

Impacto a corto plazo del absentismo docente en el desempeño escolar chileno

Tesis para optar al título de
Ingeniería Comercial con mención en Economía

Andreas Aron Winkler Alvaro Carril Rubio^ⁿ

Profesor guía:
Valentina Paredes Haz

27 de enero de 2015

Resumen

Analizamos el absentismo docente y el impacto que tiene en el desempeño de los estudiantes, medido como puntaje en la prueba SIMCE. Este es el primer estudio en Chile en analizar de forma sistemática el absentismo docente como una variable relevante para determinar desempeño, lo que aporta nuevo conocimiento a la literatura relacionada. Usando una base de datos de elaboración propia encontramos que el absentismo docente en Chile, ligeramente subreportado en el cuestionario SIMCE, es alrededor del 5% y por tanto cercano al de países desarrollados. Además, encontramos que el efecto del absentismo es negativo y significativo bajo distintas especificaciones, aunque el coeficiente es pequeño.

Códigos JEL: I21, I28, J24

Palabras clave: Educación, Absentismo, SIMCE

^ⁿAgradezco a Valentina Paredes por la confianza que depositó en esta investigación, además de su guía y entusiasmo constantes durante el desarrollo de la misma. Agradezco también a amigos, profesores y equipo directivo del Colegio Antares, que aportaron valiosa discusión y comentarios que hicieron de este un trabajo con un significado real y cercano, además de su flexibilidad para permitirme compatibilizar mis labores de profesor y estudiante. También agradezco a mis alumnos mágicos, que le dan sentido a todos los esfuerzos realizados. Finalmente agradezco a Alejandrina Correa, que siempre me recuerda del lado humano de las cosas.

Índice

1. Introducción	1
2. Contexto	2
3. Datos	5
3.1. Datos de SIMCE	5
3.2. Datos Absentismo Docente	6
3.3. Construcción de la base de datos	8
3.4. Estadística Descriptiva	8
4. Marco teórico	10
4.1. Modelo de efecto fijo	10
4.2. Modelo multinivel mixto	13
5. Estimaciones	16
5.1. Modelo de efecto fijo	16
5.2. Modelo multinivel mixto	19
6. Conclusión	21
Apéndices	25
A. Tablas	25
B. Test de hipótesis para para coeficientes de los regresores	29
C. Test de hipótesis para varianza intra-grupos	30

Índice de tablas

1.	Absentismo docente por países	3
2.	Días de absentismo promedio por asignatura, urbano vs. rural	3
3.	Comparación de estadísticas para variables seleccionadas.	8
4.	Inasistencias de profesores por tramos y asignaturas	9
5.	Inasistencias auto-reportadas vs. registros municipales	9
6.	Resultados del modelo de efecto fijo (panel de asignaturas) usando como único control la variable de interés y con distintos efectos fijos.	17
7.	Resultados del modelo de efecto fijo (panel de asignaturas), errores robustos agrupados por alumno.	18
8.	Resultados del modelo multinivel mixto con interceptos aleatorios por colegio. Modelos separados por asignatura.	20
A.1.	Consolidación de la base de datos utilizada	25
A.2.	Inasistencias de profesores por tramos y grupo socioeconómico	25
A.3.	Resumen del panel, variables seleccionadas	26
A.4.	Resultados del modelo multinivel mixto con interceptos aleatorios por año y colegio.	27
A.5.	Resultados del modelo multinivel mixto con interceptos aleatorios por colegio y curso. Modelos separados por asignatura.	28

Índice de figuras

1.	Correlación entre puntajes de matemática y lectura por alumno.	11
2.	Variación intra-grupo de desviaciones del puntaje SIMCE con respecto a la media individual contra desviaciones de días de inasistencia del profesor con respecto a la media individual.	12

1. Introducción

La importancia que tiene el profesor para el aprendizaje de sus estudiantes es innegable. Más allá de la calidad de su enseñanza, el requisito básico para que se produzca aprendizaje es que el profesor esté presente. Por lo tanto, medir el absentismo docente —los días que el profesor se encuentra ausente— y su impacto en el desempeño de los estudiantes es de suma importancia para evaluar la salud de un sistema educativo y diseñar políticas de manera informada.

A pesar de lo anterior, no existe en Chile ninguna medición sistemática y fidedigna de absentismo docente. Solo en algunos años el absentismo es una variable medida en el cuestionario SIMCE¹ para profesores, pero al ser auto-reportada presenta cierta subestimación, como evidenciamos más adelante. Quizás como resultado de ello, no existe en Chile ninguna medición del efecto que tiene el absentismo docente sobre el desempeño SIMCE. Muchos estudios nacionales y extranjeros se han abocado a estudiar los determinantes del desempeño estudiantil en pruebas estandarizadas y sin embargo el absentismo ha sido dejado de lado.

En esta investigación abordamos el problema de falta de datos utilizando una base de datos de elaboración propia, generada a partir de los registros de asistencia reportados por las Corporaciones Municipales, empleadores de los profesores de establecimientos municipales. Esta base es única y, a nuestro conocimiento, la más completa medición del absentismo docente, tanto en cuanto a la cantidad de profesores que incluye y como al rango de años que abarca. Utilizando esta base de datos, en conjunto con los datos SIMCE, medimos el impacto del absentismo docente en el desempeño SIMCE de un estudiante. Para esto planteamos dos modelos teóricos que nos permiten resolver de distinta forma problemas que se presentan al hora de explorar los datos. El primero es un modelo de panel con observaciones de alumnos por asignatura y que incluye efectos fijos para controlar las variaciones entre las asignaturas. El segundo modelo es uno de multinivel mixto, y se basa en que las observaciones de los datos se encuentran agrupadas en distintas jerarquías o “niveles”.

Este trabajo se distribuye de la siguiente manera. En la Sección 2 se discute acerca de la motivación y relevancia para estudiar el absentismo, además de discutir literatura previa sobre el tema. En la Sección 3 nos referimos a las fuentes de datos de nuestra investigación. La Sección 4 presenta los modelos teóricos utilizados, mientras que la Sección 5 presenta los resultados de las estimaciones. La Sección 6 concluye.

¹Esta investigación utilizó como fuente de información las bases de datos de la Agencia de Calidad de la Educación. Los autores agradecen a la Agencia de Calidad de la Educación el acceso a la información. Todos los resultados del estudio son de responsabilidad de los autores y en nada comprometen a dicha institución.

