiempo de desplazamiento afecta al rendimiento escolar se basa en que un mayor tiempo de desplazamiento, reduce el tiempo dedicado al estudio sea en la vivienda y/o al centro educativo, en consecuencia, se afecta el rendimiento escolar. A continuación, se presenta el modelo propuesto por Levin y Tsang (1987) adaptándolo a los fines de esta investigación.

$$T = t_d + t_e \quad (1)$$

$$Q = f(C, S, t_e) \quad (2)$$

Donde Q es el rendimiento del escolar; C es la capacidad del estudiante como las habilidades innatas; S es la cantidad y el tipo de recursos disponibles para el escolar, como las variables del centro educativo, hogar y comunidad; y t_e es el tiempo dedicado al estudio (Levin & Tsang, 1987). Además, t_d es el tiempo de desplazamiento al centro educativo (ida y vuelta); y T es el tiempo total de un día. En ese sentido, como se puede observar en la ecuación (1), el escolar destina todo su tiempo a estudiar y a desplazarse al centro educativo. Asimismo, se observa que el tiempo de estudio es un insumo en la función de producción, el cual tiene un efecto positivo sobre el rendimiento escolar. Sin embargo, si aumenta el tiempo de desplazamiento, reduce el tiempo de estudio.

Este modelo sostiene que un mayor tiempo de desplazamiento reduce el tiempo de estudio del escolar. Sin embargo, no toma en cuenta que algunos padres realizan

² Levin y Tsang (1987) se basan en el aporte de Brown y Saks (1984).

un *trade-off* entre el tiempo de desplazamiento y la calidad del centro educativo al que asiste el estudiante. En el Perú rural, los padres toman una decisión entre tres opciones respecto a la educación de sus hijos escolares cada año (Gertler & Glewwe, 1989). Los padres escogen la opción que más les beneficie entre las siguientes: “(1) send their child to a local public school for the year; (2) send the child away from home to a more distant, but presumably higher quality, school; or (3) do not send the child to school” [(1) enviar al escolar a un centro educativo público local por un año; (2) enviar al escolar a un centro educativo más lejano, el cual es de mejor calidad; y (3) no enviar al niño al centro educativo] (Gertler & Glewwe, 1989, p. 8). Entonces, ciertos padres están dispuestos a que su hijo escolar se desplace por un mayor tiempo al centro educativo, a cambio de obtener un mejor rendimiento escolar, dada la buena calidad de dicho centro educativo. Si bien hay una pérdida en el tiempo de estudio, esta es compensada por una mejor calidad educativa que ofrece el centro educativo lejano. En esa línea, Gertler y Glewwe (1989) y Tigre et al. (2017) coinciden en que los padres evalúan si enviar a su hijo escolar a un centro educativo local o un centro educativo lejano, siendo este último de mejor calidad que el centro educativo local.

Por otra parte, el modelo de Levin y Tsang (1987) no incluye a otros factores de producción como el efecto de los compañeros de clase o el puntaje de un test previo como lo señalan Hanushek (1979) y Todd y Wolpin (2007). En ese sentido, se procede a presentar los aportes de Hanushek (1972, 1979, 1992) y Todd y Wolpin (2007), los cuales son de mayor utilidad para abordar el efecto del tiempo de desplazamiento sobre el rendimiento en el marco de las funciones de producción educativas.

2.2. Funciones de producción educativas

Las funciones de producción educativas son relevantes para enmarcar al rendimiento escolar como un producto de la combinación de insumos a lo largo de un proceso acumulativo y continuo (Hanushek, 1972). Además, resalta que, a nivel conceptual, no hay una forma funcional clara de las funciones de producción (Hanushek, 1972). Sin embargo, los insumos que intervienen principalmente en las funciones de producción son los de la familia, el centro educativo, los compañeros de clase, y las habilidades innatas (Hanushek, 1972, 1979, 1992; Todd & Wolpin, 2007).

En ese sentido, se pueden distinguir distintas especificaciones de las funciones de producción. En primer lugar, se presenta la especificación de insumos contemporáneos, donde solo se considera relevante considerar los insumos actuales que afectan el rendimiento escolar. En segundo lugar, se presenta la especificación de insumos acumulados, el cual considera importante incluir todos los insumos (pasados y contemporáneos). En tercer lugar, se presenta la especificación de valor añadido, donde se considera el puntaje de alguna prueba previa, para solucionar problemas de omisión o falta de datos en periodos pasados. El objetivo de presentar estas especificaciones es tener un panorama amplio sobre los supuestos, limitaciones y ventajas de cada especificación al momento de incluir la variable tiempo de desplazamiento en las funciones de producción.

2.2.1. Especificación de insumos contemporáneos

La especificación de insumos contemporáneos considera que solo son relevantes los insumos actuales para explicar el rendimiento actual. A continuación, se presenta la especificación de los insumos contemporáneos que proponen Todd y Wolpin (2007).

$$P_{ija} = \alpha_1 X_{ija} + e_{ija} \quad (3)$$

Donde,

P_{ija} : Puntaje de una prueba estandarizada del escolar i del hogar j , a la edad de a años.

α_1 : Vector de parámetros.

X_{ija} : Insumos contemporáneos observados del escolar i del hogar j a la edad de a años.

e_{ija} : Término residual que captura cualquier efecto de los insumos omitidos, rezagados (observados y no observados), y la habilidad innata del escolar.

El supuesto necesario para que α_1 sea estimado consistentemente, se basa en que los insumos omitidos, rezagados y la habilidad innata, sean ortogonales a los insumos contemporáneos observados. Este modelo es útil cuando solo se tenga disponibles los datos de insumos contemporáneos y únicamente un puntaje de una

prueba estandarizada actual. Si bien este modelo considera relevante los insumos actuales, la limitación del mismo radica en que no permite capturar los efectos de los insumos educativos en periodos anteriores sobre el rendimiento actual, debido a la ausencia de datos pasados.

A partir de esta especificación, si bien es cierto que el tiempo de desplazamiento no es propiamente un factor de producción, sin embargo, reduce el tiempo dedicado al estudio, el cual sí afecta directamente al rendimiento (Tigre et al., 2017). En ese sentido, en la especificación de insumos contemporáneos, el tiempo de desplazamiento afecta al rendimiento en un periodo específico. Usualmente un cierto año, o a una determinada edad, se toma como un periodo. Sin embargo, si se considera que el escolar asiste al centro educativo en distintos periodos de su vida, entonces, se podría observar un efecto acumulado del tiempo de desplazamiento sobre su rendimiento escolar. En ese sentido, es pertinente presentar la especificación de insumos acumulados, la cual puede brindar un marco de distintos periodos. En esa línea, el tiempo de desplazamiento afecta en más de un periodo al rendimiento escolar, puesto que el escolar pierde tiempo de estudio en cada periodo.

2.2.2. Especificación de insumos acumulados

La especificación de insumos acumulados considera que es importante incluir todos los insumos acumulados (pasados y contemporáneos) provistos por la familia, el centro educativo y otros factores exógenos (Hanushek, 1992). Según Hanushek (1992, p. 89), “si se ignora cualquier aleatoriedad, el rendimiento del alumno i en el tiempo t (P_{it}) resulta”:

$$P_{it} = \Phi\left(F_i^{(t)}, S_i^{(t)}, X_i^{(t)}, I_i\right) \quad (4)$$

Donde,

P_{it} : Puntaje de una prueba estandarizada del alumno i en el tiempo t .

$F_i^{(t)}$: Insumos acumulados de la familia del alumno i hasta el tiempo t .

$S_i^{(t)}$: Insumos acumulados del centro educativo del alumno i hasta el tiempo t .

$X_i^{(t)}$: Insumos acumulados de factores exógenos del alumno i hasta el tiempo t .

I_i : Vector de habilidades innatas del alumno i .

Esta especificación de insumos acumulados es importante porque establece que el rendimiento escolar es producto de la combinación de distintos insumos acumulados en el tiempo (Hanushek, 1972). Cabe mencionar que Hanushek (1972) incluye explícitamente en la función de producción un vector de influencias acumuladas de los compañeros³. Asimismo, Hanushek (1972) indica que cada vector de insumos introducido contribuye al aprendizaje, y la exclusión de alguno de estos puede generar potenciales sesgos en los resultados.

A partir de la especificación de insumos acumulados, se plantea que el tiempo de desplazamiento tiene un efecto acumulado en el rendimiento, ya que los escolares asisten al centro educativo en varios periodos. Entonces, si se considera que por cada periodo (día, mes o año) el escolar pierde un determinado tiempo de estudio, entonces, el efecto acumulado del tiempo de desplazamiento es mayor que el efecto de un solo periodo.

En la aplicación empírica de esta especificación, algunos datos de los insumos pueden estar incompletos. No obstante, en el caso de que existan datos incompletos u omitidos, se puede recurrir a la especificación de valor añadido.

2.2.3. Especificación de valor añadido

Como se ha señalado, un problema común en las aplicaciones empíricas de la especificación de insumos acumulados se debe a la omisión de datos. Ante ello, se plantea la especificación de valor añadido, la cual vincula una medida del rendimiento actual con insumos contemporáneos de la familia y centro educativo, y una medida pasada del rendimiento (Todd y Wolpin, 2007).

$$P_{ija} = X_{ija}\alpha + \gamma P_{ija-1} + e_{ija} \quad (5)$$

Donde,

³ Hanushek (1979) considera el efecto de los insumos acumulados de la familia, colegio, compañeros o pares de clase, y las habilidades innatas en la función de producción. Sin embargo, Hanushek (1992) omite los efectos de los compañeros y el de habilidades innatas, y los sustituye por una variable "X" que captura todo tipo de factores exógenos. Sin embargo, dada la importancia de las habilidades innatas, el autor de la presente investigación opta por mantenerla de forma explícita en la ecuación 4. Asimismo, cabe señalar que los efectos de los pares están incluidos en X.

P_{ija} : Puntaje de una prueba estandarizada del escolar i del hogar j , a la edad de a años.

P_{ija-1} : Puntaje de una prueba estandarizada del escolar i del hogar j , a la edad de $a - 1$ años.

X_{ija} : Insumos observados en el escolar i del hogar j a la edad de a años.

e_{ija} : Residuo que incluye los efectos de los insumos omitidos y pasados (observados y no observados).

La especificación de valor añadido es importante debido a que el puntaje de una prueba pasada (como una línea base) sirve para contrarrestar los datos omitidos o incompletos de insumos pasados (Todd & Wolpin, 2003, 2007). Ante ello, en la especificación, solo se necesita incluir los insumos contemporáneos y el puntaje de una prueba pasada, es decir, en el periodo $t - 1$. Asimismo, Todd y Wolpin (2007) realizan algunos supuestos en su especificación. El primer supuesto consiste en que los coeficientes de los insumos observados, insumos omitidos y las habilidades innatas decrecen geométricamente desde la aplicación del insumo, en otras palabras, el efecto de los insumos pasados sobre el rendimiento actual disminuye conforme transcurren los años⁴. El segundo supuesto se basa en que los insumos omitidos (pasados y contemporáneos) no se correlacionan con la medida de la prueba del periodo anterior ni con los insumos observados contemporáneos.

La importancia de esta especificación radica en que se requiere menos datos que en la especificación de insumos acumulados (Hanushek, 1992). Asimismo, una ventaja de esta especificación de valor añadido con respecto a la especificación de insumos acumulados, se basa en que los efectos de los insumos omitidos son menores debido a que se estaría incluyendo el efecto nivel de un test previo (P_{ija-1}), y sólo se omite los efectos de las habilidades innatas que hayan podido incrementar en el tiempo (Hanushek, 1979).

⁴ Según Todd y Wolpin (2007), "heritable endowments are not observed" [las habilidades innatas son inobservables] (p. 96).

A partir de esta especificación, se observa que el efecto del tiempo de desplazamiento sobre el rendimiento se da en un periodo determinado, similar a la especificación de insumos contemporáneos, ya que, al incluirse una medida de un test previo, la especificación de valor añadido contempla a las variables que afectan entre el periodo del test actual y del test previo. En consecuencia, no se considera a todos los periodos anteriores al test previo (P_{ija-1}), por lo que no se puede observar un efecto del tiempo de desplazamiento en todos los periodos. Mientras que en la especificación de insumos acumulados se observa el efecto del tiempo de desplazamiento en todos los periodos, en la especificación de valor añadido, solo se observa el efecto del tiempo de desplazamiento en uno o algunos periodos.

En esa línea, el desplazamiento debe considerarse en todos los periodos, porque es una actividad necesaria y recurrente que realiza el escolar para asistir al centro educativo (Bratti & Staffolani, 2013). Por ello, es recomendable usar la especificación de insumos acumulados en vez de la especificación insumos de valor añadido. En ese sentido, se puede emplear una especificación de insumos acumulados mediante un modelo de Efectos Fijos o Efectos Aleatorios; en la sección de Metodología se evalúa cuál es mejor.

3. Revisión de literatura empírica

Dentro de la revisión de literatura empírica nacional e internacional, se presenta los hallazgos en Perú, Brasil, Chile, Estados Unidos, Pakistán, Camboya, China y Ghana. En algunas de estas investigaciones, se usa el tiempo de desplazamiento de la vivienda de los estudiantes al centro educativo (medido en minutos u horas), y en otras, se usa la distancia de la vivienda al centro educativo (medida en kilómetros).

En ciertas investigaciones, el rendimiento escolar es medido a través del puntaje de pruebas estandarizadas. Mientras que, en otras, se usa los años de escolarización o el nivel educativo como un proxy del rendimiento escolar, dada la escasa literatura relacionada al tema. Primero, se presentan dos investigaciones en el Perú, y posteriormente, se presenta la literatura internacional. Finalmente, la revisión de la literatura culmina con mencionar las variables que influyen en el rendimiento, las cuales sirven como variables control en las estimaciones.

Jacoby, Cueto y Pollit (1999) realizan un estudio en Huaraz en 1993 y 1994, con datos de 360 escolares de cuarto y quinto grado de primaria. A los escolares se les aplica cuatro pruebas psico-educacionales, y se obtienen datos de sus características individuales, familiares, nivel socioeconómico y del centro educativo. Cabe precisar que el máximo puntaje que se puede obtener en las pruebas es de 22 puntos en matemática y 40 puntos en vocabulario y comprensión lectora, donde cada pregunta equivale a un punto. Los autores usan una regresión mediante un modelo lineal generalizado, y encuentran una significancia negativa, en la que por cada incremento de 30 minutos, caminando de la vivienda al centro educativo, hay una disminución 0.5, 0.84, y 1 punto en las pruebas de matemática, vocabulario y comprensión lectora, respectivamente.

Asimismo, Benavides (2002) realiza una investigación para conocer las variables que influyen en el rendimiento de matemática en Perú. Para ello, usa los datos de la prueba CRECER 1998⁵, y selecciona una muestra de 12 855 alumnos que cursan el cuarto grado de primaria. El autor usa un modelo jerárquico lineal, y usa una variable dummy para el tiempo de desplazamiento, donde se toma el valor de 1, si se camina más de una hora para llegar al centro educativo, y 0 en caso contrario; la cual resulta significativa. Benavides encuentra que aquellos estudiantes que caminan más de una

⁵ CRECER 1998 fue aplicado solo en colegios de áreas urbanas.

hora para llegar al centro educativo obtienen, en promedio, 3.59 puntos porcentuales menos que aquellos que caminan menos de una hora.

Respecto de la literatura internacional, Glewwe y Zhao (2010) realizan una investigación en Gansu, China, y los datos son tomados de una encuesta que contiene información desde el 2000 hasta el 2004. Los escolares tienen entre 9 y 12 años en el año 2000, y de los cuales, solo el 88% de la muestra sigue matriculado en la escuela en el 2004. En esta investigación, se usa un logit ordenado censurado, donde la variable dependiente es si el escolar ha desertado o no, lo cual sirve para observar en qué grado el escolar deserta del centro educativo y ubicar el nivel educativo logrado. En consecuencia, se observa qué variables influyen en el rendimiento. Los autores encuentran que si la distancia a la secundaria (primeros años de secundaria) aumenta en un kilómetro, la probabilidad de seguir matriculado en el siguiente año disminuye en 0.076, siendo este efecto significativo. Sin embargo, no encuentran efectos claros para la primaria ni la secundaria superior (últimos años de secundaria).

Asimismo, Tigre et al. (2017) realizan un estudio en la ciudad de Recife, Brasil, usando datos de una encuesta multinivel del 2013. Esta encuesta es aplicada a estudiantes de sexto grado de primaria de 118 centros educativos públicos. En este estudio, la variable dependiente es el puntaje de una prueba de matemática, donde los puntajes oscilan entre 0 a 100 puntos. Los autores usan Mínimos Cuadrados Ordinarios y variables instrumentales, y encuentran un efecto negativo y significativo del tiempo de desplazamiento sobre el rendimiento en matemática. Otro hallazgo importante es que, ante un aumento de una hora en el tiempo de desplazamiento al centro educativo, el resultado de la prueba de matemática disminuye, en promedio, 0.75 desviaciones estándar.