2. Contexto

Como se ha discutido bastante en Chile en el último tiempo, la calidad del sistema educativo de un país es esencial para su desarrollo. Existe abundante evidencia de esto en la literatura, tanto del efecto sobre el desarrollo económico de un país (ver Benhabib y Spiegel, 1994; Banerjee y Duflo, 2011) como en sus retornos para las personas (ver Jensen, 2010; Meier y Rauch, 2005, capítulo 5). Con respecto a este último punto existe abundante literatura —la llamada “función de producción de la educación”— que busca establecer la relación entre el desempeño escolar y los ingresos futuros de la persona.

El absentismo docente corresponde a la cantidad de días que el profesor no asiste a la escuela, por cualquier motivo —sea este justificado o no. Usamos esta definición ya que el objetivo es medir el impacto del absentismo como fuente de discontinuidad en la educación impartida, discontinuidad que se produce independientemente de los motivos detrás de la inasistencia.

El absentismo docente es un problema porque tiene un impacto negativo en el desempeño de los estudiantes a través de una serie de mecanismos (ver Ehrenberg et al., 1991; Duflo et al., 2011; Miller, 2012). Esto ocurriría porque (i) existe menor exposición directa a los profesores y por lo tanto menor cantidad de horas de clase al año, (ii) se produce una discontinuidad en la enseñanza y (iii) el absentismo de los profesores crea un mal ejemplo, desmotivando la responsabilidad de los alumnos.

Si bien una correlación negativa entre absentismo y desempeño escolar es esperable y está documentada hace décadas en la literatura (por ejemplo Beavers, 1981; Woods, 1990; Cantrell, 2003), fue el trabajo de Duflo y Hanna (2005) el primero que aportó evidencia contundente de una relación *causal* negativa. Dicha investigación aleatorizó la entrega de un incentivo financiero a profesores con buenas tasas de asistencia, encontrando que el grupo de tratamiento tenía resultados de desempeño 0,17 puntos de desviación estándar superior al de control.

Investigaciones posteriores han expandido la literatura sobre el tema. De Miller et al. (2008), el efecto del absentismo tiene un impacto negativo y significativo sobre el desempeño en matemáticas. En particular, los autores encuentran que 10 días de absentismo anuales disminuyen el desempeño de alumnos de colegios urbanos en 3,3 puntos de desviación estándar. Clotfelter et al. (2009) encuentran un resultado similar aunque ligeramente menor, usando una muestra con estudiantes del sector urbano y el rural.

Clotfelter et al. (2009) encuentran que las tasas promedio de absentismo esconden una alta heterogeneidad, donde regiones rurales o de difícil acceso tienen tasas mucho más altas que el promedio país. Dicha heterogeneidad podría ser una de las muchas fuentes de la

desigualdad de un sistema educativo, al ser un problema mucho más grave en regiones rurales o aisladas. Así, si tomamos la India, se observa que para la misma muestra de la Tabla 1 existen regiones con una tasa de absentismo de hasta el 50 %. Algo similar ocurre en Bangladesh y Ecuador (Chaudhury et al., 2006). Incluso en EE.UU., donde la tasa promedio es 5 %, existen ciudades donde la tasa llega hasta a ser 40 % (Miller, 2012).

Tabla 1: Absentismo docente por países

País	Tasa (%)
Bangladesh	16
Chile	5
Ecuador	14
Estados Unidos	5
India	25
Indonesia	19
Peru	11
Uganda	27

El dato de la tasa chilena corresponde al de nuestra propia base. La tasa de EE.UU. es de (Miller, 2012). El resto de las tasas son de Chaudhury et al. (2006)

Nuestros datos indican que existe una disparidad estadísticamente significativa, aunque no demasiado grande, entre el absentismo urbano vs. el rural. Si se descompone por asignatura resulta que en matemáticas la diferencia es significativa, pero en lectura no lo es. Información más detallada se reporta en la Tabla 2.

Tabla 2: Días de absentismo promedio por asignatura, urbano vs. rural

	Asignatura		
	Matemática	Lectura	Total
Urbano	8,7	9,4	9,0
	(11,2)	(10,9)	(10,9)
Obs.	79	74	153
Rural	12,2	14,1	13,1
	(16,5)	(21,7)	(18,7)
Obs.	12	11,0	23
Total	9,1	10	9,6
	(11,9)	(12,6)	(12,2)
Obs.	91	85	176

DE entre paréntesis.

No existe consenso aún acerca de cómo enfrentar el absentismo docente. Si tomamos

como variable de interés la tasa de absentismo promedio de los profesores, la política de incentivos monetarios analizada en Duflo y Hanna (2005) es recomendable desde el punto de vista de la eficiencia. Los autores encuentran que usando muy pocos recursos (monetarios y administrativos) es posible reducir enormemente la tasa de absentismo, mejorando consecuentemente el desempeño escolar.

Sin embargo hay que considerar que un sistema de incentivos basado en la presencia de los profesores podría no tener el efecto esperado en la asistencia. Puede ocurrir que los profesores con metas de salario mensual simplemente decidieran parar de trabajar luego de alcanzar dicha meta, como muestra la evidencia empírica del trabajo de Fehr y Goette (2007) sobre la elasticidad intertemporal de la oferta laboral. Por otro lado, cabría pensar que el simple hecho de poner incentivos puede tener un efecto de *crowding out* de la motivación intrínseca de enseñar de los profesores, argumento que usualmente es desarrollado bajo el marco de un modelo que combina heterogeneidad de altruismo y codicia individual con preocupaciones de reputación social. Siguiendo la misma línea pero en un marco más general, es relevante considerar los argumentos de monetarización de actividades que no deberían estarlo, desarrollado por Sandel (2012). También se da una discusión interesante del alcance del argumento de Sandel en Besley (2013).

Además, no es claro que aunque la asistencia aumente esto implique un mayor aprendizaje de los niños. Por lo tanto, que la reducción del absentismo docente tenga un impacto positivo en la calidad escolar es una pregunta empírica y la literatura es aún muy reciente para sacar conclusiones de nivel general.

Esto nos lleva a analizar el caso chileno donde, desafortunadamente, no es fácil encontrar datos sobre la tasa de absentismo de los profesores. Si bien hay estudios del sistema educativo chileno que hacen referencia a su importancia (por ejemplo en Eisenberg, 2008), pareciera no haber una cuantificación formal de dicha tasa. Más allá del estudio citado y algunos otros documentos técnicos, no parece haber en Chile ninguna base de datos centralizada con datos de asistencia docente. Esto podría deberse, en el mejor de los casos, a un desconocimiento de la autoridad de la importancia de medirla, o, en el peor de los casos, a una omisión deliberada, ya que los profesores serían una fuerza política poderosa y capaz de resistir intentos de un control estricto de su asistencia (Banerjee y Duflo, 2006; Duflo et al., 2012).