Asimismo, Iqbal et al. (2020) exploran el efecto de los determinantes económicos sobre el logro educativo de los niños en Pakistán. Por ello, seleccionan 30,513 niños entre 5 y 18 años, y emplean un modelo Probit ordenado censurado por dos razones. La primera, porque se quiere saber el último año de educación logrado por los niños, por ello, es necesario realizar una censura de datos de los niños que actualmente están matriculados; y la segunda, debido a que el nivel de estudios es una serie de elecciones discretas ordenadas. Los autores encuentran que, en las áreas rurales, cuando la distancia de la vivienda al centro educativo está entre 6 a 10 kilómetros, se

encuentra un efecto negativo y significativo de 0.43 sobre la probabilidad de lograr un buen nivel educativo. Sin embargo, cuando la distancia es mayor a 10 kilómetros, no se encuentra un efecto significativo.

Por otra parte, Pov et al. (2021) estudian los determinantes del rendimiento escolar en 517 alumnos de secundaria en el área rural de Battambang, Camboya. En este estudio, las variables dependientes son los resultados en la prueba de matemática y lenguaje, y se utiliza una metodología de un modelo lineal jerárquico de dos niveles. En la investigación no se encuentra significancia en el coeficiente de la variable distancia al centro educativo.

Afoakwoah y Koomson (2021) encuentran que existe un efecto negativo del tiempo de desplazamiento sobre el rendimiento en escolares rurales y urbanos de Ghana. Sin embargo, los escolares rurales se ven más afectados que sus pares urbanos, ya que el efecto del tiempo de viaje sobre el resultado de aprendizaje varía en un rango de 0.3 puntos porcentuales de las habilidades de cálculo a 1.8 puntos porcentuales en las habilidades para escribir en lenguas nativas. Además, los escolares cuyo tiempo de desplazamiento a la escuela es superior al percentil 75 (equivalente a 30 minutos) tienen peores resultados de aprendizaje que los que viajan menos de 30 minutos a la escuela. Asimismo, los que van caminando al centro educativo tienen menor rendimiento que aquellos que van en algún vehículo. Cabe señalar que, en este estudio, el rendimiento es medido por dummies tales como si el niño sabe escribir en inglés o francés, si sabe leer en inglés o francés, si sabe escribir en alguna lengua nativa, si sabe leer en alguna lengua nativa y si puede hacer cálculos. Asimismo, se usa la variable instrumental: tiempo de viaje promedio al centro educativo desde las comunidades de un mismo distrito. Esta variable captura la disponibilidad de centros educativos que son accesibles para los escolares. En ese sentido, el tiempo de viaje al centro educativo requiere que sea instrumentalizado por la variable de disponibilidad de centros educativos.

En contraposición a los efectos negativos de los estudios mencionados, Thomas (2016) usa un método de Mínimos Cuadrados Ordinarios y encuentra que una mayor distancia al centro educativo se asocia positivamente y significativamente con los puntajes en matemática en escolares estadounidenses. No obstante, no encuentra un efecto significativo entre la distancia y los puntajes en la capacidad de lectura ni las

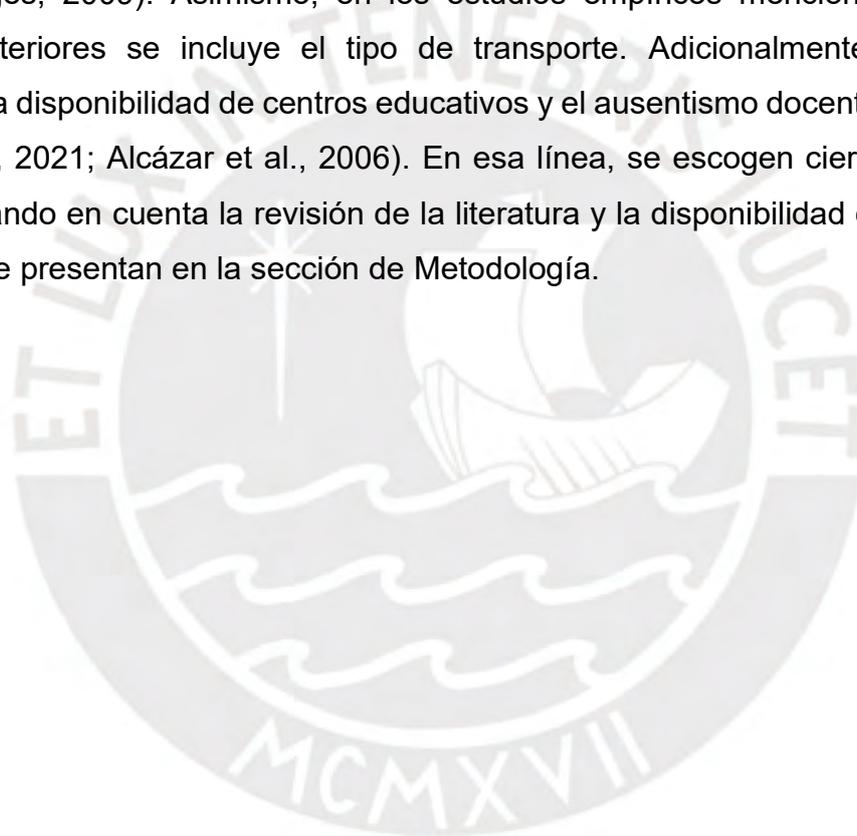
habilidades básicas de vocabulario. Este estudio considera que en Los Ángeles existen vecindarios segregados por el nivel socioeconómico (NSE) y la segregación racial. En los vecindarios más segregados por el bajo NSE y/o racialmente, los escolares no viajan largas distancias, por lo que asisten a centros educativos ubicados en su mismo vecindario, los cuales probablemente sean de baja calidad (Thomas, 2016). En cambio, los escolares que viven en vecindarios menos segregados, se desplazan a centros educativos ubicados fuera de su vecindario a cambio de asistir a un centro educativo de buena calidad. Este hallazgo refuerza el planteamiento teórico de Gertler y Gleewe (1989) y Tigre et al. (2017) respecto a que hay escolares que se desplazan largas distancias a cambio de asistir a un centro educativo de buena calidad.

Asimismo, Contreras et al. (2018) encuentra una asociación positiva y significativa del tiempo de desplazamiento y el rendimiento cuando emplea un método de mínimos cuadrados ordinarios en los datos de escolares chilenos. Cabe mencionar que los autores realizan un promedio entre los puntajes de una prueba de matemática y español. Dicho promedio es tomado como una medida de rendimiento. Los autores encuentran que el efecto del tiempo de desplazamiento se torna negativo y no significativo en la mayoría de estimaciones cuando usan variables instrumentales con efectos fijos a nivel de colegios. Los autores consideran que puede ocurrir un *trade-off* entre un mayor tiempo de desplazamiento a cambio de asistir a un centro educativo de buena calidad. Además, similar que Tigre et al. (2017), la variable instrumental que se emplea es el promedio del tiempo de desplazamiento a los dos colegios más cercanos a la vivienda del escolar, donde no se considera el colegio al que asiste el estudiante.

Tomando en cuenta el balance de todos los aportes de la literatura empírica sobre el tiempo de desplazamiento y el rendimiento escolar, se observa que en algunas investigaciones sí se encuentra un efecto negativo del tiempo de desplazamiento sobre el rendimiento como en Jacoby et al. (1999), Benavides (2002), Tigre et al. (2017), y Afoakwak y Koomson (2021). Asimismo, Gleewe y Zhao (2010), Iqbal (2020) encuentran un efecto negativo solo en ciertos rangos de distancia. Por otro lado, Thomas (2016) y Contreras et al. (2018) encuentran una asociación positiva del tiempo de desplazamiento sobre el rendimiento en matemática; aunque Contreras et

al. (2018) encuentra un efecto negativo y no significativo cuando cambia de método. Finalmente, Pov et al. (2021) no encuentra evidencia de algún efecto.

A partir de la revisión de literatura empírica, no se puede determinar con certeza el efecto del tiempo de desplazamiento sobre el rendimiento escolar, puesto que los hallazgos son diversos. Por otra parte, hay estudios que abordan las distintas variables que afectan al rendimiento escolar. Entre tales variables se encuentran el género, tipo de centro educativo, nivel socioeconómico del hogar, lengua materna del escolar, ubicación del hogar (urbano o rural), nivel educativo de los padres, número de padres e hijos en el hogar, extraedad o repetición de grado, entre otras (Cueto, 2007; Menges, 2009). Asimismo, en los estudios empíricos mencionados en los párrafos anteriores se incluye el tipo de transporte. Adicionalmente, se puede considerar la disponibilidad de centros educativos y el ausentismo docente (Afoakwah & Koomson, 2021; Alcázar et al., 2006). En esa línea, se escogen ciertas variables control tomando en cuenta la revisión de la literatura y la disponibilidad de los datos, las cuales se presentan en la sección de Metodología.



4. Hechos estilizados

A la luz de la revisión de la literatura teórica y empírica, se puede motivar una primera aproximación a la descripción estadística del tiempo de desplazamiento y el puntaje en las pruebas evaluadas. Primero, se presenta la base de datos utilizada en esta investigación. Posteriormente, se presentan estadísticas de las variables de interés.

El estudio longitudinal Niños del Milenio es un proyecto de investigación que aborda la dinámica de la pobreza infantil (Escobal & Flores, 2008). Asimismo, tiene como objetivo comprender mejor las causas y consecuencias de la pobreza infantil, analizar el efecto de las políticas sobre el bienestar infantil y brindar información sobre el desarrollo e implementación de políticas que pueden reducir la pobreza infantil (Escobal & Flores, 2008). Este estudio internacional ha realizado un seguimiento a 12,000 niños en Perú, Etiopía, Vietnam e India en los años 2002, 2006, 2009, 2013, 2016 y 2020-2021 mediante las rondas 1, 2, 3, 4, 5 y 6.

Niños del Milenio tiene dos cohortes de niños para cada país. Las personas que conforman la cohorte menor nacieron entre el 2001 y 2002, mientras que la cohorte mayor está conformada por quienes nacieron entre 1994 y 1995. Asimismo, la cohorte mayor peruana y cohorte menor peruana están compuestas por 714 y 2052 niños en la primera ronda respectivamente⁶. A los niños de la cohorte mayor, se les empieza a aplicar las encuestas desde que tenían entre 7 a 8 años, y a los niños de la cohorte menor, desde que tenían entre 6 a 18 meses (Escobal & Flores, 2008).

En cuanto a la selección de los niños peruanos que son parte del estudio longitudinal, se excluye al 5% de distritos más ricos del país según el mapa de pobreza elaborado en el año 2000 por el Fondo Nacional de Cooperación para el Desarrollo (Escobal & Flores, 2008). Asimismo, “los hogares de Niños del Milenio son muy similares al hogar medio de Perú” (Escobal & Flores, 2008, p. iv). Además, los niños o adolescentes del mencionado proyecto longitudinal mencionado viven áreas rurales y urbanas. Cabe señalar que hay 20 clusters en la base de datos de Niños del Milenio en la ronda 1 del Perú.

En esa línea, hay 553, 509 y 475 niños o adolescentes que viven en el área rural en las rondas 3, 4 y 5. Asimismo, aquellos escolares rurales que asisten a un centro educativo público de primaria o secundaria son 532, 498 y 418, lo cual representa al

⁶ Cabe mencionar que en cada ronda se pierden observaciones.

98.15, 98.61% y 97.44% de los escolares rurales que asisten a algún tipo de centro educativo en la tercera, cuarta y quinta ronda respectivamente⁷.

Referente al tiempo de desplazamiento de ida, es una variable reportada por la cuidadora principal del niño o adolescente; usualmente es la madre. En caso que no sea ella, puede ser algún familiar. La variable captura el tiempo de ida desde la vivienda al centro educativo. A partir de ello, se establece el supuesto de que el tiempo de desplazamiento total (ida y vuelta) es el doble del tiempo de ida al centro educativo. Este supuesto lo establece el autor de esta tesis para los escolares rurales de las rondas 3, 4 y 5 de la cohorte menor; y se explica en el siguiente párrafo. Cabe recordar que en esta sección de Hechos estilizados y en las siguientes, el tiempo de desplazamiento es el tiempo que emplea el escolar para ir y volver de su vivienda al centro educativo.

El supuesto mencionado en el párrafo anterior se basa en que, según la base de datos, la mayoría de escolares rurales camina al centro educativo, y se espera que retornen a sus viviendas del mismo modo. Asimismo, aquellos que usan un vehículo para ir al centro educativo, se espera que regresen con el mismo medio de transporte. Cabe mencionar que el tipo de desplazamiento lo reporta la cuidadora principal en la ronda 3, y lo reporta el niño o adolescente en las siguientes rondas.

Por otra parte, esta base de datos cuenta con información de los puntajes de las evaluaciones mencionadas, variables individuales, familiares, comunidad del niño, etcétera. La prueba de matemática contiene 29 preguntas en la tercera y cuarta ronda, y 31 preguntas en la quinta ronda. Dicha prueba es diferente en cada ronda, aunque algunos ítems son comunes entre rondas y cohortes (Leon, 2020). En cuanto a la prueba de comprensión lectora, contiene 24 y 27 preguntas en las rondas 4 y 5, y solo se aplica en tales rondas. Al igual que en la evaluación de matemática, las pruebas son diferentes en cada ronda; sin embargo, algunos ítems son comunes entre rondas y cohortes (Leon, 2020).

⁷ Los porcentajes se calculan en base al total de alumnos rurales que asisten a un centro educativo público, privado o mixto. Quienes asisten a un centro educativo privado o mixto son 10, 7 y 11 escolares en las rondas 3, 4 y 5, respectivamente. Asimismo, en la quinta ronda, no se incluye a un alumno de un centro educativo público municipal, dado que es el único caso. Además, no se considera a cuatro alumnos de un Colegio de Alto Rendimiento (COAR), ya que los COAR ofrecen un internado para sus alumnos.

Referente al PPVT, “es una prueba (...) de vocabulario receptivo, que consiste en presentar al adolescente una serie de imágenes entre las cuales debe elegir aquella que mejor represente la palabra que se le indica oralmente” (Lavado et al., 2015, p. 68). En ese sentido, el PPVT mide el lenguaje de vocabulario que va adquiriendo el niño a lo largo del tiempo. Esta prueba contiene 125 preguntas en las tres rondas, y se administra individualmente, oralmente, sin tiempo y con referencias a normas (Leon, 2020). En dicha prueba, el escolar empieza respondiendo en un ítem de acuerdo a su edad. Asimismo, no se espera que el escolar sea evaluado en todos los ítems, sino que, si comete ocho o más errores en un conjunto de doce ítems, entonces, se establece un puntaje techo, donde cada ítem por debajo de dicho límite vale un punto, y los ítems por encima de dicho techo, valen cero puntos (Leon, 2020). Esta prueba es comparable en el tiempo, mientras que las pruebas de matemática y comprensión lectora no lo son. En consecuencia, esta prueba no requiere que sea estandarizada, a diferencia de las otras dos pruebas. En la sección de Metodología se aborda la estandarización de pruebas.

A partir de lo mencionado, se presentan figuras que contienen gráficos estadísticos. En cuanto a las figuras del tiempo y tipo de desplazamiento se consideran los datos de los escolares rurales que asisten a un centro educativo público. Mientras que, para las figuras del puntaje de las pruebas y de dispersión, se considera a aquellos escolares rurales que asisten a un centro educativo público y que cumplen con ciertos criterios que son abordados posteriormente.

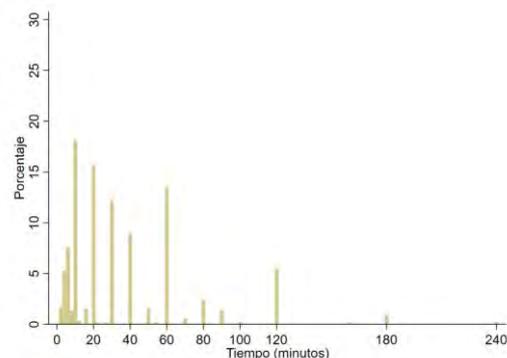
Cabe resaltar que la descripción y análisis estadístico, matrices de correlación, estimaciones, etc. presentadas mediante figuras y tablas son elaboradas por el autor de esta tesis a partir de la base de datos del estudio longitudinal Niños del Milenio. Esta base de datos es administrada por Boyden (2018), Woldehanna et al. (2018) y Sanchez et al. (2018) en las rondas 3, 4 y 5 del mencionado estudio longitudinal, respectivamente.

4.1. Tiempo de desplazamiento

A continuación, se presentan los histogramas del tiempo de desplazamiento de los escolares rurales que asisten a un centro educativo público. Respecto a dichos alumnos, los que tienen registrado un tiempo de desplazamiento son 512, 492 y 404

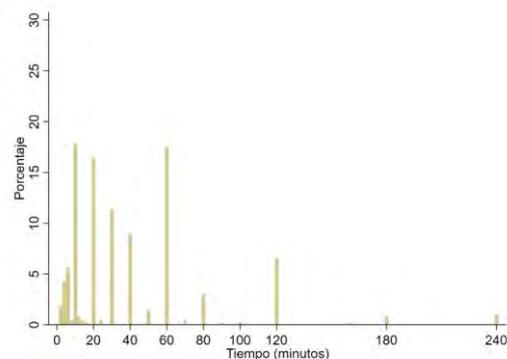
en las rondas 3, 4 y 5 respectivamente. Se observa una distribución asimétrica, donde hay una mayor frecuencia cuando se emplean 10, 20 y 60 minutos en las tres rondas.

Figura 1
Tiempo de desplazamiento en el área rural en la ronda 3



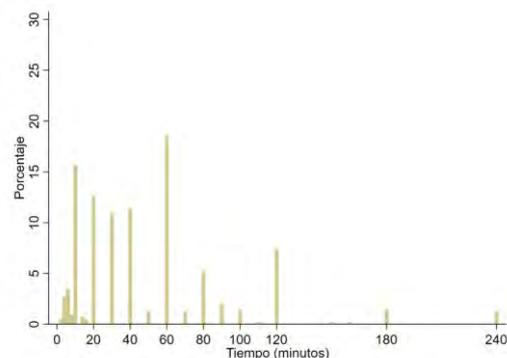
Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 2
Tiempo de desplazamiento en el área rural en la ronda 4



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 3
Tiempo de desplazamiento en el área rural en la ronda 5



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

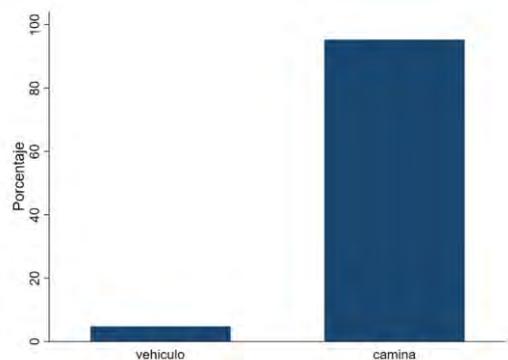
El promedio del tiempo de desplazamiento resulta 35, 39 y 48 minutos en la tercera, cuarta y quinta ronda respectivamente, lo cual es coherente por la mayor oferta de centros educativos que cuentan con el nivel de primaria en comparación con los que cuentan con el nivel de secundaria (MINEDU, ca. 2016). Al haber menor cantidad de centros educativos que ofrezcan este último nivel, los escolares tendrán que recorrer mayores distancias para asistir a uno.

Con respecto a la moda del tiempo de desplazamiento, resulta 10 minutos en la tercera y cuarta ronda, y una hora en la quinta ronda. Asimismo, las medianas son 20, 30 y 40 minutos en la tercera, cuarta y quinta ronda respectivamente. Además, hay 24.8%, 29.88% y 39.36% de escolares que se desplazan una hora o más al centro educativo (ida y vuelta) en la tercera, cuarta y quinta ronda respectivamente. Referente al tiempo mínimo y máximo de desplazamiento, resulta dos minutos y cuatro horas en las tres rondas. Cabe mencionar que si se emplea la misma base de datos con escolares urbanos que asisten a centros educativos públicos, privados o mixtos, los promedios del tiempo de desplazamiento son 24, 26 y 29 minutos en las rondas 3, 4 y 5 respectivamente. A partir de ello, se desprende que el promedio del tiempo de desplazamiento en el área rural es mayor que el del área urbana en las tres rondas.

A todo esto, es relevante considerar el tipo de desplazamiento al centro educativo. En las siguientes figuras se observa que la mayoría de alumnos se desplaza caminando en las rondas 3 y 4. Mientras que en la ronda 5, aumenta la cantidad de alumnos que usan algún tipo de vehículo.

Figura 4

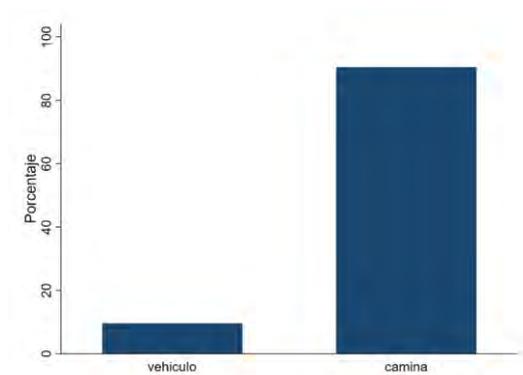
Tipo de desplazamiento en la ronda 3



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 5

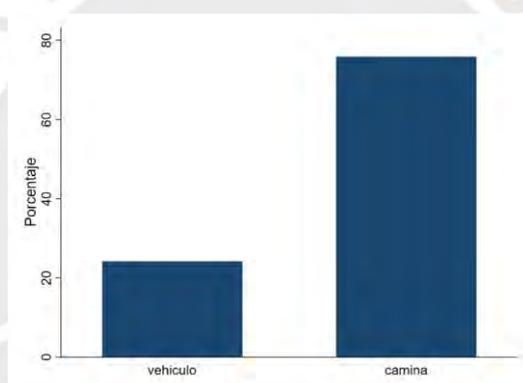
Tipo de desplazamiento en la ronda 4



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 6

Tipo de desplazamiento en la ronda 5



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

En la tercera, cuarta y quinta ronda hay 504, 444 y 314 alumnos que caminan al centro educativo público, respectivamente. Asimismo, quienes usan algún vehículo son 25, 47 y 100 escolares. En efecto, se reduce la cantidad de los alumnos que caminan entre la cuarta y quinta ronda. En esta última ronda se requiere usar algún tipo de vehículo para asistir a un centro educativo con secundaria, ya que implica un mayor desplazamiento para ciertos escolares.

4.2. Puntaje en pruebas escolares

En esta subsección se abordan la descripción estadística de las evaluaciones. Para ello, se considera los datos de aquellos escolares rurales que reúnan cuatro criterios simultáneamente. Aquellos datos que no los cumplen, no se consideran en la

descripción y análisis estadístico ni en las estimaciones que se presentan en la sexta y séptima sección. En ese sentido, los escolares seleccionados cumplen con los cuatro criterios en las rondas 3, 4 y 5 para las pruebas de matemática y PPVT. En el caso de comprensión lectora, solo se requiere que se cumpla con los criterios en las rondas 4 y 5.

El primer criterio se basa en que sean escolares rurales que asistan a un centro educativo público en las rondas mencionadas, ya que casi todos los alumnos rurales estudian en dicho tipo de centro educativo. El segundo criterio se basa en que los escolares sean evaluados en las pruebas de matemática, PPVT y comprensión lectora. El tercer criterio descansa en que los escolares deben tener información registrada en la variable tiempo de desplazamiento, las variables del puntaje de las pruebas y en todas las variables control simultáneamente en las tres rondas. Las variables control se detallan en la subsección 6.1. El cuarto criterio se basa en que los escolares sean los mismos en las tres rondas. A partir del cumplimiento de estos cuatro criterios, la cantidad de observaciones es la misma en cada ronda.

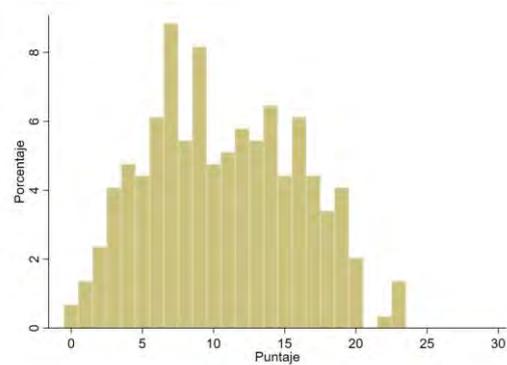
En esa línea, la cantidad de niños o adolescentes rurales obtenida para analizar el efecto del tiempo de desplazamiento sobre el rendimiento escolar son 294, 280 y 324 en cada ronda, lo cual implica tener 882, 840 y 648 observaciones en las evaluaciones de matemática, PPVT y comprensión lectora respectivamente. A continuación, se presentan figuras de las pruebas de Matemática y PPVT en las rondas 3, 4 y 5. En comprensión lectora, se usa solo la ronda 4 y 5.

4.2.1. Puntaje en matemática

La prueba de matemática tiene diversas preguntas en las tres rondas. En la ronda 3, se incluyen preguntas como identificación de números y operaciones básicas. En la ronda 4, se incluyen el tipo de preguntas de la ronda 3, y adicionalmente, operaciones básicas con números enteros y decimales, operaciones combinadas, uso de fracciones. Además, se requiere leer enunciados u oraciones para resolver los ejercicios. En la ronda 5, a parte del tipo de preguntas de la ronda 4, se incluyen preguntas como operaciones con porcentajes, regla de tres simple, razonamiento matemático, estadística, uso de promedios, fracciones en ecuaciones, entre otras. Cabe mencionar que hay algunas preguntas iguales en las tres rondas. Asimismo, los puntajes mínimos y máximos son 0 y 23 puntos en la tercera ronda; 0 y 27 puntos en

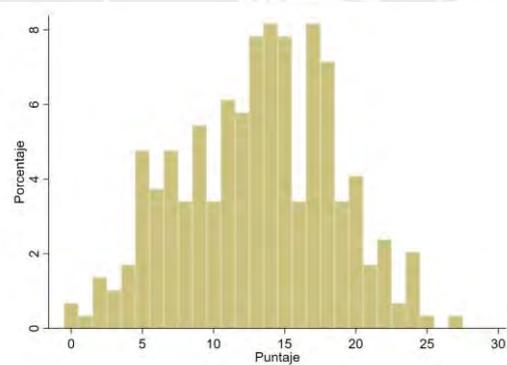
la cuarta; y 0 y 24 en la quinta. En las siguientes figuras se puede observar el puntaje bruto de matemática por rondas. Cabe mencionar que en el anexo A, se presentan los histogramas de los puntajes estandarizados de los escolares rurales en las pruebas.

Figura 7
Puntaje en matemática en la ronda 3



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

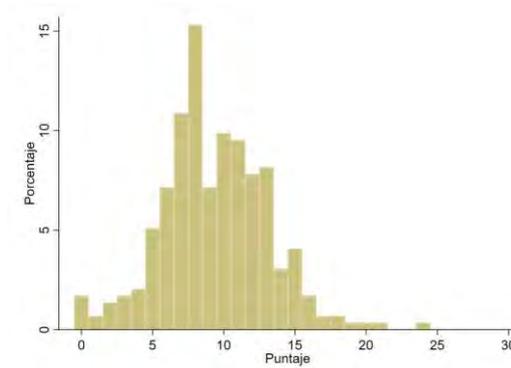
Figura 8
Puntaje en matemática en la ronda 4



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 9

Puntaje en matemática en la ronda 5



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

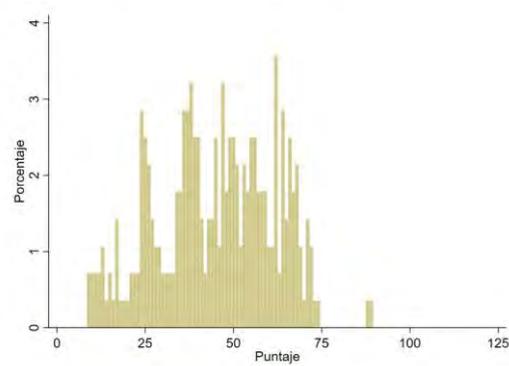
En primer lugar, respecto a matemática, se observa que los promedios son 10.67, 13.11 y 9.35 en las rondas 3, 4 y 5 respectivamente. Asimismo, las medianas son 10, 13 y 9 puntos en las rondas 3, 4 y 5 respectivamente. En ese sentido, se puede observar que el rendimiento en matemática durante la primaria, incrementa entre la ronda 3 y 4. Sin embargo, cuando la gran mayoría de escolares está en la secundaria, el rendimiento cae. Una posible explicación se debe a que en la ronda 3 y 4, las preguntas son similares, como la identificación de números, operaciones básicas y/o combinadas, entre otras. Sin embargo, en la ronda 5, aparte de realizar el tipo de preguntas de las rondas previas, se incluyen otro tipo de preguntas como reglas de tres simple, porcentajes, interpretación de gráficos estadísticos, resolución de fracciones en ecuaciones, comprensión de un texto y a partir de ahí, elaborar operaciones combinadas, razonamiento lógico, etcétera.

4.2.2. Puntaje en PPVT

La prueba de PPVT es la misma en las tres rondas. Asimismo, esta prueba mide el lenguaje de vocabulario que va adquiriendo el niño a lo largo del tiempo como se ha mencionado anteriormente. Cabe mencionar que los puntajes mínimos y máximos son 9 y 89 puntos en la tercera ronda; 2 y 109 puntos en la cuarta; y 2 y 125 puntos en la quinta. En las siguientes figuras se puede observar los puntajes a lo largo de las rondas.

Figura 10

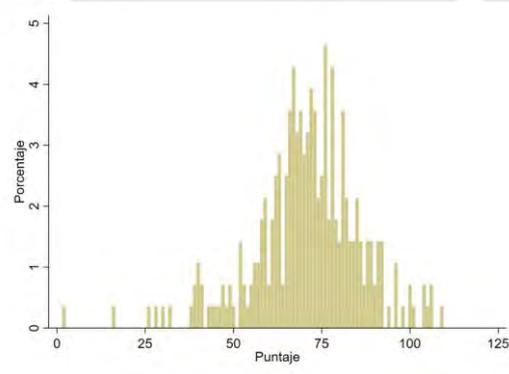
Puntaje en PPVT en la ronda 3



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 11

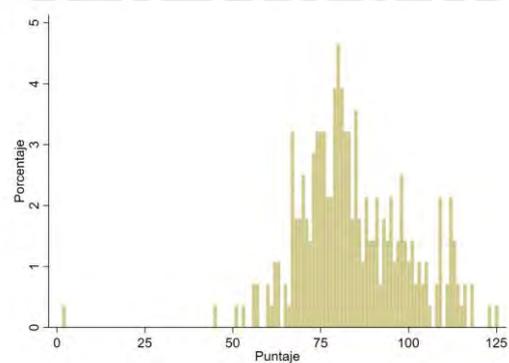
Puntaje en PPVT en la ronda 4



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 12

Puntaje en PPVT en la ronda 5



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

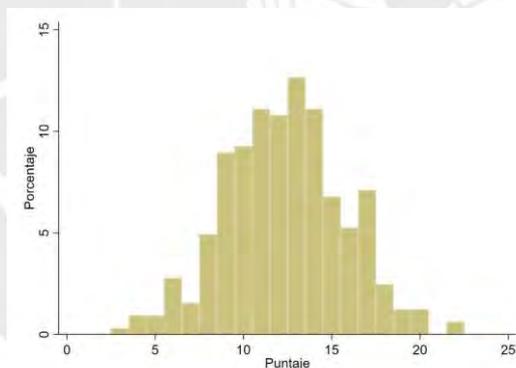
En segundo lugar, en PPVT, los promedios resultan 45.03, 70.86 y 84.28 en las rondas 3, 4 y 5. Además, las medianas son 46.5, 72 y 82 en la tercera, cuarta y quinta ronda respectivamente. A comparación de los resultados en matemática, los escolares han incrementado los promedios de sus puntajes en PPVT en el transcurso de las rondas, lo cual tiene sustento en que ellos van adquiriendo un mayor vocabulario conforme aumenta su edad.

4.2.3. Puntaje en comprensión lectora

En cuanto a la prueba de comprensión lectora, se indican palabras y oraciones donde se debe marcar una opción que contenga la imagen correspondiente en la cuarta y quinta ronda. Asimismo, hay preguntas de comprensión a partir de un anuncio, noticia o textos. Cabe señalar que los puntajes mínimos y máximos son 3 y 22 puntos en la cuarta ronda, mientras que son 2 y 25 puntos en la quinta. Seguidamente, se presentan los histogramas correspondientes a esta prueba.

Figura 13

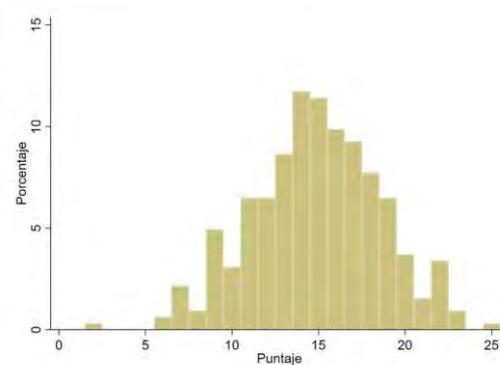
Puntaje en comprensión lectora en la ronda 4



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 14

Puntaje en comprensión lectora en la ronda 5



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

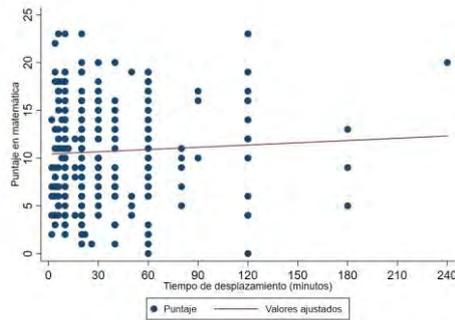
Los promedios resultan 12.34 y 14.87 en las rondas 4 y 5 respectivamente. Asimismo, las medianas obtenidas son 12 y 15 en dichas rondas. Similarmente al caso de PPVT, los puntajes se han incrementado en el tiempo. Una posible explicación se debe a que la dificultad es similar entre las pruebas de la ronda 4 y 5 en dicha materia. En la ronda 4, se le pide al escolar que lea, comprenda y responda preguntas acerca de un anuncio, una noticia y un texto breve. Asimismo, en la ronda 5, se les pide marcar opciones que contienen imágenes que representan el enunciado que leen, y, también, se les pide leer y comprender cuatro textos breves y resolver las preguntas sobre los textos, donde uno de ellos, es el mismo que el de la ronda 4. En ese sentido, se espera que los escolares puedan tener un mejor puntaje en la ronda 5 con respecto a la ronda previa, tomando en cuenta que los niveles de dificultad no varían mucho.