Nuestra investigación aporta a la literatura existente porque mide el impacto del absentismo en colegios públicos usando una rica base de pruebas estandarizadas, proveniente de los resultados del SIMCE. Además, utiliza datos del absentismo docente proveídos por los registros de las Corporaciones Municipales, empleadores de los profesores, por lo que se cuenta con una base de datos única para observar esta variable. Explicamos estas fuentes de datos en detalle a continuación.

3. Datos

Esta investigación utilizó dos fuentes de datos. La primera corresponde al SIMCE, disponible públicamente y compuesta de varias bases individuales de información. Nos referimos en detalle a éstas en la sección 3.1. La segunda fuente corresponde a una base de datos de elaboración propia, construida a partir de información solicitada a Municipalidades del país y cuyo principal objetivo es proveer información del absentismo docente. Nos referimos en detalle a ésta en la sección 3.2.

3.1. Datos de SIMCE

Todos los datos excepto a los correspondientes al absentismo docente provienen de la bases de datos SIMCE (Sistema de Medición de la Calidad de la Educación), la cuales incluyen resultados estandarizados tanto para Matemática como para Lenguaje y Comunicación (en adelante, “Lectura”²) a nivel de estudiante. Las bases son proveídas por la Agencia de la Calidad de la Educación³ y corresponden a los segundos medios (en adelante, “2M”) de los años 2010, 2012 y 2013.⁴

Los datos en sí mismos están distribuidos en diferentes bases. Para cada año disponible hay bases con características de los estudiantes, como también del establecimiento educacional, padres, y características de los profesores. También hay datos de características a nivel comunal, provincial y regional. Para cada año se unieron estas bases usando un identificador transversal único.

Nuestra mayor preocupación fue encontrar inconsistencias internas con RBD, el cual es un identificador único para las colegios chilenos. Eliminar duplicados a nivel de estudiante es muy difícil⁵, pero el RBD usa una variación del algoritmo de Luhn para generar un número de validación⁶. Esto es incluido como una variable distintiva en nuestra base, y para corroborar los datos, generamos manualmente el dígito verificador, contrastándolo con el de la base. La cantidad de observaciones que no pasaron la prueba de validación es insignificante y fueron eliminadas.

²Así es como se denomina a la prueba SIMCE de Lenguaje y Comunicación.

³Organismo responsable de evaluar logros de aprendizaje y proporcionar información a la comunidad, desde el año 2012.

⁴El año 2011 no se realizó prueba SIMCE a segundo medio.

⁵Cada estudiante de la base está identificado con un número único a lo largo de los años, el MRUN, que corresponde a un RUT falso (para proteger su identidad). Este MRUN no tiene dígito verificador en nuestros datos, por lo que resulta imposible distinguir si dos observaciones con el mismo MRUN corresponden a un alumno que ha repetido de curso o se trata de un error de tipeo.

⁶Específicamente, es un algoritmo módulo 11. Ver Luhn (1960).

3.2. Datos Absentismo Docente

Dada la inexistencia de una fuente de datos estandarizados confiable para la variable absentismo docente, se procedió a construir una base de datos a partir de información entregada por Municipalidades, sostenedores de los establecimientos educacionales públicos. Para ello, nos apoyamos en la Ley de Transparencia y Acceso a la Información Pública (Ley 20.285), solicitando a través del Portal de Transparencia los datos a dichos organismos. El Portal de Transparencia⁷ permite realizar solicitudes de información de toda información que sea elaborada con presupuesto público o que esté en poder de los organismos regulados. Estos son: El Consejo para la Transparencia, Empresas Públicas, Fundaciones, Ministerios, Municipalidades, Organismos Autónomos, Presidencia de la República y Universidades. Como esta ley es válida solo para organismos públicos y no pudiendo acceder a información de establecimientos particulares subvencionados y particulares pagados⁸, nuestra investigación se acotó a los establecimientos públicos.

Para hacer uso de esta herramienta solo es necesario identificar al solicitante, al organismo al cual se le pide la información, señalar el formato en que se requiere la información y detallar exhaustivamente qué información se solicita. En nuestro caso particular, las oficinas de transparencia de las municipalidades tuvieron que derivar la solicitud a las Corporaciones Educacionales Municipales, quienes son los encargados de administrar la educación municipal.⁹

Para efectos de esta investigación, se solicitó el número de días de inasistencias o, en su defecto, porcentaje de asistencia de los docentes de matemática y lenguaje de segundo medio para los años 2010, 2012 y 2013. Esta información, para el caso chileno, aún no se registra de manera digital. La información debía contener además algún tipo de identificación del profesor (nombre/RUT), cursos a los que hace clases y RBD del establecimiento educacional, con el objeto de poder cruzar estos datos con aquellos de la base SIMCE. Adjunta a la solicitud, se agregó un archivo Excel tipo para facilitar y estandarizar el ingreso de datos. En total, se levantaron 133 solicitudes, puesto que no todas las Municipalidades están registradas en el Portal de Transparencia.

De estas 133 solicitudes aproximadamente 90 respondieron con algún tipo de dato

⁷<http://www.portaltransparencia.cl/PortalPdT/>

⁸En la búsqueda preliminar de los datos de absentismo se exploró la alternativa de solicitar los datos a fundaciones sostenedoras de colegios vulnerables subvencionados, sin embargo la petición no pudo concretarse por la inexistencia de los datos o simple rechazo.

⁹Cuando esto sucede, las “observaciones adicionales” a la solicitud, y los detalles del formato, en conjunto con los datos adjuntos a la solicitud, pueden no ser derivadas, por lo que se recomienda al investigador especificar todas las características relevantes de los datos y la entrega de estos en la solicitud de información misma, no obviando ninguna de ellas y siendo lo más claro posible. De esta forma, de ser derivada la solicitud a otra oficina dependiente del organismo al que se le solicitó la información, esta incluirá los aspectos relevantes de los datos y el formato de entrega, lo que ayuda a la construcción de la base final.

adjunto, mientras que las demás apelaron a un artículo de ley¹⁰ que los excusa de remitir los datos. Cabe destacar que durante el proceso de recopilación de los datos se produjo un paro docente en gran parte del país, agregando mayor dificultad a la entrega de la información. Por otro lado, en febrero del 2010, varios establecimientos educacionales fueron destruidos por motivo del terremoto que afectó la zona centro-sur del país, lo que habría producido una pérdida de registros según municipalidades del sector. Adicionalmente, varias solicitudes fueron respondidas pero omitiendo información importante para enlazarla con los datos SIMCE, como son el curso, el RBD o la asignatura que dicta el docente.