4.3. Gráficos de dispersión

En esta subsección se presentan los gráficos de dispersión entre el tiempo de desplazamiento y los puntajes de las pruebas en las rondas 3, 4 y 5 de los escolares rurales que cumplen con los cuatro criterios mencionados anteriormente. Cabe señalar que en el anexo B, se presentan los gráficos de dispersión de dichos escolares mediante logaritmos en el tiempo de desplazamiento y en los puntajes de las pruebas. A continuación, se presentan los gráficos de dispersión de matemática y PPVT en las rondas 3, 4 y 5. En el caso de comprensión lectora, se presentan los gráficos de dispersión en las rondas 4 y 5.

Figura 15

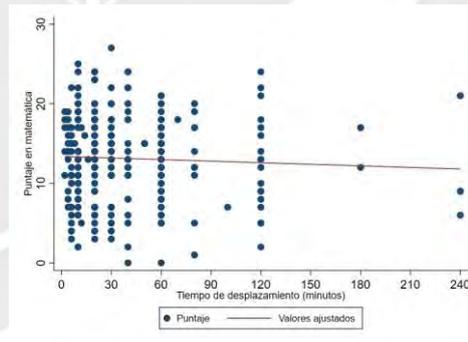
Dispersión entre el tiempo de desplazamiento y el puntaje en matemática en la ronda 3



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 16

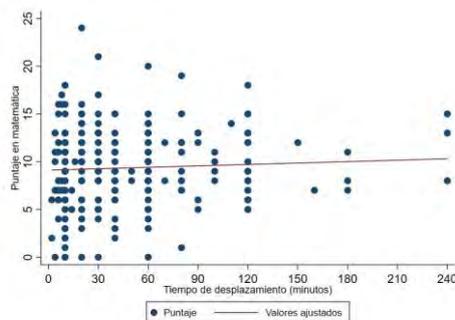
Dispersión entre el tiempo de desplazamiento y el puntaje en matemática en la ronda 4



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 17

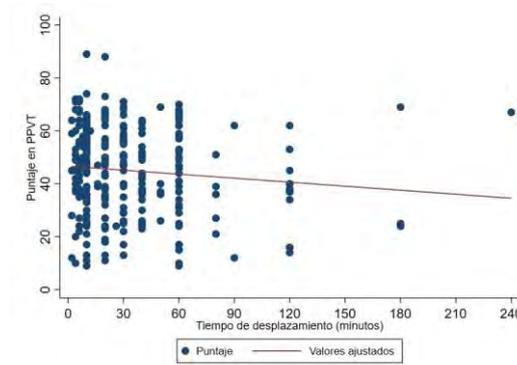
Dispersión entre el tiempo de desplazamiento y el puntaje en matemática en la ronda 5



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 18

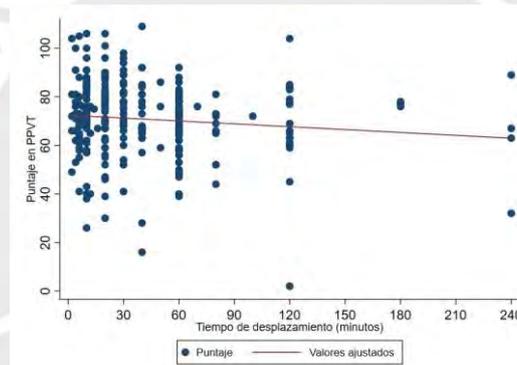
Dispersión entre el tiempo de desplazamiento y el puntaje en PPVT en la ronda 3



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 19

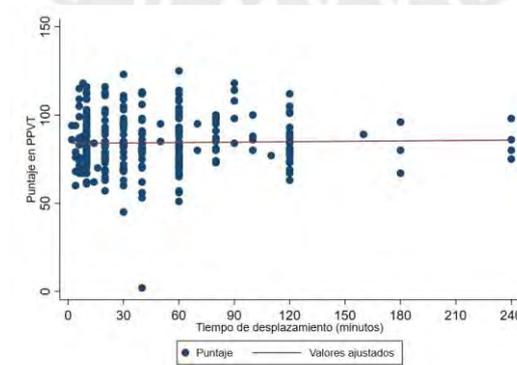
Dispersión entre el tiempo de desplazamiento y el puntaje en PPVT en la ronda 4



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 20

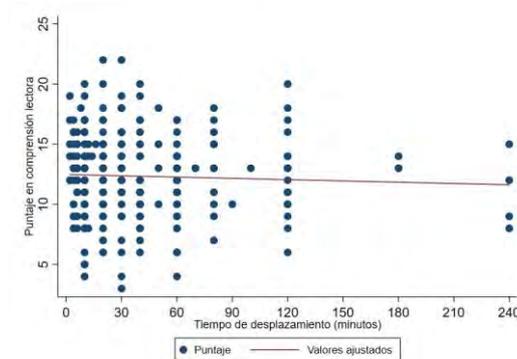
Dispersión entre el tiempo de desplazamiento y el puntaje en PPVT en la ronda 5



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 21

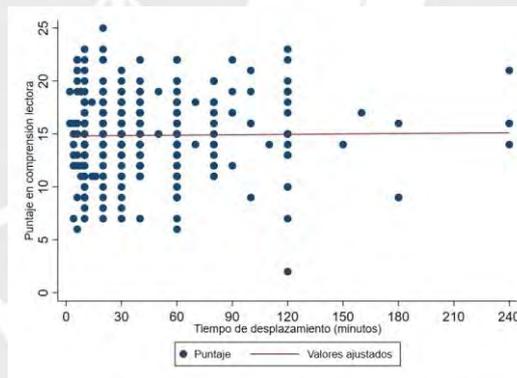
Dispersión entre el tiempo de desplazamiento y el puntaje en comprensión lectora en la ronda 4



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 22

Dispersión entre el tiempo de desplazamiento y el puntaje en comprensión lectora en la ronda 5



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

En matemática, se tienen correlaciones de 0.05, -0.05 y 0.06 en las rondas 3, 4 y 5 respectivamente. Mientras que para PPVT, son de -0.1, -0.11 y 0.02 en la tercera, cuarta y quinta ronda respectivamente. Asimismo, en comprensión lectora, las pendientes son -0.04 y 0.01 en las rondas 4 y 5. En ese sentido, en matemática y comprensión lectora, aparentemente no hay una relación inversa entre el tiempo de desplazamiento y el puntaje en dichas pruebas. Sin embargo, en PPVT, se puede observar una asociación negativa en las rondas 3 y 4.

5. Hipótesis

La presente tesis tiene como objetivo conocer cuál es el efecto del tiempo de desplazamiento de la vivienda al centro educativo, ida y vuelta, sobre el rendimiento escolar en matemática, PPVT y comprensión lectora en el área rural del Perú.

En ese sentido, la hipótesis de esta tesis plantea que, si aumenta el tiempo de desplazamiento de la vivienda al centro educativo (ida y vuelta), disminuye el rendimiento escolar. Se espera esta relación, ya que un mayor tiempo empleado en el desplazamiento de la vivienda al centro educativo (ida y vuelta), implica tener menos tiempo dedicado al estudio, en consecuencia, puede afectarse el rendimiento escolar.



6. Metodología

La presente sección consta de las siguientes partes. En la primera, se describen las variables a analizar. Posteriormente, se construye la descripción estadística de las principales variables de análisis. Luego, se presenta un análisis estadístico. Finalmente, se describe el modelo a ser usado, en base a la revisión de literatura teórica.

6.1. Variables a analizar

A continuación, se presentan las principales variables a considerarse en las estimaciones. Asimismo, después de la tabla 1, se aborda cómo estandarizar las pruebas de matemática y comprensión lectora.

Tabla 1

Variables a analizar

Variable	Nombre de la variable	Descripción
Prueba estandarizada de matemática	std_mate	Variable que captura el puntaje estandarizado de la prueba de matemática. Luego de la presentación de esta tabla, se aborda cómo estandarizar el puntaje bruto de la prueba.
Prueba estandarizada de comprensión lectora	std_cmpr	Variable que captura el puntaje estandarizado de la prueba de comprensión lectora.
Prueba de vocabulario en imágenes Peabody (PPVT)	ppvt_raw	Variable que captura el puntaje bruto en PPVT. A diferencia de la prueba de matemática y comprensión lectora, PPVT es una prueba comparable en el tiempo, por lo que no se necesita estandarizar el puntaje.
Tiempo de desplazamiento (minutos)	desplazamiento	Variable continua que captura la cantidad de minutos que el escolar se desplaza de su vivienda al centro educativo (ida y vuelta).
Camina	camina	Variable dummy que toma el valor de 0 si el escolar camina al centro educativo; 1, si usa un vehículo o bicicleta.
Grado	grado	Variable que captura el grado educativo en el que se encuentra el escolar. El grado 1 equivale al primer grado, y así sucesivamente hasta el grado 11 que equivale a quinto de secundaria. La categoría base es primer grado.

Índice de riqueza	riqueza	Variable continua que toma valores entre 0 a 1, donde 1 se considera el máximo Nivel socioeconómico (NSE). Para hallar este índice, se toma el promedio de los sub índices de la calidad de la vivienda, acceso a servicios y consumo de bienes durables.
Presencia del centro educativo	prsc_col	Es una dummy que toma el valor de 1 si se cumplen dos escenarios: si hay algún centro educativo que se ubique en la misma comunidad del escolar, y si dicho centro enseña el grado o año que le corresponde al escolar. En el caso de que no se cumpla al menos uno de los escenarios, la dummy toma el valor de 0. Por ejemplo, si un escolar cursa el tercer año de secundaria, y en su comunidad, hay un centro educativo que ofrezca el nivel de secundaria, la dummy toma el valor de 1. En caso que no haya un centro educativo que enseñe dicho nivel, la dummy toma el valor de 0. La definición de “comunidad” es ad-hoc al estudio Niños del Milenio. Se define como comunidad al área donde hay al menos un centro educativo y un puesto de salud al que acuden las familias de 2 escolares o más. Sin embargo, puede ocurrir que no necesariamente haya centros educativos en la comunidad del escolar. Se podría usar como un instrumento.

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

En cuanto a la estandarización de pruebas, para el caso de la prueba de matemática y comprensión lectora, dado que son diferentes pruebas en el tiempo, se emplea la estandarización. Este método es conocido como *linear equating*, el cual consiste en comparar a grupos de escolares que rinden una prueba mediante la media y desviación estándar. El procedimiento consiste en que se resta la media del grupo seleccionado a la nota del escolar y, posteriormente, el resultado es dividido entre la desviación estándar (Ryan & Brockmann, 2018). Este procedimiento es replicado en cada ronda. A continuación, se presenta la fórmula de la estandarización.

$$PE_{it} = \frac{P_{it} - \bar{P}_t}{\sigma_t}$$

Donde,

PE_{it} : Prueba estandarizada de un niño i en un periodo t .

P_{it} : Puntaje de una prueba del niño i en el periodo t .

\bar{P}_t : Media de todos los puntajes de las pruebas de los escolares rurales en el periodo t.

σ_t : Desviación estándar de todos los puntajes de las pruebas de los escolares rurales en el periodo t.

Las pruebas de matemática y comprensión lectora se estandarizan, debido a que no son pruebas comparables en el tiempo; sin embargo, algunos ítems de dichas pruebas son comunes para las rondas y las cohortes de edad de Niños del Milenio, lo cual sirve para equiparar puntajes (Leon, 2020). En cambio, el PPVT, es la misma prueba en cada ronda, lo cual permite que sea una prueba comparable en el tiempo. Por ello, se toma el puntaje bruto en dicha prueba.

6.2. Descripción estadística de las principales variables de análisis

Como se ha mencionado en los párrafos iniciales de la sección de Metodología, se selecciona a los escolares rurales de centros educativos públicos. A continuación, se presentan las variables de aquellos escolares rurales de centros educativos públicos que cumplen con los cuatro criterios mencionados en la sección de Hechos estilizados.

Para la presentación de la tabla 2, se considera los datos de los escolares que han sido seleccionados mediante los criterios mencionados anteriormente y tomando en cuenta a aquellos escolares que tienen registrada una nota en la prueba de matemática. Se usa dicha prueba, ya que tiene mayor cantidad de observaciones que las demás pruebas. Si bien en las estimaciones de matemática se consideran los datos de 294 escolares, se consideran los datos de 280 y 324 escolares para las pruebas de PPVT y comprensión lectora. En las siguientes tablas, se presenta la información de dichas variables. Posteriormente, se presentan variables de lengua nativa y educación de los padres.

Tabla 2

Estadística de las variables

Variable	Ronda 3		Ronda 4		Ronda 5	
	Obs.	Media	Obs.	Media	Obs.	Media
Género (1=hombre)	294	0.49	294	0.49	294	0.49
Puntaje en matemática	294	10.67	294	13.11	294	9.35
Puntaje estandarizado en matemática	294	0	294	0	294	0
Puntaje en PPVT	284	45.01	294	70.85	289	84.13
Puntaje en Comprensión lectora	293	5.44	294	12.39	288	14.91
Puntaje estandarizado en comprensión lectora			288	0.03	288	0.01
Desplazamiento (minutos)	294	33	294	40	294	47
Camina (1=sí camina)	294	0.96	294	0.9	294	0.78
Grado	294	2.14	294	5.7	294	8.59
Índice de riqueza	294	0.34	294	0.4	294	0.44
Número de hermanos	294	2.74	294	2.46	294	2.29
Presencia de centro educativo público (1=sí hay)	294	0.89	294	0.84	294	0.57

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

En primer lugar, hay variaciones en el rendimiento por cada prueba. Por un lado, en matemática, aumenta el promedio del puntaje bruto entre la ronda 3 a la ronda 4; sin embargo, luego cae de la ronda 4 a la ronda 5. Por otro lado, en PPVT y comprensión lectora, los promedios han aumentado en el transcurso de las rondas. Cabe mencionar que las posibles explicaciones sobre las variaciones en los promedios de cada prueba se han señalado en la sección de Hechos estilizados. Finalmente, cabe mencionar que el puntaje estandarizado de las pruebas en las rondas es 0, ya que cuando se hace la estandarización, la media obtenida es 0 en cada prueba y en las tres rondas.

En segundo lugar, el promedio del tiempo de desplazamiento aumenta 14 minutos entre la ronda 3 a la ronda 5. Este aumento tiene coherencia con la oferta de centros educativos públicos de primaria y secundaria en las rondas 3 y 5. En estas rondas, el 90 % y 57% de escolares tienen un centro educativo en su comunidad. Este descenso porcentual se debe a que todos los escolares cursan la primaria en la ronda 3, y la

gran mayoría cursa la secundaria en la ronda 5. En ese sentido, en algunos casos, aquellos escolares que transitan de la primaria a la secundaria deben desplazarse por un mayor tiempo a un centro educativo.

En tercer lugar, el índice de riqueza aumenta a lo largo de las rondas. El promedio del índice de riqueza de los hogares de los escolares resulta 0.34, 0.4 y 0.44 en las rondas 3, 4 y 5 respectivamente. En cuarto lugar, el número de hermanos que viven en el hogar se reduce ligeramente entre la ronda 3 a la ronda 5, lo cual puede deberse a que los hermanos mayores pueden mudarse por motivos de trabajo, estudio, independencia de la familia o algún otro motivo conforme transcurre el tiempo.

En quinto lugar, los escolares rurales cursan el segundo grado en la ronda 3 principalmente. Tomando en cuenta que la ronda 3, 4 y 5 se llevan a cabo en los años 2009, 2013 y 2016, se espera que los escolares avancen un grado por cada año. En el anexo C, se muestran los grados que cursan los escolares en cada ronda con mayor detalle. Los estudiantes se encuentran en quinto grado (20.75%), sexto grado (47.96%), y primero de secundaria (18.37%) en la ronda 4, y los demás escolares cursan grados inferiores al quinto grado, excepto uno. Además, los alumnos cursan el segundo año de secundaria (22.79%), el tercer año (45.24%) y cuarto año (16.67%) en la ronda 5. A partir de estos datos, se observa que hay algunos escolares que pueden repetir de grado o desertar del sistema educativo.

Por otra parte, es importante considerar a las variables asociadas al escolar como la lengua materna del escolar y la educación de sus padres. Por ello, se usan los datos de la ronda 3, ya que dicha ronda contiene mayor cantidad de observaciones que las rondas 4 y 5. La lengua materna es la primera lengua que los escolares aprendieron en el hogar. Si bien esta variable es invariante en el tiempo se puede considerar como un proxy de la condición étnica del niño o adolescente. La condición de indígena coloca en una situación de desventaja frente aquel que no es indígena en el Perú (Cueto et al., 2012).

Asimismo, la educación de los padres, especialmente la educación de la madre, es una variable que según la revisión de la literatura influye en el rendimiento escolar. No obstante, así como la lengua materna es una variable que no cambia en el tiempo, la educación de los padres no varía en el tiempo o cambia muy poco, más aún en un

contexto rural. En consecuencia, la educación de la madre no es incluida en las estimaciones de Efectos fijos, puesto que si se incluye genera sesgo.

Si bien ambas variables no se incluyen en las estimaciones, es necesario tener una aproximación a dichas variables. A continuación, se presenta la tabla 3 que las presenta.

Tabla 3
Variables asociadas al escolar

Variable	Obs.	Porcentaje
Lengua materna	292	100%
Castellano	165	56.51%
Nativa	127	40.63%
Educación de la madre	268	100%
Sin educación	61	22.76%
Primaria incompleta	117	43.66%
Primaria completa	42	15.67%
Secundaria incompleta	29	10.82%
Secundaria completa	19	7.09%
Educación del padre	206	100%
Sin educación	7	3.40%
Primaria incompleta	83	40.29%
Primaria completa	54	26.21%
Secundaria incompleta	30	14.56%
Secundaria completa	32	15.53%

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

A partir de la información de la tabla 3 se observa que los escolares que poseen una lengua nativa representan al 40.63%. Sin embargo, no es la lengua que hablan en las rondas 3, 4 y 5, necesariamente, ya que el escolar puede comunicarse en castellano en las interacciones en el centro educativo, la vivienda, o en otros entornos.

En cuanto a la educación de los padres, la mayoría de madres y padres posee primaria incompleta. Las madres y padres que poseen primaria incompleta representan el 43.66% y 40.29%. Asimismo, los padres que poseen el nivel de secundaria completa representan al 15.53%. Mientras que las madres que alcanzan dicho nivel, representan al 7.09% del total de madres de la muestra.

6.3. Análisis estadístico

A continuación, se presenta las matrices de correlación con la finalidad de conocer cómo se asocian los puntajes de las pruebas con la variable de interés y las variables control en cada ronda.

Tabla 4

Matriz de correlación de las variables en la ronda 3

Variables	Puntaje en matemática	Puntaje en PPVT	Desplazamiento	Riqueza	Hermanos	Grado
Puntaje en matemática	1.00					
Puntaje en PPVT	0.60	1.00				
Desplazamiento	0.05	-0.11	1.00			
Riqueza	0.25	0.32	-0.13	1.00		
Hermanos	-0.27	-0.18	-0.07	-0.06	1.00	
Grado	0.45	0.29	0.02	0.26	0.00	1.00

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Tabla 5

Matriz de correlación de las variables en la ronda 4

Variables	Puntaje en matemática	Puntaje en PPVT	Puntaje en CL	Desplazamiento	Riqueza	Hermanos	Grado
Puntaje en matemática	1.00						
Puntaje en PPVT	0.49	1.00					
Puntaje en comprensión lectora (CL)	0.51	0.51	1.00				
Desplazamiento	-0.05	-0.12	-0.06	1.00			
Riqueza	0.20	0.33	0.21	-0.07	1.00		
Hermanos	-0.14	-0.14	-0.13	0.04	-0.08	1.00	
Grado	0.47	0.41	0.39	0.01	0.27	0.02	1.00

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Tabla 6

Matriz de correlación de las variables en la ronda 5

Variables	Puntaje en matemática	Puntaje en PPVT	Puntaje en CL	Desplazamiento	Riqueza	Hermanos	Grado
Puntaje en matemática	1.00						
Puntaje en PPVT	0.45	1.00					
Puntaje en CL	0.53	0.56	1.00				
Desplazamiento	0.06	0.03	0.03	1.00			
Riqueza	0.17	0.28	0.21	-0.18	1.00		
Hermanos	-0.05	-0.17	-0.16	0.06	-0.04	1.00	
Grado	0.33	0.38	0.50	0.04	0.25	0.05	1.00

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Como se observa en las tablas 4, 5 y 6, la variable grado se asocia positivamente con los puntajes de las pruebas y con el tiempo de desplazamiento en cada ronda, excepto en la ronda 5. Asimismo, el índice de riqueza se asocia positivamente con los puntajes de las pruebas en cada ronda, y negativamente con el tiempo de desplazamiento. Por otra parte, el número de hermanos se asocia negativamente con el puntaje de las pruebas en cada ronda, y positivamente con el tiempo de desplazamiento en las rondas 4 y 5.

6.4. Modelo de estimación

Se pueden mencionar distintos modelos de estimación para trabajar con datos de panel. Los datos de panel tienen ventajas respecto a los datos de corte transversal, porque “pueden modelar formalmente los efectos dinámicos y la heterogeneidad entre grupos que son típicos en los datos microeconómicos” (Greene, 2018, p. 459). En ese sentido, las estimaciones a través de modelos de panel son más eficientes que aquellas que son por corte transversal. Se puede usar Mínimos Cuadrados Ordinarios (MCO) en datos de panel; sin embargo, es necesario conocer sus supuestos y limitaciones.

En esa línea, el Pool-MCO tiene los mismos supuestos que el MCO de corte transversal: linealidad en parámetros, exogeneidad estricta, homocedasticidad, no autocorrelación y no multicolinealidad (Greene, 2018). El Pool-MCO ofrece un

estimador insesgado, consistente y eficiente (Greene, 2018). A continuación, se presenta el modelo mencionado.

$$Y_{it} = x_{it}\beta + u_{it} \quad (6)$$

$$\text{donde, } i = 1, 2, \dots, N; t = 1, 2, \dots, T; x_{it} = [1 \ X_{2it} \ X_{3it} \ \dots \ X_{kit}]$$

Como se observa en este modelo el Y_{it} depende de las variables que varían según cada individuo y periodo, x_{it} , y un término de error, u_{it} . A este modelo se le puede hacer la crítica de que el u_{it} , contiene variables que no son observables en el individuo, y que son constantes a lo largo del tiempo o que cambian muy poco (Greene, 2018). En consecuencia, puede existir una autocorrelación intra individuo, más aún, en el contexto de esta investigación, donde los insumos actuales de los escolares se pueden relacionar con insumos pasados (Hanushek, 1972). A partir de esta precisión, se recurre a un modelo con heterogeneidad inobservable, donde el término de error se descompone en:

$$u_{it} = a_i + \varepsilon_{it}$$

Donde, a_i es el error invariante en el tiempo y ε_{it} es el error idiosincrático. Ante ello, se puede emplear el modelo de Efectos Aleatorios (EA) o Efectos Fijos (EF), donde la única diferencia radica en el supuesto de que a_i no se correlaciona con x_{it} en EA, mientras que a_i sí se correlaciona con x_{it} en EF. Además, si $E[a_i|X_i] \neq 0$, entonces, la correlación entre X_i y a_i es distinta de 0, i.e., $\text{Cov}[a_i|X_i] \neq 0$.

En el contexto de esta investigación longitudinal, donde los escolares son los mismos en el tiempo, hay variables no observables que no cambian en el tiempo, o que cambian muy poco, e.g., habilidades innatas, expectativas educativas de los escolares y padres, preferencias por la educación, etc., entonces, es posible creer que el error que no cambia en el tiempo, a_i , se correlaciona con x_{it} . En consecuencia, es preferible emplear una estimación por EF. En el caso de que a_i , no se correlacione con x_{it} , la estimación se realiza mediante EA.

Los modelos de EA y EF se pueden emplear siempre y cuando se cumple con el supuesto de exogeneidad estricta entre los regresores, x_{it} , y el error idiosincrático, ε_{it} (Greene, 2018; Wooldridge, 2010). Si no se cumple con dicho supuesto, x_{it} resulta endógeno en los modelos de EA y EF, y se obtiene un sesgo en las estimaciones de

dichos modelos, y se requiere usar una variable instrumental (VI) para una estimación consistente (Greene, 2018).

A la luz de estos métodos, primero, se menciona el modelo a estimarse y las variables empleadas. Posteriormente, se aplican los test de Breush-Pagan y Hausman para determinar si es mejor un Pool-MCO, EA o EF. Luego, se propone una VI considerando los aportes de la revisión de la literatura. Finalmente, se aplica un test de Hausman para escoger entre un modelo sin VI versus un modelo con VI.

El siguiente modelo de estimación considera la especificación de insumos acumulados propuesto por Hanushek (1972), ya que los insumos intervienen a lo largo del tiempo.

$$P_{it} = \beta_1 T_{it} + \beta_2 X_{it} + \alpha_i + \varepsilon_{it} \quad (7)$$

Donde,

P_{it} : Puntaje estandarizado de una prueba del escolar i en el tiempo t .

T_{it} : Tiempo de desplazamiento del escolar i en el tiempo t .

X_{it} : Variables control del escolar i en el tiempo t .

α_i : Variables del escolar i que no cambian en el tiempo.

ε_{it} : Error idiosincrático del escolar i en el tiempo t .

En el modelo econométrico planteado, se espera que un mayor tiempo de desplazamiento reduzca el puntaje estandarizado de las pruebas de matemática, PPVT y comprensión lectora, ya que reduce el tiempo dedicado al estudio. Asimismo, las variables control que se emplean son: camina, índice de riqueza, número de hermanos y grado.

A partir de la ecuación 7, se aplican el test de Breusch-Pagan y el test de Hausman para conocer qué modelo es el más adecuado para las estimaciones. Además, los resultados de dichos test se obtienen cuando se usan las variables control mencionadas en el párrafo anterior⁸. En los anexos 6, 7 y 8, se puede observar con detalle los resultados y significancias de los dos tests. Cabe resaltar que se usan los

⁸ En los anexos 6, 7 y 8, también se muestran los resultados de ambos test usando las variables control.

puntajes estandarizados en las áreas de matemática y comprensión lectora, y el puntaje bruto en PPVT en ambos tests.

Cuando se aplica el test de Breusch-Pagan, se obtiene que es mejor usar un modelo de heterogeneidad inobservable frente al Pool-MCO. Los resultados obtenidos en dicho test, resultan 170.34***, 86.43*** y 39.32*** para las pruebas de matemática, la prueba PPVT y comprensión lectora, respectivamente⁹.

Posteriormente, se aplica el test de Hausman para seleccionar entre un modelo de EA o EF. Como resultado de este último test, se obtiene: 123.76***; 21.71***; 380.91*** en matemática, PPVT y comprensión lectora, respectivamente. Por lo tanto, se opta por usar un modelo de EF en las tres evaluaciones. En ese sentido, el modelo de EF es el siguiente:

$$(P_{it} - \bar{P}_i) = \beta_1(T_{it} - \bar{T}_i) + \beta_2(X_{it} - \bar{X}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i) \quad (8)$$

Por otro lado, el tiempo de desplazamiento puede estar influenciado por la presencia o ausencia de centros educativos en la comunidad donde vive el escolar (Afoakwah & Koomson, 2021). En ese sentido, la posible endogeneidad del tiempo de desplazamiento se debe a que, si en la comunidad rural del escolar hay, al menos, un centro educativo que ofrezca el nivel educativo que le corresponde al escolar, el tiempo de desplazamiento será menor en comparación al caso en que no hubiese ningún centro educativo. En ese sentido, el escolar que no tiene un centro educativo en su comunidad, tendrá que recorrer mayores distancias para acceder al nivel de primaria o secundaria, aunque, es más frecuente que sea el nivel de secundaria, ya que existen menos centros educativos con nivel de secundaria en el área rural (MINEDU, ca. 2016).

En esa línea, se busca una variable instrumental (VI), porque hay la posibilidad de que el tiempo de desplazamiento sea endógeno, lo cual puede generar sesgo en las estimaciones de EA o EF. Ante ello, si se usa una VI, se obtiene una estimación consistente a comparación de usar EA o EF (Greene, 2018).

A partir de ello, se selecciona la VI, “presencia de centro educativo”. Esta variable indica si hay o no hay un centro educativo que ofrezca el grado o año correspondiente

⁹ Estos resultados consideran como variable control al tipo de desplazamiento. En los anexos 6, 7 y 8 también se muestra el resultado sin incluir tal variable control. Cabe señalar que *** indica significancia al 1%; **, significancia al 5%; y *, significancia al 10%.

al escolar. Además, dicho centro educativo debe estar en la comunidad donde vive el estudiante, tal como se ha señalado en la tabla 1. Esta VI indica que, si hay un centro educativo en la misma comunidad del escolar, el tiempo de desplazamiento que el escolar emplea para ir a su centro educativo, probablemente sea bajo. En cambio, si no hay un centro educativo en su comunidad, el escolar tiene que emplear un mayor tiempo de desplazamiento. Asimismo, la VI no se relaciona con ε_{it} porque la presencia de centro educativo se asume como una variable dada, según las características sociales y/o geográficas de la comunidad, y no depende del escolar o de su hogar.

En esa línea, se usa la variable instrumental y se regresiona en dos etapas, donde, T_{it}^0 : representa al tiempo de desplazamiento, el cual es un regresor endógeno correlacionado con ε_{it} . Además, la VI (z_{it}) toma el valor de 1 si en la comunidad del escolar existe algún centro educativo según el grado del escolar, y 0, en caso que no haya ningún centro educativo según el grado. La siguiente ecuación emplea Efectos fijos con la VI. En la primera etapa se regresiona a $(T_{it}^0 - \bar{T}_1^0)$ contra $(z_{it} - \bar{z}_1)$ y los demás regresores. Posteriormente, se calcula $(\widehat{T_{it}^0} - \bar{T}_1^0)$, y luego se estima la segunda etapa, y se obtiene el coeficiente β^0 sin sesgo e inconsistencia (Greene, 2018; Wooldridge, 2011). A continuación, se presenta el siguiente modelo.

$$P_{it} - \bar{P}_1 = \beta^0 (\widehat{T_{it}^0} - \bar{T}_1^0) + \beta(X_{it} - \bar{X}_t) + \xi_{it} \quad (9)$$

El modelo de la ecuación (9) se emplea para ciertas especificaciones de los modelos en los que se usa la VI. En caso que no se use la VI, se emplea el modelo de la ecuación 8. Asimismo, la estimación por VI debe cumplir las condiciones de exogeneidad y relevancia, sin embargo, si no las cumple, se obtiene estimaciones sesgadas e inconsistentes (Greene, 2018). Ante ello, se usa la “regla de oro” de Staiger y Stock (1997), la cual señala que el estadístico F debe ser mayor a 10 en el caso de un solo regresor endógeno.

7. Estimaciones: Principales resultados

En esta sección, se presentan los resultados para cada prueba. Para matemática y PPVT, se emplean tres rondas, y en comprensión lectora, dos rondas¹⁰. Asimismo, en matemática y comprensión lectora se toman los puntajes estandarizados, ya que no son pruebas comparables en el tiempo, mientras que en PPVT, se toma el puntaje bruto, puesto que sí es una prueba comparable en el tiempo.

Además, para escoger entre los modelos que se prefieren utilizar para las estimaciones, se usan los test de Breusch-Pagan y test de Hausman. En ambos test, se puede rechazar la hipótesis nula hasta el diez por ciento. Similarmente, se reconoce el nivel de significancia de las variables hasta un diez por ciento en las especificaciones.

En el análisis de cada prueba, primero, se presentan los resultados. Luego, se compara los resultados de las especificaciones, la validez del instrumento y el test de Hausman. Después, se comparan los resultados con los de otros autores. Posteriormente, se presenta una subsección de balance de los resultados obtenidos.

7.1. Resultados en matemática

En Matemática, se descarta el modelo de MCO agrupado en favor del modelo de heterogeneidad inobservable mediante el test de Breusch-Pagan. A partir de ello, se compara entre un modelo de EA y EF mediante el test de Hausman, y se rechaza la hipótesis nula al uno por ciento. En consecuencia, se prefiere usar EF. Cabe mencionar que los resultados y significancias de ambos test se observan en el anexo D.

En los resultados de matemática, se presentan distintas especificaciones. Las dos primeras especificaciones usan EF, y la tercera y cuarta especificación usan Efectos Fijos con variable instrumental (EFVI). Cabe mencionar que, la especificación más apropiada es la segunda, ya que controla por el tipo de desplazamiento al centro educativo.

¹⁰ Como se ha mencionado anteriormente, la prueba de comprensión lectora es evaluada solo en la ronda 4 y 5. En cambio, las demás pruebas son evaluadas en las rondas 3, 4 y 5.

Tabla 7

Resultados en matemática

VARIABLES	(1) EF	(2) EF	(3) EFVI	(4) EFVI
desplazamiento	0.0011 (0.0009)	0.0010 (0.0009)	0.0001 (0.0031)	-0.0004 (0.0032)
camina		-0.1219 (0.1112)		-0.1334 (0.1144)
riqueza	-0.2013 (0.3343)	-0.2092 (0.3343)	-0.2033 (0.3347)	-0.2128 (0.3351)
hermanos	-0.0834** (0.0334)	-0.0862** (0.0335)	-0.0852** (0.0339)	-0.0891*** (0.0341)
grado	-0.0080 (0.0105)	-0.0113 (0.0110)	-0.0060 (0.0124)	-0.0087 (0.0125)
Constante	0.2878* (0.1496)	0.4256** (0.1954)	0.3198* (0.1808)	0.4849** (0.2365)
Observaciones	882	882	882	882
R-squared	0.015	0.017		
Número de escolares	294	294	294	294
Validez del instrumento			48.72	46.52

Errores estándar en paréntesis

*** p<0.01, ** p<0.05, * p<0.1

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Tabla 8

Test de Hausman para estimaciones en matemática

Especificaciones	EFVI (3) versus EF (1)	EFVI (4) versus EF (2)
Resultado	0.1	0.2

*** p<0.01, ** p<0.05, * p<0.1

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

En primer lugar, no se observa un efecto negativo y significativo del tiempo de desplazamiento sobre el rendimiento escolar en matemática en las estimaciones de la tabla 7. Asimismo, los coeficientes del tiempo de desplazamiento en dichas especificaciones son bajos en comparación a los coeficientes de otras variables como hermanos o riqueza.

En las dos primeras especificaciones se observa que los coeficientes del tiempo de desplazamiento tienen valores 0.0011 y 0.001 en la primera y segunda especificación respectivamente; los cuales son bajos respecto a las variables control. Sin embargo, no se puede aseverar que hay un efecto positivo y significativo del tiempo de desplazamiento sobre el rendimiento en matemática, ya que la variable tiempo de desplazamiento puede ser endógena, lo cual se aborda en párrafos posteriores.