A pesar de estas contrariedades, se lograron recopilar 529 observaciones de profesores al número de inasistencias anuales y 362 observaciones correspondientes al porcentaje de asistencia anual. Ambas mediciones son a nivel profesor. Procedimos a convertir el porcentaje de asistencia a número de inasistencias tomando en cuenta el total de días lectivos de los años considerados, recopilando un total de 691 observaciones a de absentismo docente.¹¹ Estas observaciones pueden ser presentadas en distintos niveles. A nivel de curso, existen 427 observaciones. A nivel de profesor, se cuenta con datos de 181 profesores los cuales están presentes en al menos 2 observaciones a nivel curso, donde 89 son de lenguaje y los restantes 92 de matemática. Finalmente a nivel de establecimiento educacional, la base contiene datos de 56 RBD distintos.

Una ventaja de esta base frente a, por ejemplo, la auto-reportada que se encuentra en el cuestionario al profesor en el SIMCE, es que fue construida por información reportada por la Corporación Municipal de cada comuna, organismo empleador de los profesores del sistema público, por lo que podemos suponer que los datos son veraces. Se debe recordar que los profesores pueden ser multados por inasistencias injustificadas, siendo estas controladas al igual que las licencias médicas y permisos administrativos.

En estricto rigor, la variable de interés, más que número de inasistencia o porcentaje de asistencia, debiera ser la cantidad de horas no realizadas respecto al total de horas lectivas de la asignatura para cada año. Sin embargo, ya que los registros de asistencia no contienen tanto detalle, la recopilación de este dato resulta extremadamente difícil. A pesar de esto, creemos que la ausencia diaria por parte del profesor es una aproximación razonable de las horas no realizadas, independiente de la heterogeneidad de horarios de clase que pueda existir, porque las asignaturas consideradas se imparten casi todos los días de clases.¹²

¹⁰ Artículo 21 n°1, letra «c» de la ley 20.285: “La obtención de dicha información es de aquellas de carácter genérico, cuya atención requiere distraer indebidamente a los funcionarios del cumplimiento regular de sus labores habituales”.

¹¹ Esta ocurre porque hubo solicitudes que entregaron información tanto del número de inasistencias como del porcentaje de asistencia.

¹² En estricto rigor el problema es muy complejo, ya que requeriría saber qué días faltó el profesor y el alumno, además de conocer los días en que el profesor imparte la materia al alumno. De esta manera podríamos generar un *matching* perfecto, pero un estudio así de detallado requiere datos que no es posible conseguir a través de las municipalidades.

3.3. Construcción de la base de datos

Las bases SIMCE combinadas para 2M en los años 2010, 2012 y 2013 corresponden a 474.517 observaciones. Al unirla a la base de asistencia de profesores conservamos 8.748 alumnos que fueron emparejados con algún profesor con datos de absentismo en alguna de las dos asignaturas. Luego eliminamos a aquellos alumnos que no poseían resultados SIMCE en *ninguna* de las dos pruebas (312 observaciones). Por construcción de la base de absentismo docente, todos los alumnos tenían datos del absentismo de al menos uno de sus dos profesores relevantes. Finalmente, de las 474.517 observaciones originales disponibles, nuestra base de datos consiste en 8.436 observaciones, es decir, cercano a una muestra del 1,8%.

Tabla 3: Comparación de estadísticas para variables seleccionadas.

	Base SIMCE	Muestra
Ptje. Matemáticas	255,2 (63,3)	239,3 (56,2)
Obs.	460.838	8.364
Ptje. Lectura	250,3 (53,1)	244,8 (50,7)
Obs.	460.728	8.374
Educ. Madre	11,3 (3,4)	10 (3,3)
Obs.	461.649	8.126

DE en paréntesis.

3.4. Estadística Descriptiva

Los promedios de días de inasistencia desagregados por asignatura se encuentran especificados en la Tabla 4. De esta podemos sacar dos conclusiones. No detectamos una diferencia estadísticamente significativa entre la inasistencia de los profesores de lenguaje y matemática.

Un dato interesante que corroboramos con nuestros datos es el de la imprecisión que tiene la medición del absentismo, registrado como una pregunta del cuestionario SIMCE. Dado que dicho absentismo es auto-reportado, cabría esperar que no fuera fidedigno, porque (i) no corresponde a una medición sistemática y se confía en la memoria del docente y (ii) existe una tendencia documentada a sobreportar asistencia en el ambiente laboral. El efecto conjunto de ambos mecanismos indicaría que el absentismo podría estar subreportado en el cuestionario SIMCE.

Tabla 4: Inasistencias de profesores por tramos y asignaturas

Días ausente	Asignatura					
	Lectura		Matemática		Total	
	No.	%	No.	%	No.	%
0	42	21	50	23	92	22
1 a 5	53	26	73	34	126	30
6 a 10	39	20	31	14	70	17
11 a 15	30	15	25	12	55	13
16 a 20	8	4	11	5	19	5
21 o más	28	14	26	12	54	13
Total	200	100	216	100	416	100

Tabla 5: Inasistencias auto-reportadas vs. registros municipales

Días ausente	Auto-reportado		Registro municipal		Diferencia
	No.	%	No.	%	
0	23	20.9	32	29.1	-8.2
1 a 4	46	41.8	27	24.6	17.2
5 a 15	26	23.6	36	32.7	-9.1
16 a 30	10	9.1	8	7.3	1.8
21 a 60	2	1.8	3	2.7	-0.9
60 o mas	3	2.7	4	3.6	-0.9
Total	110	100	110	100	

Se consideraron solamente aquellos profesores para los que se tenían datos tanto auto-reportados (del cuestionario SIMCE) y de los registros municipales. El primer criterio por sí solo acotó la muestra solamente a profesores del 2010.

La pregunta para el auto-reporte de días de absentismo solo se encuentra en las bases del año 2010, por lo que solo utilizamos una fracción de nuestra base de datos, correspondiente a 3.101 observaciones. Se comparan los datos de absentismo según registros municipales y auto-reportados en la Tabla 5.

4. Marco teórico

Para identificar correctamente el efecto de los días de ausencia de un profesor sobre el desempeño SIMCE el principal desafío es poder controlar por características no observables que son relevantes para determinar el puntaje del estudiante. Las principales son la habilidad no observada del alumno y la calidad del docente. Desarrollamos a continuación dos modelos que nos permiten lidiar con estas dificultades.