A partir de ello, se introduce la VI, presencia de centro educativo, porque el tiempo de desplazamiento puede ser influenciado por la presencia de un centro educativo en la comunidad del escolar. Si hay un centro educativo en la comunidad, el escolar emplea un cierto tiempo para asistir al centro educativo. No obstante, si no lo hay, entonces, el escolar tiene que emplear un mayor tiempo de desplazamiento.

Cuando se estima la tercera y cuarta especificación se observa que no hay un efecto negativo y significativo del tiempo de desplazamiento. Además, en la cuarta especificación se controla por la variable camina, por lo que el signo del coeficiente del tiempo de desplazamiento cambia de positivo a negativo. Asimismo, en ambas especificaciones, el tamaño de los coeficientes del tiempo de desplazamiento es bajo.

En segundo lugar, en la tercera y cuarta especificación, se observa que el F estadístico de la regla de oro resulta 48.72 y 46.52, respectivamente, con lo cual, se cumple dicha regla. No obstante, cuando se emplea el test de Hausman para comparar las especificaciones de EF versus EFVI, se observa que es mejor emplear las especificaciones que no incluyen la VI. En la tabla 8, se observa que no se puede rechazar la hipótesis nula ni al 10%, por lo que es mejor usar EF.

En tercer lugar, en las especificaciones de EF, si bien la magnitud de los coeficientes del tiempo de desplazamiento es pequeña, el signo de estos resulta positivo, lo cual, puede tener sentido cuando se compara con el planteamiento teórico de Gertler y Glewwe (1989), quienes señalan que, en el Perú rural, hay ciertos padres que están dispuestos a enviar a sus hijos a centros educativos lejanos a la comunidad donde viven, a cambio de que reciban una educación de mejor calidad. Asimismo, estos resultados son similares al hallazgo por Thomas (2016) y Contreras et al. (2018) cuando emplea MCO, ya que se observa que hay una asociación positiva entre el tiempo de desplazamiento y el rendimiento en matemática. Entonces, si bien el escolar

se desplaza un mayor tiempo, lo cual, disminuye su tiempo dedicado al estudio, recibe una educación de mejor calidad (Gertler & Glewwe, 1989; Tigre et al., 2017). En esa línea, aquellos escolares que emplean un mayor tiempo de desplazamiento para asistir a un centro educativo, pueden reflejar que tienen un alta motivación o compromiso con su rendimiento escolar.

Cuando se compara los resultados de la tabla 7 frente a otros estudios nacionales previos, se puede llegar a los siguientes puntos. El estudio de Jacoby et al. (1999) usa datos de 360 escolares de Huaraz, mientras que, en el presente estudio, se emplean datos de escolares de todos los departamentos y regiones. Por otro lado, el estudio de Benavides (2002) tiene una ventaja sobre este estudio, ya que cuenta con alrededor de 13,000 observaciones; sin embargo, solo estudia a escolares urbanos en un solo periodo y grado específico. Asimismo, la ventaja de la presente investigación es que usa datos de panel en el rendimiento en matemática.

7.2. Resultados en PPVT

En cuanto a PPVT, se descarta el modelo de homogeneidad total en favor del modelo de heterogeneidad mediante el test de Breusch-Pagan. A partir de ello, se compara entre un modelo de EA y EF mediante el test de Hausman, y se rechaza la hipótesis nula al uno por ciento. En consecuencia, se prefiere usar EF. Cabe mencionar que los resultados y significancias de ambos test se observan en el anexo E.

En ese sentido, se hacen las estimaciones en PPVT mediante EF. En las dos primeras, no se usa la VI, y en las siguientes dos, sí se emplea. Posteriormente, se presentan las comparaciones entre EFVI y EF mediante el test de Hausman. Cabe señalar que la especificación más adecuada es la segunda, puesto que controla por el tipo de desplazamiento al centro educativo.

Tabla 9

Resultados en PPVT

VARIABLES	(1) EF	(2) EF	(3) EFVI	(4) EFVI
desplazamiento	-0.0082 (0.0146)	-0.0080 (0.0147)	-0.0853 (0.0555)	-0.0860 (0.0569)
camina		0.4410 (1.8783)		-0.1807 (1.9746)
riqueza	5.1371 (5.6001)	5.1673 (5.6064)	4.9297 (5.7402)	4.9156 (5.7504)
hermanos	-1.4374** (0.5686)	-1.4264** (0.5710)	-1.5359*** (0.5867)	-1.5412*** (0.5910)
grado	5.8895*** (0.1769)	5.9014*** (0.1842)	6.0615*** (0.2170)	6.0580*** (0.2186)
Constante	36.3003*** (2.5337)	35.8004*** (3.3111)	38.7595*** (3.1060)	38.9847*** (4.0658)
Observaciones	840	840	840	840
R-squared	0.747	0.747		
Número de escolares	280	280	280	280
Validez del instrumento			43.54	41.55

Errores estándar en paréntesis

*** p<0.01, ** p<0.05, * p<0.1

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Tabla 10

Test de Hausman para estimaciones en PPVT

Especificaciones	EFVI (3) versus EF (1)	EFVI (4) versus EF (2)
Resultado	2.07	2.02

*** p<0.01, ** p<0.05, * p<0.1

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

En primer lugar, no se observa un efecto negativo y significativo del tiempo de desplazamiento sobre el rendimiento en PPVT en las estimaciones de la tabla 9. Cabe mencionar que el signo del coeficiente del tiempo de desplazamiento resulta negativo en las cuatro estimaciones. En la especificación 1 y 2, los coeficientes resultan -0.0082

y -0.008, aunque no son significativos. Sin embargo, el signo del efecto del tiempo de desplazamiento es negativo como se espera en la hipótesis. A partir de estas especificaciones, se introduce la VI, y el signo del tiempo de desplazamiento se mantiene en negativo, pese a que los coeficientes no son significativos. El tamaño de los coeficientes en EFVI resulta alrededor de 0.08, el cual es mayor que los coeficientes estimados por EF. Cabe señalar que las variables control que resultan significativas son el número de hermanos y grado.

En segundo lugar, si bien el instrumento es válido, ya que el F estadístico es 43.54 y 41.55 en las especificaciones 3 y 4, con lo cual se cumple con la regla de oro, según el test de Hausman, cuando se comparan las especificaciones de EF vs EFVI, es mejor usar las especificaciones de EF, ya que no se puede rechazar la hipótesis nula ni al 10%, como se observa en la tabla 10. En tercer lugar, este resultado en PPVT es similar a lo encontrado por Jacoby et al. (1999), ya que el signo del coeficiente del tiempo de desplazamiento es negativo; aunque en Jacoby et al. (1999) el coeficiente resulta significativo.

7.3. Resultados en comprensión lectora

Similar al caso de PPVT, en comprensión lectora, se descarta el modelo de MCO agrupado en favor del modelo de heterogeneidad inobservable mediante el test de Breusch-Pagan. A partir de ello, se compara entre un modelo de EA y EF mediante el test de Hausman, y se rechaza la hipótesis nula al uno por ciento. En consecuencia, se prefiere usar EF. Los resultados y significancias de ambos test se observan en el anexo F.

A continuación, se presentan las estimaciones en comprensión lectora. Todas las especificaciones usan EF. En las dos primeras no se usa la VI, y en las siguientes dos, sí se emplea. Posteriormente, se presentan las comparaciones usando el test de Hausman. La segunda especificación es la más adecuada, ya que controla por la forma en que se desplaza el escolar al centro educativo.

Tabla 11

Resultados en comprensión lectora

VARIABLES	(1) CMP1	(2) CMP2	(3) VI_CMP1	(4) VI_CMP2
desplazamiento	-0.0004 (0.0012)	-0.0004 (0.0012)	-0.0027 (0.0049)	-0.0021 (0.0051)
camina		0.1032 (0.1297)		0.0921 (0.1339)
riqueza	-0.2292 (0.5069)	-0.2120 (0.5076)	-0.2033 (0.5127)	-0.1939 (0.5119)
hermanos	0.0367 (0.0561)	0.0397 (0.0562)	0.0404 (0.0569)	0.0423 (0.0569)
grado	0.0056 (0.0210)	0.0104 (0.0218)	0.0125 (0.0258)	0.0153 (0.0259)
Constante	-0.0104 (0.2650)	-0.1468 (0.3157)	0.0172 (0.2729)	-0.1107 (0.3328)
Observaciones	648	648	648	648
R-squared	0.002	0.004		
Número de escolares	324	324	324	324
Validez del instrumento			20.09	19.00

Errores estándar en paréntesis

*** p<0.01, ** p<0.05, * p<0.1

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Tabla 12

Test de Hausman para estimaciones en comprensión lectora

Especificaciones	EFVI (3) versus EF (1)	EFVI (4) versus EF (2)
Resultados	0.22	0.13

*** p<0.01, ** p<0.05, * p<0.1

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

En primer lugar, no se observa un efecto negativo y significativo del tiempo de desplazamiento sobre el rendimiento en comprensión lectora en las estimaciones de la tabla 11. Cabe señalar que el signo del coeficiente del tiempo de desplazamiento resulta negativo en las cuatro estimaciones. En la primera y segunda especificación,

el coeficiente resulta -0.0004. Asimismo, cuando se introduce la VI, los coeficientes resultan -0.0027 y -0.0021 en las especificaciones 3 y 4, respectivamente.

En segundo lugar, si bien el instrumento resulta válido, ya que su valor es de 20.09 y 19 en las especificaciones 3 y 4, según el test de Hausman, es mejor usar las estimaciones de EF con respecto EFVI, ya que no se rechaza la hipótesis nula ni al 10%. Los resultados del test, se pueden observar en la tabla 12.

En tercer lugar, cuando se compara estos hallazgos con los de Jacoby et al. (1999) se encuentra que dichos autores también encuentran un signo negativo; aunque en su estudio el coeficiente resulta significativo. Asimismo, Afoakwa y Koomson (2021) encuentran un efecto negativo y significativo sobre una dummy si los escolares son capaces de leer en inglés, francés y una lengua nativa. Por lo que sería mejor si solo se compara con los hallazgos de Jacoby et al. (1999), ya que el estudio de Afoakwa y Koomson (2021) usan una dummy, pero no un puntaje de un test como variable dependiente.

7.4. Balance de los resultados

A partir de las estimaciones mediante EF sin usar la variable instrumental, las cuales son presentadas en las tablas 7, 9 y 11, el coeficiente del tiempo de desplazamiento no es significativo y negativo en ninguna evaluación. No obstante, el signo del coeficiente varía según la evaluación. En matemática se observa un signo positivo porque podría ocurrir que los escolares intercambian un mayor tiempo de desplazamiento a cambio de recibir una educación de mejor calidad en un centro educativo lejano a su comunidad. Sin embargo, en PPVT y comprensión lectora cambia el signo. Estas dos últimas evaluaciones se asocian especialmente con variables del entorno familiar (Jacoby et al., 1999).

Por otra parte, en el anexo G, se presentan dos estimaciones por cada evaluación, donde el tiempo de desplazamiento es categorizado mediante dummies. La primera categoría abarca a aquellos escolares que emplean entre un minuto a 29 minutos, y cuyo valor es 0. Respecto a la segunda categoría, si el escolar se desplaza entre 30 a 59 minutos, toma el valor de 1. Referente a la tercera categoría, si se desplaza entre

una hora a cuatro horas, toma el valor de 2. Cabe mencionar que la segunda estimación, a diferencia de la primera, controla por la variable camina.

A partir de ello, se encuentran los siguientes resultados. Respecto a la primera estimación en matemática, el coeficiente del tiempo de desplazamiento de la tercera categoría resulta 0.1552 y es significativo al diez por ciento cuando no se controla por la variable camina. Este resultado refleja el *trade-off* que puede ocurrir entre un mayor tiempo de desplazamiento a cambio de recibir una educación de buena calidad. En la segunda estimación, cuando se controla por la variable camina, la significancia desaparece; sin embargo, el signo del coeficiente se mantiene positivo. En cuanto a las dos estimaciones de PPVT, el efecto resulta negativo y no significativo para las categorías del tiempo de desplazamiento. Referente a comprensión lectora, pese a que el signo de las categorías del tiempo de desplazamiento varía según los intervalos de desplazamiento, no se encuentran efectos significativos en ambas estimaciones.

En conclusión, a partir de los resultados de las tablas 7 a la 12 y las tablas del anexo G, en esta investigación se señala que no hay evidencia de un efecto negativo y significativo del tiempo de desplazamiento sobre el rendimiento escolar en las evaluaciones de matemática, PPVT y comprensión lectora en el área rural peruana. En ese sentido, se rechaza la hipótesis planteada. No obstante, la ventaja del presente estudio se basa en que usa datos de panel, mientras que los estudios mencionados en la literatura empírica emplean datos de corte transversal principalmente¹¹. En ese sentido, hay una mejora en el método de estimación, ya que emplear un modelo como el de EF genera una ganancia en eficiencia y ofrece mejores estimaciones que un MCO. Además, en este estudio se incluyen escolares que viven en distintos departamentos del Perú.

¹¹ En el estudio de Jacoby et al. (1999), se usa datos de panel que abarca dos periodos. No obstante, el efecto del tiempo de desplazamiento sobre el rendimiento escolar es analizado en un periodo.

8. Limitaciones

En esta sección se presentan las limitaciones de la presente investigación. En primer lugar, una limitación consiste en que no se puede acceder a la información de la calidad de los centros educativos a los que asisten los escolares. Asimismo, no se puede conocer los nombres de los centros educativos, ya que hay cierta confidencialidad en la base de datos utilizada. La información de los centros educativos sirve para controlar por variables de oferta educativa, lo cual se recomienda en la literatura teórica y empírica. En segundo lugar, la base de datos no contiene datos de los compañeros de clase del niño o adolescente en las rondas 3, 4 y 5 de la cohorte menor, por lo que no se ha usado variables asociadas a los pares en las rondas mencionadas.

En tercer lugar, si bien las pruebas se pueden hacer comparables mediante la estandarización, sería de utilidad que las pruebas de matemática y comprensión lectora sean comparables en el tiempo como en el caso del PPVT. Asimismo, la prueba de comprensión lectora solo cuenta con puntajes para la ronda 4 y 5. En ese sentido, ayudaría que se cuente con tres puntajes como en el resto de las pruebas.

En cuarto lugar, no se puede probar la condición de exclusión del instrumento, ya que es un caso exactamente identificado, donde solo se dispone de una VI, dada la naturaleza de los datos. Finalmente, otra limitación radica en la baja cantidad de observaciones de escolares rurales.

9. Conclusiones

La presente investigación estudia el efecto del tiempo de desplazamiento entre vivienda y el centro educativo, ida y vuelta, sobre el rendimiento escolar en el área rural peruana. Este tema es muy interesante, pero poco estudiado en la literatura económica, educativa y en otras disciplinas a nivel nacional. Asimismo, escasos estudios usan datos longitudinales para explorar el efecto del tiempo de desplazamiento sobre el rendimiento. Además, el tiempo de desplazamiento no suele ser una variable que se considera como un factor o determinante del rendimiento escolar en el Perú. En ese sentido, esta investigación contribuye al análisis del efecto del tiempo de desplazamiento sobre el rendimiento escolar mediante el uso de datos longitudinales.

A la luz de la literatura teórica, se enmarca un modelo donde un mayor tiempo de desplazamiento reduce el tiempo dedicado al estudio, lo cual puede reducir el rendimiento escolar (Levin & Tsang, 1987). Sin embargo, en el Perú rural, ciertos padres pueden enviar a sus hijos a centros educativos lejanos a su comunidad a cambio de que dichos centros sean de mejor calidad (Gertler & Glewwe, 1989). Si bien, algunos estudiantes recorren largas distancias, lo cual puede reducir su tiempo de estudio, a cambio de ello, obtienen una educación de mejor calidad, la cual incide favorablemente en su rendimiento escolar.

Asimismo, en la literatura empírica, se pueden observar diversos hallazgos. Por un lado, Jacoby et al. (1999), Benavides (2002), Tigre et al. (2017) y Afoakwah (2021) encuentran un efecto negativo del tiempo de desplazamiento sobre el rendimiento. Sin embargo, Thomas (2016) encuentra una asociación positiva entre el tiempo de desplazamiento, ya que los escolares que recorren mayores distancias son quienes asisten a centros educativos de buena calidad. En ese sentido, el hallazgo de Thomas (2016) refuerza el aporte teórico de Gertler y Gleewe (1989).

A partir de los resultados obtenidos en la presente investigación, el coeficiente del tiempo de desplazamiento resulta positivo, aunque no significativo en el caso de matemática. Por otro lado, en el caso de PPVT y comprensión lectora, el coeficiente del tiempo de desplazamiento no es significativo, sin embargo, el signo es negativo. Esta diferencia entre las evaluaciones puede deberse a que las variables del centro educativo están asociadas al rendimiento en matemática, especialmente; mientras

que las áreas de PPVT y comprensión lectora están asociadas a variables de la familia (Jacoby et al., 1999).

Por lo tanto, en esta investigación se concluye que no hay evidencia de un efecto negativo y significativo del tiempo de desplazamiento sobre el rendimiento escolar en matemática, PPVT y comprensión lectora. En ese sentido, se rechaza la hipótesis planteada que señala que un mayor tiempo de desplazamiento de la vivienda al centro educativo (ida y vuelta) disminuye el rendimiento escolar.