4.1. Modelo de efecto fijo

Siguiendo el trabajo de Paredes (2014), consideramos un panel de alumnos por asignatura controlando por las variaciones de cada alumno entre las asignaturas. Los datos se estructuran de manera que el indicador j representa la asignatura del puntaje obtenido por el alumno k , por lo que tenemos un panel “corto” ($j = 1, 2$) con la particularidad de que la variable “tiempo” es la asignatura.

Los estimadores del modelo planteado en esta sección se basan en el supuesto de que el error idiosincrásico ε se distribuya $\varepsilon \sim (0, \sigma_\varepsilon^2)$. Si bien este supuesto no necesariamente se cumple para nuestros datos, sabemos que los estimadores seguirán siendo consistentes¹³ pero los errores serán incorrectos. Por lo tanto todos los modelos son estimados usando errores estándar robustos de grupo.

Finalmente se incluyen efectos fijos, cuyo uso descansa en el supuesto que los regresores y el término de error pueden presentar cierta correlación, supuesto que exploramos a continuación. Tenemos un modelo de la forma:

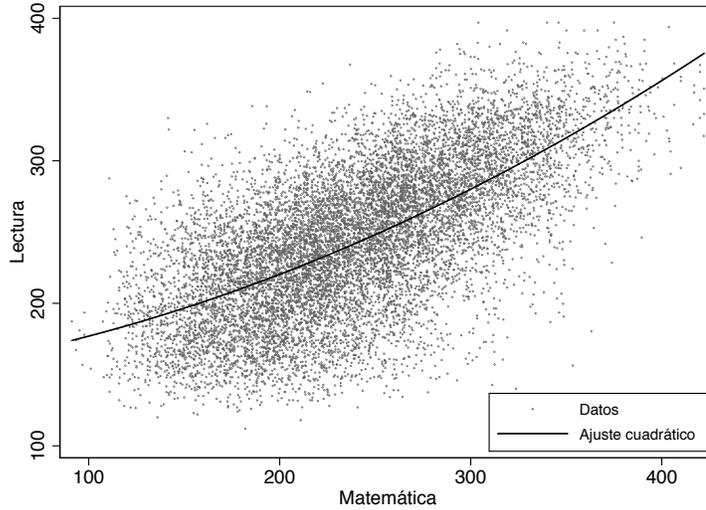
$$y_{kj} = \alpha_k + x'_{kj}\beta + \varepsilon_{kj} \quad [1]$$

donde y_{kj} es el puntaje SIMCE de la prueba j ($j = 1, 2$) para el estudiante k , x_{kj} es un vector de variables independientes, α_k son efectos aleatorios específicos a cada estudiante k y ε_{kj} es un término de error idiosincrásico. La variable de interés es “días de absentismo” y se encuentra contenida en x_{kj} .

¹³Siempre que se cumpla que $E(\varepsilon_{kj}\varepsilon_{iy}) = 0$ para $k \neq i$, $E(\varepsilon_{kj}\varepsilon_{ky})$ no esté restringido y se permite que ε_{kj} sea heterocedástico.

El término de error puede ser interpretado como $\mu_{kj} = \alpha_k + \varepsilon_{kj}$, permitiéndose que x_{kj} correlacione con la parte invariante en j de μ_{kj} , es decir, con α_k . Intuitivamente esto significa que si los regresores de la estimación de puntaje SIMCE están correlacionados con, por ejemplo, “habilidad académica” no observada, estos están correlacionados solamente con el componente de habilidad que es invariante entre asignaturas.¹⁴ El supuesto anterior puede analizarse gráficamente observando la correlación entre el puntaje de lectura y matemáticas por alumno, la que se encuentra en la Figura 1.

Figura 1: Correlación entre puntajes de matemática y lectura por alumno.



Para estimar β deben eliminarse los efectos fijos de α_k . Stata consigue esto realizando MCO sobre los datos diferenciados en la media, de manera que tiene que haber variación de los datos en j . Esto se cumple fácilmente en nuestros datos, dada la gran cantidad de puntajes posibles.

Los efectos fijos α_k en [1] pueden ser eliminados restando el modelo de medias individuales $\bar{y}_k = \bar{x}'\beta + \varepsilon_k$, lo que resulta en el modelo de media-diferenciada:

$$(y_{kj} - \bar{y}_k) = (x_{kj} - \bar{x}_k)'\beta + (\varepsilon_{kj} - \bar{\varepsilon}_k) \quad [2]$$

Al eliminar α_k , la estimación de β por MCO será consistente a pesar de la correlación de α_k con x_{kj} . En particular, esta estimación será consistente con regresores endógenos siempre que x_{kj} esté correlacionado solamente con la parte invariante en la asignatura j del error.¹⁵ Una desventaja de este método es que no permite estimar coeficientes de un regresor

¹⁴Si bien este supuesto relativamente fuerte, es más débil que el de un modelo de efecto aleatorio, donde se asume que α_i no correlaciona de ninguna forma con los regresores.

¹⁵Es decir, con α_k y no con ε_{kj} .

invariante en j (por ejemplo, género). Incluso para los regresores que sí varían en j , como es nuestra variable de interés, el estimador recién planteado será relativamente impreciso si ésta varía poco.

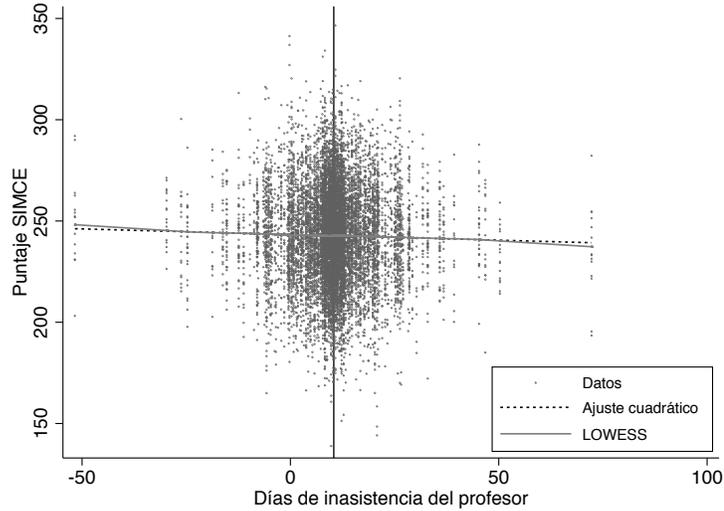
Para controlar por la calidad del docente se utilizan *dummies* por profesor. Al incluir estos efectos fijos controlamos por las diferencias promedio entre profesores en características observables y no observables, tales como diferencias en calidad de la enseñanza, años de escolaridad, títulos, etc., reduciendo el riesgo de sesgo por variable omitida relevante. Análogamente, también planteamos especificaciones con efecto fijo por asignatura, colegio, profesor y/o curso.