El valor agregado de esta investigación consiste en que usa datos de panel, lo cual ofrece mejores estimaciones que los datos de corte transversal. Por ello, se recomienda que puedan surgir nuevas investigaciones que empleen datos longitudinales en la educación peruana.



10. Referencias bibliográficas

- Adell, M. (2006). *Estrategias para mejorar el rendimiento académico de los adolescentes* (2da ed.). Pirámide.
- Afoakwah, C., & Koomson, I. (2021). How does school travel time impact children's learning outcomes in a developing country? *Review of Economics of the Household*, 19(4), 1077–1097. <https://doi.org/10.1007/s11150-020-09533-8>
- Alcázar, L., Rogers, F. H., Chaudhury, N., Hammer, J., Kremer, M., & Muralidharan, K. (2006). Why are teachers absent? Probing service delivery in Peruvian primary schools. *International Journal of Educational Research*, 45(3), 117–136. <https://doi.org/10.1016/j.ijer.2006.11.007>
- Alvaro, M., Bueno, M., Calleja, J., Cerdán, J., Echevarría, M., García, C., Gaviria, J., Gómez, C., Jiménez, S., López, B., Javato, L., Mínguez, A., Sánchez, A., & Trillo, C. (1990). *Hacia un modelo causal del rendimiento académico* (Vol. 43). Ministerio de Educación y Ciencia. https://www.researchgate.net/profile/Carmuca-Gomez-Bueno-2/publication/39127951_Hacia_un_modelo_causal_del_rendimiento_academico/links/586cbb5308ae6eb871bb8000/Hacia-un-modelo-causal-del-rendimiento-academico.pdf
- Arregui, P., Benavides, M., Cueto, S., Hunt, B., Saavedra, J., & Secada, W. (2004). *¿Es posible mejorar la educación peruana?: evidencias y posibilidades*. Grupo de Análisis para el Desarrollo. <https://www.grade.org.pe/publicaciones/638-es-posible-mejorar-la-educacion-peruana-evidencias-y-posibilidades/>
- Becker, G. S. (1965). A theory of the allocation of time. *The Economic Journal*, 75(299), 493–517. <https://doi.org/https://doi.org/10.2307/2228949>
- Benavides, M. (2002). Para explicar las diferencias en el rendimiento en matemática de cuarto grado en el Perú urbano: análisis de resultados a partir de un modelo básico. En J. Rodríguez y S. Vargas (Eds.), *Análisis de los resultados y metodología de las pruebas CRECER 1998* (pp. 83–107). Ministerio de Educación. <https://repositorio.minedu.gob.pe/bitstream/handle/20.500.12799/225/108.%20El%20curr%C3%ADculo%20implementado%20como%20indicador%20del%20proceso%20educativo.pdf?sequence=1&isAllowed=y>
- Boyden, J. (2018). *Young Lives: an International Study of Childhood Poverty: Round 3, 2009* (SN: 6853; 3rd Edition) [Data set]. UK Data Service. <http://doi.org/10.5255/UKDA-SN-6853-3>
- Bratti, M., & Staffolani, S. (2013). Student Time Allocation and Educational Production Functions. *Annals of Economics and Statistics*, 111/112, 103–140. <https://doi.org/https://doi.org/10.2307/23646328>
- Contreras, D., Hojman, D., Matas, M., Rodríguez, P., & Suárez, N. (2018). *The impact of commuting time over educational achievement: a machine learning approach* (Series Documentos de Trabajo No. 472). Universidad de Chile. <https://repositorio.uchile.cl/handle/2250/153416>

- Cueto, S. (2007). Las evaluaciones nacionales e internacionales de rendimiento escolar en el Perú: balance y perspectivas. En E. Zegarra, V. Minaya, J. Escobal, C. Ponce, J. Kuramoto, M. Glave, L. Alcázar, M. Jaramillo, H. Ñopo, J. Díaz, N. Birdsall, R. Menezes, M. Torero, J. Deustua, M. Hernández, S. Cueto, M. Benavides, E. Pollit, J. León, ... N. Valdivia (Eds.), *Investigación, políticas y desarrollo en el Perú* (pp. 405–455). Grupo de Análisis para el Desarrollo. <https://www.grade.org.pe/publicaciones/771-las-evaluaciones-nacionales-e-internacionales-de-rendimiento-escolar-en-el-peru-balance-y-perspectivas/>
- Cueto, S., Escobal, J., Penny, M., & Ames, P. (2012). *¿Quién se queda atrás? Resultados iniciales del estudio Niños del Milenio Tercera ronda de encuestas en el Perú*. Grupo de Análisis para el Desarrollo. <https://www.grade.org.pe/publicaciones/1039-quien-se-queda-atras-resultados-iniciales-del-estudio-ninos-del-milenio-tercera-ronda-de-encuestas-en-el-peru/>
- Escobal, J., & Flores, E. (2008). *An Assessment of the Young Lives Sampling Approach in Peru* (Young Lives Technical Note No. 3). Young Lives. <https://www.younglives.org.uk/sites/www.younglives.org.uk/files/YL-TN3-Escobal-Sampling-Approach-In-Peru.pdf>
- Gertler, P., & Glewwe, P. (1989). *The Willingness to Pay for Education in Developing Countries: Evidence from Rural Peru* (Living Standards Measurement Study No. 54). The World Bank.
- Greene, W. (2018). *Econometric Analysis* (8va ed.). Pearson. <https://www.pearson.com/en-us/subject-catalog/p/econometric-analysis/P200000005909/9780134811932>
- Hanushek, E. (1972). *Education and race: An analysis of the educational production process*. Lexington, Mass., Lexington Books. <https://archive.org/details/educationraceana0000hanu/page/n7/mode/2up>
- Hanushek, E. (1979). Conceptual and Empirical Issues in the Estimation of Educational Production Functions. *The Journal of Human Resources*, 14(3), 351–388. <https://doi.org/10.2307/145575>
- Hanushek, E. (1992). The Trade-off between Child Quantity and Quality. *The Journal of Political Economy*, 100(1), 84–117. <https://www.jstor.org/stable/2138807>
- Iqbal, A., Sajid, G. M., & Khan, S. U. (2020). The Socio-Economic Determinants of Child Educational Attainment in Pakistan. *Journal of Applied Economics and Business Studies*, 4(4), 75–96. <https://doi.org/10.34260/jaebs.444>
- Jacoby, E., Cueto, S., & Pollitt, E. (1999). Determinants of School Performance Among Quechua Children in the Peruvian Andes. *International Review of Education*, 45(1), 27–43. <https://doi.org/10.1023/A:1003521804362>
- Lavado, P., Aragón, C., & Gonzales, M. (2015). ¿Cuál es la relación entre las habilidades cognitivas y no cognitivas y la adopción de comportamientos de riesgo? Un estudio para el Perú. *Apuntes: Revistas de Ciencias Sociales*, 42(76), 59–93. <https://doi.org/https://doi.org/10.21678/apuntes.76.729>
- Leon, J. (2020). *Equating Cognitive Scores across Rounds and Cohorts for Young Lives in Ethiopia, India, Peru and Vietnam* (Technical Note 51). Young Lives. https://www.younglives.org.uk/sites/default/files/migrated/YL-TN51_0.pdf

- Levin, H., & Tsang, M. (1987). The Economics of Student Time. *Economics of Education Review*, 6(4), 357–364. [https://doi.org/10.1016/0272-7757\(87\)90019-7](https://doi.org/10.1016/0272-7757(87)90019-7)
- Menges, B. (2009). *An analysis of the determinants of educational achievement in Peru*. [Tesis de maestría, Pontificia Universidad Católica del Perú].
- Ministerio de Educación. (ca. 2016). *Padrones de instituciones educativas. Padrones de asignaciones temporales. Padrones de asignaciones temporales 2016. Resolución Ministerial N.° 069-2016-MINEDU. Anexo 4: Padrón de Instituciones Educativas Públicas ubicadas en zona rural y su grado de ruralidad* [Data set]. Ministerio de Educación. <https://www.minedu.gob.pe/reforma-magisterial/padrones-instituciones-educativas.php>
- Ministerio de Educación. (2023a). *Evaluaciones nacionales. 2.° grado de primaria. Resultados Generales 2007-2022 (MC). Según nivel de desempeño. Resultados Generales 2007-2022 (MC)* [Data set]. Ministerio de Educación. <http://umc.minedu.gob.pe/bases-de-datos/>
- Ministerio de Educación. (2023b). *Evaluaciones nacionales. 2.° grado de secundaria. Resultados Generales 2015-2022 (ECE). Según nivel de desempeño. Resultados Generales 2015-2022* [Data set]. Ministerio de Educación. <http://umc.minedu.gob.pe/bases-de-datos/>
- Montero, C., Oliart, P., Ames, P., Cabrera, Z., & Uccelli, F. (2001). *La escuela rural: modalidades y prioridades de intervención*. Ministerio de Educación del Perú.
- Montero, C., & Uccelli, F. (2023). *De ilusiones, conquistas y olvidos: La educación rural en el Perú*. Instituto de Estudios Peruanos. <https://fondoeditorial.iep.org.pe/producto/de-ilusiones-conquistas-y-olvidos-la-educacion-rural-en-el-peru/>
- Pov, S., Kawai, N., & Matsumiya, N. (2021). Determinants of student achievement at lower secondary schools in rural Cambodia. *Educational Research for Policy and Practice*, 20(2), 207–222. <https://doi.org/10.1007/s10671-020-09276-4>
- Ryan, J., & Brockmann, F. (2018). *A Practitioner's Introduction to Equating* (1ra edición revisada). Council of Chief State School Officers. <https://ccsso.org/resource-library/practitioners-introduction-equating>
- Sanchez, A., Woldehanna, T., Duc, L. T., Boyden, J., Penny, M., & Galab, S. (2018). *Young Lives: an International Study of Childhood Poverty: Round 5, 2016* (SN: 8357) [Data set]. UK Data Service. <http://doi.org/10.5255/UKDA-SN-8357-1>
- Staiger, D., & Stock, J. (1997). *Instrumental Variables Regression with Weak Instruments*. 65(3), 557–586. <https://doi.org/10.2307/2171753>
- Thomas, C. (2016). *Going the Distance! How Distance to School Relates to Education Outcomes* [Tesis de doctorado, Universidad de California]. <https://escholarship.org/uc/item/97f0n3mq>
- Tigre, R., Sampaio, B., & Menezes, T. (2017). The impact of commuting time on youth's school performance. *Journal of Regional Science*, 57(1), 28–47. <https://doi.org/10.1111/jors.12289>

- Todd, P., & Wolpin, K. (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement. *The Economic Journal*, 113(485), F3–F33. <https://www.jstor.org/stable/3590137>
- Todd, P., & Wolpin, K. (2007). The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital*, 1(1), 91–136. <https://doi.org/10.1086/526401>
- Woldehanna, T., Galab, S., Sanchez, A., Penny, M., Duc, L. T., & Boyden, J. (2018). *Young Lives: an International Study of Childhood Poverty: Round 4, 2013-2014* (SN: 7931; 2nd Edition) [Data Set]. UK Data Service. <http://doi.org/10.5255/UKDA-SN-7931-2>
- Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data* (2da ed.). The MIT Press. <https://www.jstor.org/stable/j.ctt5hhcfr>
- Zhao, M., & Glewwe, P. (2010). What determines basic school attainment in developing countries? Evidence from rural China. *Economics of Education Review*, 29(3), 451–460. <https://doi.org/10.1016/j.econedurev.2009.10.008>

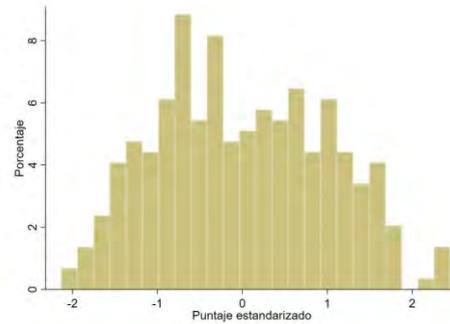


11. Anexos

Anexo A: Puntaje estandarizado en el área rural

Figura 23

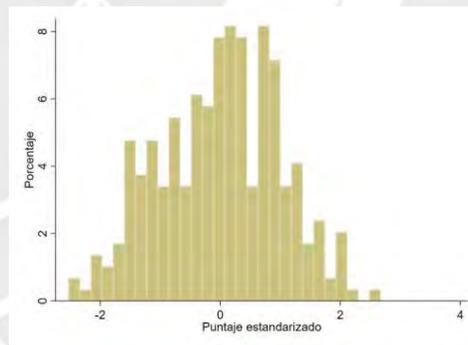
Puntaje estandarizado en matemática en el área rural en la ronda 3



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 24

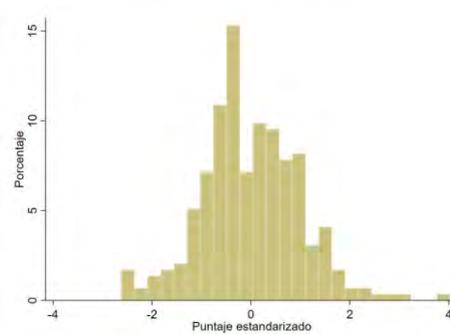
Puntaje estandarizado en matemática en el área rural en la ronda 4



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 25

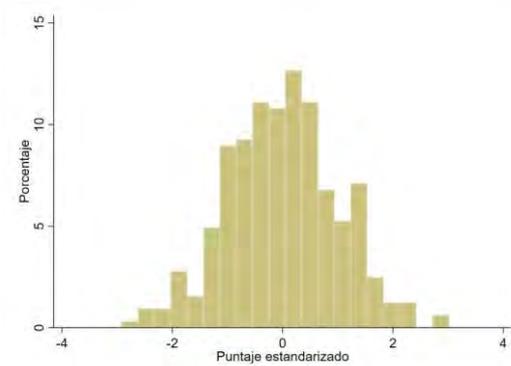
Puntaje estandarizado en matemática en el área rural en la ronda 5



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 26

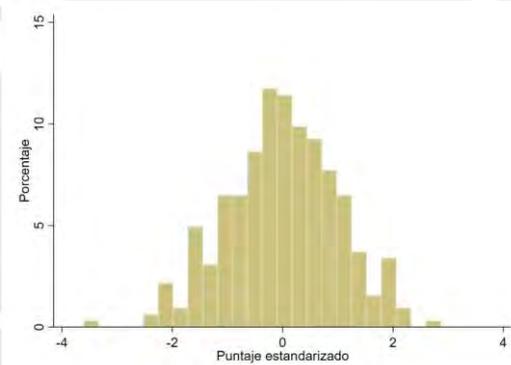
Puntaje estandarizado en comprensión lectora en el área rural en la ronda 4



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 27

Puntaje estandarizado en comprensión lectora en el área rural en la ronda 5

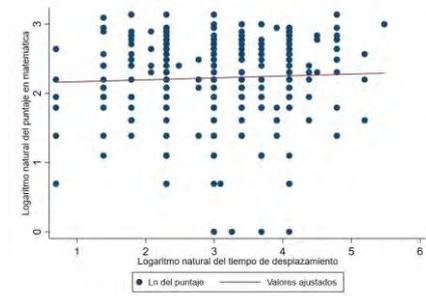


Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Anexo B: Gráficos de dispersión entre el logaritmo del tiempo de desplazamiento y el logaritmo del puntaje bruto de las pruebas

Figura 28

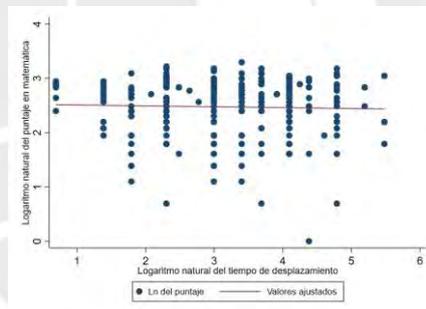
Dispersión entre el logaritmo del tiempo de desplazamiento y el logaritmo del puntaje bruto de la prueba de matemática en la ronda 3



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 29

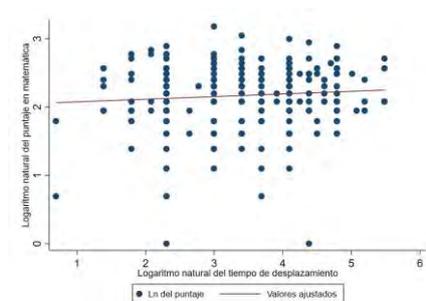
Dispersión entre el logaritmo del tiempo de desplazamiento y el logaritmo del puntaje bruto de la prueba de matemática en la ronda 4



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 30

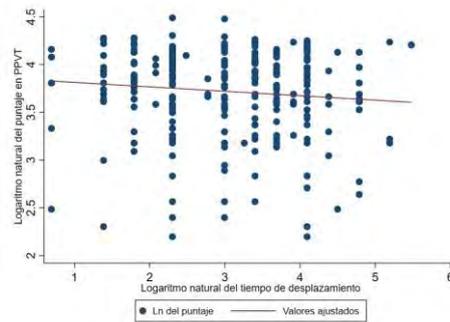
Dispersión entre el logaritmo del tiempo de desplazamiento y el logaritmo del puntaje bruto de la prueba de matemática en la ronda 5