Finalmente, el modelo general a estimar es:

$$(y_{kj} - \bar{y}_k + \bar{y}) = \alpha + (x_{kj} - \bar{x}_k + \bar{x}) + (\varepsilon_{kj} - \bar{\varepsilon}_k + \bar{\varepsilon}) \quad [3]$$

donde $\bar{z} = (1/N)\bar{z}_i$ es el promedio de z_{kj} , con z una variable cualquiera. Esta forma funcional tiene la ventaja de entregar estimadores para el intercepto y para el promedio de los efectos individuales α_i , al tiempo que conserva la misma pendiente β del modelo planteado en [2].

Figura 2: Variación intra-grupo de desviaciones del puntaje SIMCE con respecto a la media individual contra desviaciones de días de inasistencia del profesor con respecto a la media individual.



Graficamos la desviación del puntaje para cada estudiante en la Figura 2. Al implementar efecto fijo se crean desviaciones de la media, de manera que $(y_{kj} - \bar{y}_k + \bar{y})$ es graficado contra $(x_{kj} - \bar{x}_k + \bar{x})$, lo que explica los valores negativos en el eje de las abscisas (ver Tabla A.3). Ésta última cantidad se encuentra centrada en $\bar{x} = 10,4$, que corresponde al promedio de días de absentismo en nuestros datos.

La línea punteada es una estimación MCO con ajuste cuadrático del puntaje SIMCE predicho por los días de inasistencia. Si bien la pendiente no es pronunciada (lo que anticipa que el efecto no es grande), ésta es claramente negativa. La línea gris es una regresión no paramétrica del mismo modelo, la cual tiene una pendiente ligeramente más pronunciada.

Por último, consideramos una variación de [1] que permita al intercepto variar tanto para los alumnos k como por asignatura j . De esta forma, la ecuación a estimar es

$$y_{kj} = \alpha_k + \gamma_j + x'_{kj}\beta + \varepsilon_{kj} \quad [4]$$

donde γ_j es el componente adicional del intercepto que varía por asignatura. Sin embargo, si se asume que los efectos de la asignatura γ_j son fijos podemos simplemente incluir una *dummy* para asignatura en x_{kj} y así volver a la forma funcional planteada en [1].

4.2. Modelo multinivel mixto

Las unidades de observación de nuestros datos —los alumnos— se agrupan en distintas categorías anidadas o “niveles”. Cada alumno es parte de un curso, el que a su vez es parte de un colegio. Análogamente, puede considerarse que cada alumno se encuentra bajo la tutela de un profesor (para una asignatura dada), el que también es parte de un colegio.¹⁶ Es natural pensar que los resultados dentro de un mismo nivel estén correlacionados. Por ejemplo, cabe pensar que los resultados dentro de un mismo colegio estén correlacionados ya que comparten la misma infraestructura, insumos y, quizás más importante aún, los alumnos ya se diferencian por el hecho de estar en ese colegio y no otro (sea por selección y/o autoselección).

La notación usualmente usada en la literatura de modelos multiniveles considera a las variables más anidadas dentro de la jerarquía como variables de “nivel 1” (Rabe-Hesketh y Skrondal, 2012). En nuestros datos el nivel 1 corresponde al estudiante i . En este trabajo analizamos una jerarquía donde sólo hay alumnos anidados en colegio (obviando la agrupación por curso o profesor¹⁷). Esto nos permite tomar en cuenta las posibles diferencias existentes entre colegios, que consideramos que son la unidad de agrupación clave. Por lo tanto, el nivel 2 corresponde al colegio k .

Para construir este modelo comenzamos tomando la ecuación

$$y_{ik} = \beta + \xi_{ik}$$

¹⁶En nuestros datos no existen profesores inter-colegios, esto es, que hagan clases en más de un colegio de la muestra.

¹⁷Obviamos dicha agrupación en el planteamiento de los niveles del modelo, pero no de los errores.

donde y_{ik} es puntaje SIMCE, β es un intercepto que corresponde a la media poblacional y ξ_{ik} son las desviaciones individuales con respecto de esta media.

Pensando en la agrupación jerárquica de los datos, es poco razonable suponer que las desviaciones con respecto de la media ξ_{ik} no están correlacionadas con y_{ik} dentro del mismo colegio. Esta dependencia intra-colegio puede modelarse descomponiendo el residuo en dos partes no correlacionadas: un componente permanente ζ_k específico a cada colegio y un componente idiosincrásico ϵ_{ik} , de forma que tenemos

$$y_{ik} = \beta + \zeta_k + \epsilon_{ik}$$

El componente ζ_k es una desviación aleatoria del promedio del colegio k con respecto al promedio total β . En la literatura este componente a veces es llamado “efecto aleatorio”, si bien este *no* es el mismo efecto aleatorio de un modelo de panel. Para evitar confusiones preferimos llamarlo “intercepto aleatorio”, otro término que es común en la literatura. Este término tiene media poblacional cero y no presenta correlación a través de los colegios. Puede ser interpretado entonces como las características que diferencian a los colegios y que no son incluidas como variables en el modelo.

El componente ϵ_{ik} usualmente es llamado el “residuo de nivel 1” o “residuo intra-colegio” y es la desviación aleatoria de y_{ik} del promedio del colegio k . Este residuo tiene media poblacional cero y no presenta correlación a través de los estudiantes ni a través de los colegios. El intercepto aleatorio ζ_k tiene varianza ψ , que puede interpretarse como la varianza inter-colegio, al que podemos referirnos también como “residuo nivel 2”. El residuo ϵ_{ik} tiene varianza constante θ , que puede interpretarse como la varianza intra-colegio.

Ahora, para incorporar variables explicativas al modelo basta considerar una regresión lineal simple y aplicarle la misma descomposición del residuo, es decir, usar que $\xi_{ik} \equiv \zeta_k + \epsilon_{ik}$. Haciendo esto, la ecuación a estimar es

$$\begin{aligned} y_{ik} &= \beta_1 + \beta_2 x_{2ik} + \cdots + \beta_p x_{pik} + \zeta_k + \epsilon_{ik} \\ &= (\beta_1 + \zeta_k) + \sum_{p=2}^P \beta_p x_{pik} + \epsilon_{ik} \end{aligned} \quad [5]$$

donde x_{2ik} hasta x_{pik} son variables explicativas.