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 31

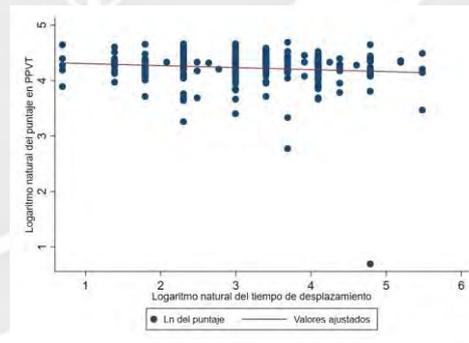
Dispersión entre el logaritmo del tiempo de desplazamiento y el logaritmo del puntaje bruto de la prueba de PPVT en la ronda 3



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 32

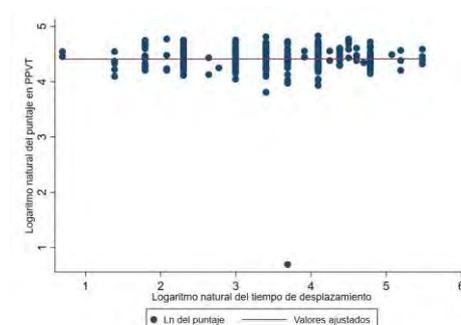
Dispersión entre el logaritmo del tiempo de desplazamiento y el logaritmo del puntaje bruto de la prueba de PPVT en la ronda 4



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 33

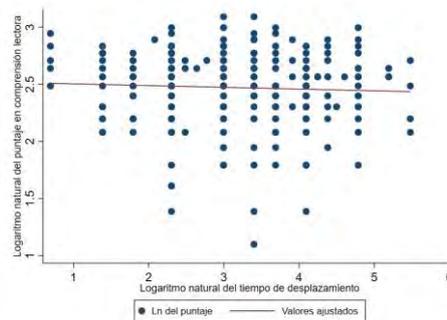
Dispersión entre el logaritmo del tiempo de desplazamiento y el logaritmo del puntaje bruto de la prueba de PPVT en la ronda 5



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 34

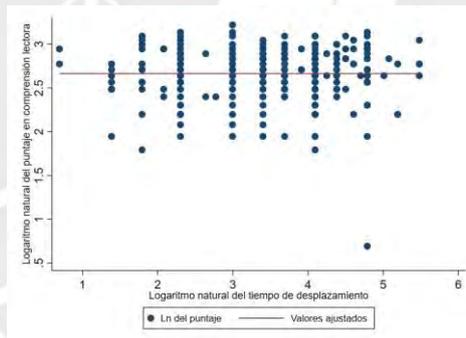
Dispersión entre el logaritmo del tiempo de desplazamiento y el logaritmo del puntaje bruto de la prueba de comprensión lectora en la ronda 4



Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Figura 35

Dispersión entre el logaritmo del tiempo de desplazamiento y el logaritmo del puntaje bruto de la prueba de comprensión lectora en la ronda 5

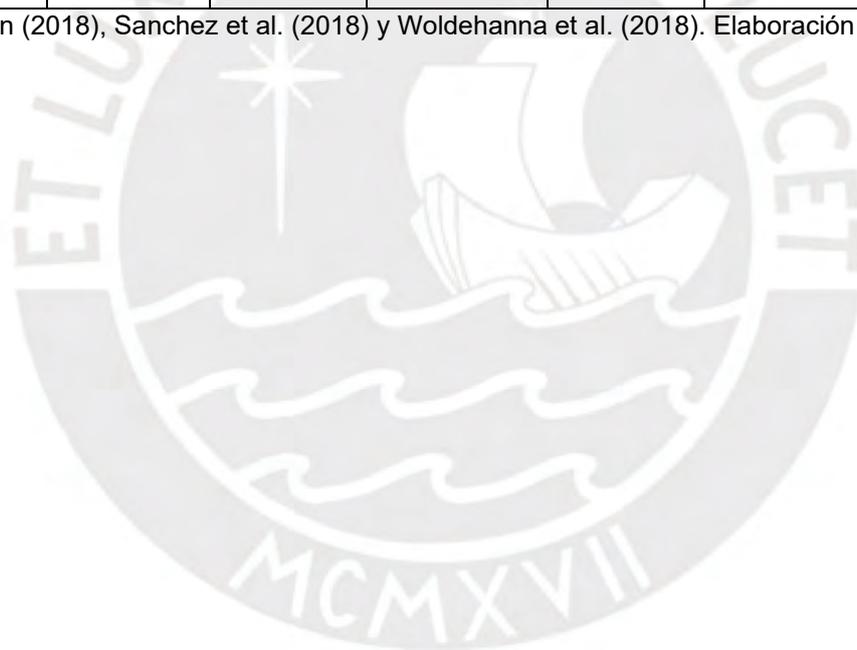


Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Anexo C: Grados de los escolares en cada ronda

Grado	Ronda 3		Ronda 4		Ronda 5	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Primer grado	23	7.82%	0	0%	0	0%
Segundo grado	207	70.41%	0	0%	0	0%
Tercer grado	63	21.43%	8	2.72%	0	0%
Cuarto grado	1	0.34%	29	9.86%	1	0.34%
Quinto grado	0	0%	61	20.75%	3	1.02%
Sexto grado	0	0%	141	47.96%	7	2.38%
Primer año	0	0%	54	18.37%	33	11.22%
Segundo año	0	0%	1	0.34%	67	22.79%
Tercer año	0	0%	0	0%	133	45.24%
Cuarto año	0	0%	0	0%	49	16.67%
Quinto año	0	0%	0	0%	1	0.34%
Total	294	100%	294	100%	294	100 %

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.



Anexo D: Test de Breusch-Pagan y Test de Hausman en Matemática

Tabla 13

Test de Breusch-Pagan para estimaciones en matemática

Especificaciones	MCO Agrupado versus EA ^a	MCO Agrupado versus EA ^b
Resultado	169.7	170.34
p-value	0.0000	0.0000

^a En esta comparación de modelos no se controla por la variable camina.

^b En esta comparación de modelos se controla por la variable camina.

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Tabla 14

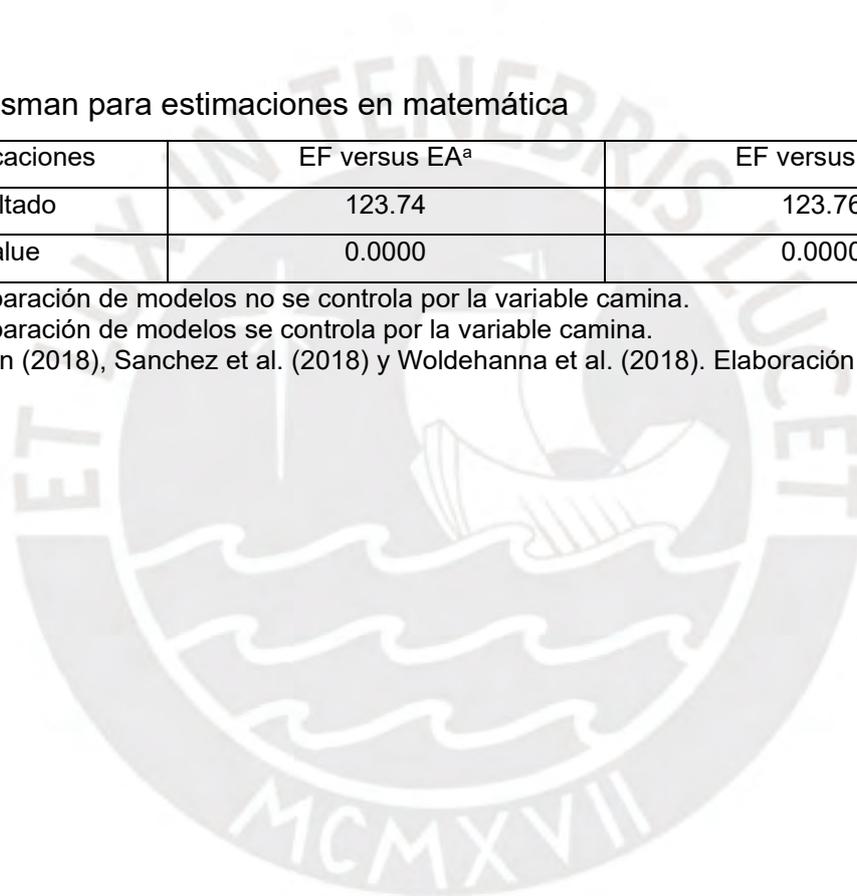
Test de Hausman para estimaciones en matemática

Especificaciones	EF versus EA ^a	EF versus EA ^b
Resultado	123.74	123.76
p-value	0.0000	0.0000

^a En esta comparación de modelos no se controla por la variable camina.

^b En esta comparación de modelos se controla por la variable camina.

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.



Anexo E: Test de Breusch-Pagan y Test de Hausman en PPVT

Tabla 15

Test de Breusch-Pagan para estimaciones en PPVT

Especificaciones	MCO Agrupado versus EA ^a	MCO Agrupado versus EA ^b
Resultado	90.38	86.43
p-value	0.0000	0.0000

^a En esta comparación de modelos no se controla por la variable camina.

^b En esta comparación de modelos se controla por la variable camina.

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Tabla 16

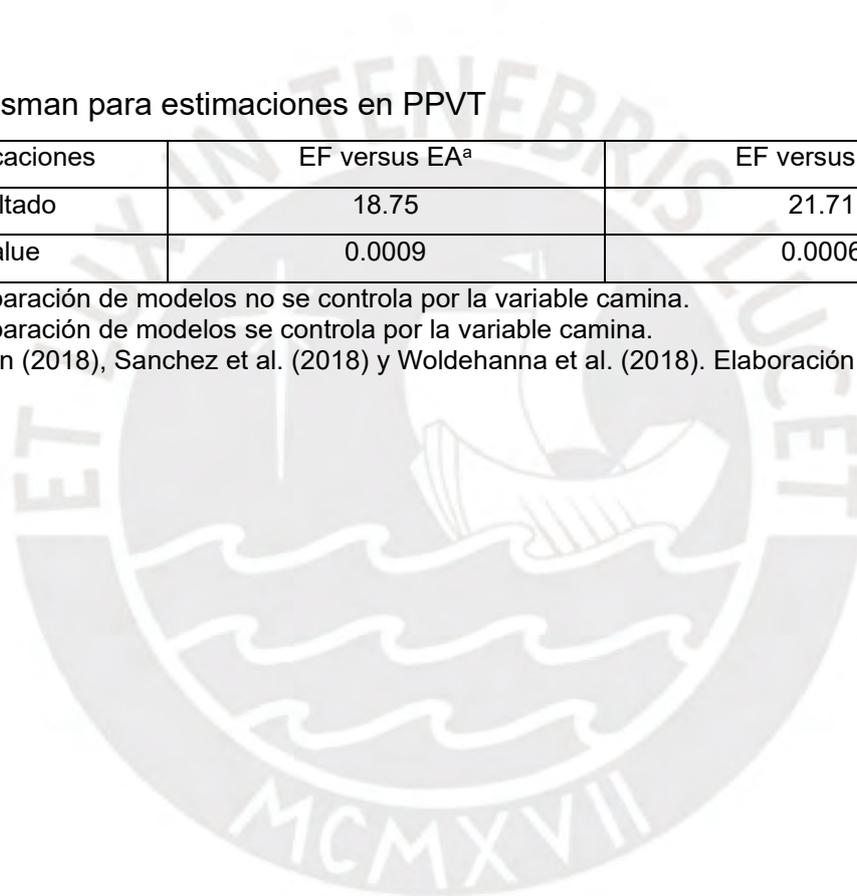
Test de Hausman para estimaciones en PPVT

Especificaciones	EF versus EA ^a	EF versus EA ^b
Resultado	18.75	21.71
p-value	0.0009	0.0006

^a En esta comparación de modelos no se controla por la variable camina.

^b En esta comparación de modelos se controla por la variable camina.

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.



Anexo F: Test de Breusch-Pagan y Test de Hausman en comprensión lectora.

Tabla 17

Test de Breusch Pagan para estimaciones en comprensión lectora

Especificaciones	MCO Agrupado versus EA ^a	MCO Agrupado versus EA ^b
Resultado	39.3	39.32
p-value	0.0000	0.0000

^a En esta comparación de modelos no se controla por la variable camina.

^b En esta comparación de modelos se controla por la variable camina.

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Tabla 18

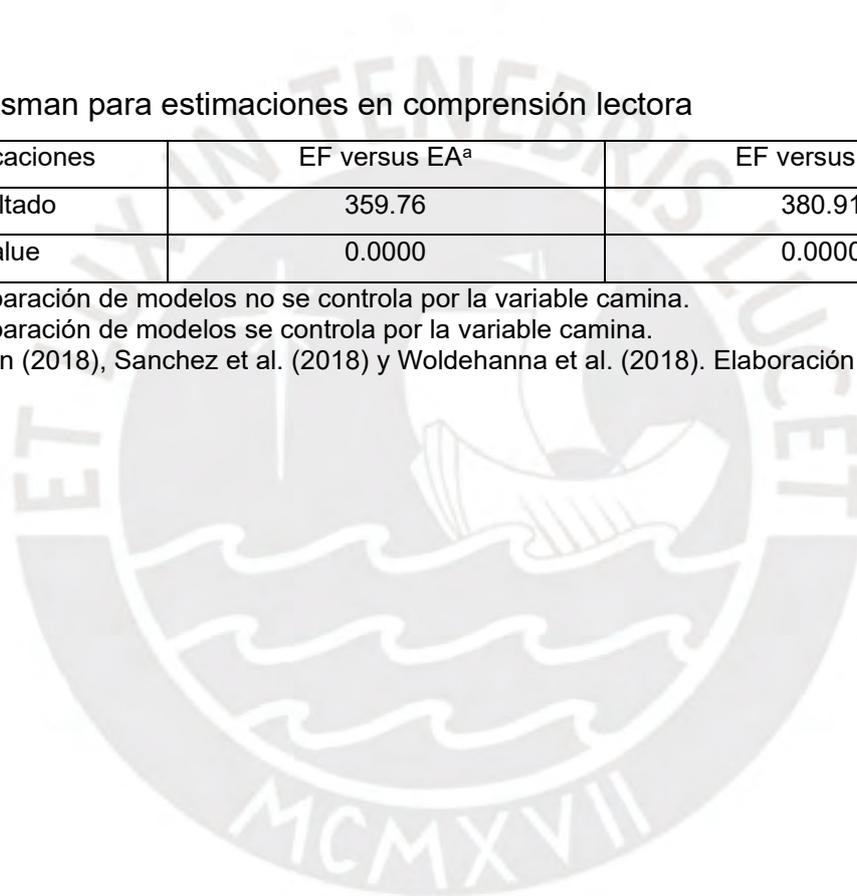
Test de Hausman para estimaciones en comprensión lectora

Especificaciones	EF versus EA ^a	EF versus EA ^b
Resultado	359.76	380.91
p-value	0.0000	0.0000

^a En esta comparación de modelos no se controla por la variable camina.

^b En esta comparación de modelos se controla por la variable camina.

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.



Anexo G: Estimaciones con dummies del tiempo de desplazamiento por pruebas

Tabla 19

Estimaciones con dummies del tiempo de desplazamiento en matemática

VARIABLES	(1)	(2)
	EF	EF
desplazamiento (30-59)	0.0378 (0.0903)	0.0238 (0.0915)
desplazamiento (60-240)	0.1552* (0.0864)	0.1423 (0.0874)
camina		-0.1094 (0.1127)
riqueza	-0.2073 (0.3353)	-0.2179 (0.3355)
hermanos	-0.0825** (0.0333)	-0.0851** (0.0334)
grado	-0.0095 (0.0106)	-0.0122 (0.0110)
Constante	0.2841* (0.1511)	0.4123** (0.2007)
Observaciones	882	882
R-squared	0.018	0.020
Número de escolares	294	294

Errores estándar en paréntesis

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Tabla 20

Estimaciones con dummies del tiempo de desplazamiento en PPVT

VARIABLES	(1)	(2)
	EF	EF
desplazamiento (30-59)	-0.8681 (1.5158)	-0.8280 (1.5357)
desplazamiento (60-240)	-0.6308 (1.4471)	-0.5940 (1.4648)
camina		0.3213 (1.9058)
riqueza	4.9052 (5.6260)	4.9378 (5.6343)
hermanos	-1.4271** (0.5690)	-1.4191** (0.5715)
grado	5.8918*** (0.1788)	5.8997*** (0.1850)
Constante	36.4041*** (2.5660)	36.0267*** (3.4067)
Observaciones	840	840
R-squared	0.747	0.747
Número de escolares	280	280

Errores estándar en paréntesis

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.

Tabla 21

Estimaciones con dummies del tiempo de desplazamiento en comprensión lectora

VARIABLES	(1) CMP1	(2) CMP2
desplazamiento (30-59)	0.0195 (0.1291)	0.0349 (0.1306)
desplazamiento (60-240)	-0.0459 (0.1243)	-0.0325 (0.1255)
1.camina		0.1067 (0.1313)
riqueza	-0.2243 (0.5088)	-0.2018 (0.5099)
hermanos	0.0368 (0.0562)	0.0398 (0.0563)
grado	0.0057 (0.0214)	0.0102 (0.0221)
Constante	-0.0210 (0.2673)	-0.1658 (0.3213)
Observaciones	648	648
R-squared	0.003	0.005
Número de escolares	324	324

Errores estándar en paréntesis

*** p<0.01, ** p<0.05, * p<0.1

Fuente: Boyden (2018), Sanchez et al. (2018) y Woldehanna et al. (2018). Elaboración propia.