Este modelo puede ser interpretado como uno de regresión con un residuo adicional ζ_k de nivel 2 o con un intercepto $(\beta_1 + \zeta_k)$ específico para cada colegio. Preferiremos esta última interpretación, donde ζ_k representa la desviación del intercepto del colegio k con respecto al

intercepto promedio β_1 . Este residuo nivel 2 es un componente del error específico a cada colegio y se asume que no está correlacionado entre colegios, mientras que el residuo nivel 2 ϵ_{ik} es específico a cada estudiante y se asume que no está correlacionado entre colegios ni entre estudiantes. Finalmente, se asume que ambos residuos no correlacionan entre ellos.

En este punto cabe preguntarse si efectivamente es necesario un modelo multinivel para explicar los datos. Para esto consideramos la hipótesis nula de que la varianza inter-grupo es cero:

$$H_0 : \psi = 0$$

$$H_a : \psi > 0$$

Esta nula es equivalente a la hipótesis de que $\zeta_k = 0$, es decir, que no hay intercepto aleatorio en el modelo. De no rechazarse la nula, el modelo de intercepto (y, en consecuencia, de coeficiente) aleatorio no sería necesario. El test de máxima verosimilitud para verificar la hipótesis nula es

$$L = 2(l_1 - l_0)$$

donde l_1 es el logaritmo de verosimilitud maximizada del modelo de intercepto aleatorio (que incluye ζ_k) y l_0 es el logaritmo de de verosimilitud maximizada de un modelo de regresión común (sin ζ_k). El p -value correcto se obtiene al dividir $\chi^2(1)$ por 2. Todas las estimaciones presentadas en la Subsección 5.2 rechazan esta nula. Para más detalles ver el apéndice C.

El modelo planteado en [5] es un modelo multinivel lineal de *intercepto* aleatorio, donde se permite que el nivel general de respuesta varíe entre colegios, controlando por ciertos regresores. En otras palabras, los colegios pueden tener distintos interceptos, pero se asume que sus pendientes son iguales. Ahora relajamos ese supuesto, planteando un modelo multinivel de *coeficientes* aleatorios. En términos de nuestra variable de interés, construimos una especificación que incorpore las relaciones entre días de absentismo docente y desempeño SIMCE para todos los colegios de la muestra.

Tomando [5] como punto de partida y restringiéndonos solamente a una variable explicativa, un método parsimonioso de escribir la relación entre puntaje SIMCE y días de absentismo es especificar un intercepto aleatorio ζ_{1k} específico para cada colegio y un coeficiente aleatorio ζ_{2k} para la variable de interés (x_{ik}):

$$y_{ik} = (\beta_1 + \zeta_{1k}) + (\beta_2 + \zeta_{2k})x_{ik} + \epsilon_{ik} \tag{6}$$

donde ζ_{1k} se interpreta igual que ζ_k en [5] y ζ_{2k} representa la desviación del coeficiente del colegio k del coeficiente (o pendiente) promedio β_2 .

Para evaluar si efectivamente es necesario incluir interceptos aleatorios se testea la siguiente hipótesis:

$$H_0 : \psi_{22} = 0$$

$$H_a : \psi_{22} > 0$$

donde H_0 es equivalente a la hipótesis que todos los coeficientes aleatorios son 0 e implica que $\psi_{21} = 0$, porque una variable que no varía —además de ser un oxímoron impresionante— tampoco covaría con otras variables.

5. Estimaciones

5.1. Modelo de efecto fijo

Para poder identificar con precisión los coeficientes de un modelo de efecto fijo es necesario que exista variación intra-grupo en el panel. Reportamos la variación intra-grupo existente en nuestros datos en la Tabla A.3. Además, los errores estándar de este modelo requieren que ε_{kj} sea i.i.d., por lo que calculamos errores robustos a nivel agrupado. Ahora, esto requiere que las observaciones sean independientes en k y que $N \rightarrow \infty$.

La Tabla 6 presenta las estimaciones de la variable de interés, controlando por distintos efectos fijos. Dada la naturaleza del panel que construimos no es posible estimar directamente los puntajes de Matemática y Lectura por separado, por lo que la variable dependiente es simplemente el puntaje SIMCE del alumno i , controlando por asignatura.

Todas las columnas de la Tabla 6 fueron estimadas con efecto fijo de alumno. Desde la columna (2) agregamos efectos fijos adicionales, los que se detallan en la parte inferior de dicha tabla. Cabe destacar que la inclusión de efecto fijo por profesor aumenta significativamente la desviación estándar del coeficiente de la variable de interés, lo que se debe a la alta correlación existente entre ambas variables.

Los coeficientes estimados indican que los días de absentismo docente tienen un impacto negativo y significativo sobre el desempeño SIMCE de los alumnos, controlando por una gran variedad de efectos fijos. Esto es consistente con la intuición de que es perjudicial para el aprendizaje de los alumnos que su profesor falte.

Tabla 6: Resultados del modelo de efecto fijo (panel de asignaturas) usando como único control la variable de interés y con distintos efectos fijos.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Controles							
Días	-0,096*** (0,034)	-0,15*** (0,034)	-0,15*** (0,034)	-0,15*** (0,034)	-0,12** (0,053)	-0,15*** (0,034)	-0,12** (0,053)
Constante	243,6*** (0,34)	247,3*** (0,52)	247,3*** (0,52)	247,3*** (0,52)	244,1*** (8,29)	247,3*** (0,52)	244,1*** (8,29)
Efectos Fijos							
Alumno	✓	✓	✓	✓	✓	✓	✓
Asignatura		✓	✓	✓	✓	✓	✓
Colegio				✓	✓	✓	✓
Profesor					✓		✓
Curso						✓	✓

Errores robustos agrupados por alumno en paréntesis.

* p<0.10, ** p<0.05, *** p<0.01

En la Tabla 7 se presentan coeficientes estimados para distintas especificaciones del modelo, controlando por efecto fijo de alumno, asignatura, colegio y curso. Por tanto, la columna (1) de la Tabla 7 es equivalente a la columna (6) de la Tabla 6.

Exploramos la posibilidad de que la variable de interés afecte de manera no lineal al desempeño. Para esto incluimos en la regresión la variable de interés al cuadrado en la columna (2). Un coeficiente positivo significaría que la relación es decreciente a tasa decreciente, mientras que un coeficiente negativo significaría que la relación es decreciente a tasa creciente. Sin embargo, el componente cuadrático parece no tener ninguna significancia estadística. En la columna (3) comprobamos que la variable de interés no tuviera un efecto distinto por sexo, pero la significancia del coeficiente estimado para la variable interactiva entre ambas es muy baja.

Tabla 7: Resultados del modelo de efecto fijo (panel de asignaturas), errores robustos agrupados por alumno.

	(1)	(2)	(3)	(4)
Controles				
Días	-0,15*** (0,034)	-0,17** (0,067)	-0,17*** (0,043)	-0,16*** (0,039)
Días ²		0,00020 (0,00070)		
Días×Mujer			0,047 (0,069)	
Días×Asignatura				0,034 (0,054)
Efectos Fijos				
Alumno, asignatura, colegio, curso	✓	✓	✓	✓

Errores robustos agrupados por alumno en paréntesis.

* p<0.10, ** p<0.05, *** p<0.01

Si bien no podemos estimar directamente el efecto de la variable de interés en los puntajes de matemáticas y lectura por separado, esto se puede realizar indirectamente por medio de la inclusión de una variable interactiva entre la *dummy* de asignatura y los días de absentismo. En esta especificación “Asignatura” es una variable que vale 0 si es lectura o 1 si es matemática. El coeficiente estimado indica que, bajo el modelo de efecto fijo, el efecto del absentismo docente es el mismo entre ambas asignaturas.

5.2. Modelo multinivel mixto

No obstante el resultado anterior, interesa estimar un modelo donde los coeficientes invariantes por alumno no tengan que ser rescatados indirectamente. En parte, esta limitación es la que nos motivó a desarrollar el modelo multinivel. Los resultados de la especificación con interceptos aleatorios planteada en [5] se presentan en la Tabla 8. Como siempre, la variable a estimar es el puntaje SIMCE del alumno y la variable de interés es “Días”. Esta estimación considera a los alumnos como variables de nivel 1 y a los colegios como variable de nivel 2. Las especificaciones se encuentran separadas por asignatura, pero son simétricas: la columna (1) es equivalente a la (5), y así sucesivamente.

La parte fija de la Tabla 8 incluye estimaciones para los controles x_{pik} . Luego se presenta la estimación para el intercepto aleatorio de nivel 2 ζ_k (IA N2) y su desviación estándar $\sqrt{\psi}$ (colegios). Finalmente se presenta la estimación del error idiosincrásico ϵ_{ik} (error N1) y su desviación estándar $\sqrt{\theta}$ (alumnos).

En este punto podemos hacer la diferencia entre un coeficiente estimado por efecto fijo, el cual solo muestra el efecto intra-grupo, y un coeficiente del modelo de intercepto aleatorio. Tomando como ejemplo la estimación para la variable de interés en la columna (1), tenemos que por cada día de absentismo del profesor, en promedio, el puntaje SIMCE de matemáticas es 0,28 puntos menor, manteniendo todas las otras variables constantes. Esta estimación representa o una comparación entre estudiantes de distintos colegios o una comparación entre niños del mismo colegio. Es decir, no es puramente ni una comparación intra-grupos ni inter-grupos.

Sobre la variable de interés resulta interesante observar que ahora que separamos por asignatura —algo que el modelo anterior solo era posible usando variables interactivas— existe una diferencia relativamente grande entre los coeficientes estimados para matemáticas y lectura. Si antes el coeficiente conjunto era cercano a $-0,15$, ahora parece separarse en coeficientes mayores para matemáticas ($\approx -0,3$) y menores para lectura ($\approx -0,15$). Todos los coeficientes son significativos al 99 % de confianza.¹⁸

La *dummy* que indica si el estudiante es mujer es negativa y significativa para matemáticas, mientras que es igualmente significativa pero positiva para lectura. Este resultado es consistente con lo reportado por Mizala y Romaguera (2002); Flórez Petour (2013); Paredes (2014) y otros. Controlar por año (tomando como base el año 2010) resulta significativo solo para matemáticas, resultado que exploramos en mayor profundidad con la especificación planteada para la Tabla A.4. Finalmente, la educación de la madre medida en años de escolaridad resulta positiva y significativa.

¹⁸Esta significancia está calculada con un estadístico z —no t , como en una regresión lineal estándar. Ver apéndice B.

Tabla 8: Resultados del modelo multinivel mixto con interceptos aleatorios por colegio. Modelos separados por asignatura.

	Matemática				Lectura			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Parte fija								
Días	-0,28*** (0,056)	-0,29*** (0,056)	-0,30*** (0,056)	-0,30*** (0,057)	-0,16*** (0,046)	-0,16*** (0,046)	-0,15*** (0,046)	-0,14*** (0,046)
Año 2012		-0,22 (1,51)	-0,31 (1,50)	-0,20 (1,53)		0,36 (1,38)	0,27 (1,38)	0,69 (1,40)
Año 2013		3,23** (1,59)	3,12** (1,58)	3,49** (1,59)		-0,71 (1,52)	-0,76 (1,51)	-0,71 (1,53)
Mujer			-12,9*** (1,19)	-12,5*** (1,21)			5,74*** (1,13)	5,94*** (1,15)
Educ. madre				1,79*** (0,20)				1,45*** (0,19)
Constante	234,5*** (4,05)	233,6*** (4,15)	253,2*** (4,55)	235,2*** (4,83)	239,3*** (3,71)	239,3*** (3,81)	230,6*** (4,14)	216,0*** (4,46)
IA N2								
Constante	28,7*** (2,86)	28,7*** (2,86)	29,0*** (2,89)	27,3*** (2,75)	25,1*** (2,66)	25,1*** (2,66)	24,9*** (2,64)	23,6*** (2,54)
Error N1								
Constante	47,1*** (0,41)	47,1*** (0,41)	46,7*** (0,41)	46,5*** (0,41)	43,9*** (0,39)	43,9*** (0,39)	43,8*** (0,39)	43,7*** (0,39)
<i>N</i>	6687	6687	6687	6443	6413	6413	6413	6166

Errores robustos en paréntesis.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Las desviaciones estándar estimadas para el intercepto aleatorio son menores a los 3 puntos. Esto quiere decir que los interceptos estimados para los colegios tienen una desviación estándar *estimada* de 2,7 puntos aproximadamente, lo que es bastante bajo. Recordar sin embargo que bajo un entorno multinivel los interceptos ζ_k son tratados como variables aleatorias y no como parámetros del modelo, por lo que las mediciones corresponden a $\hat{\psi}$.

6. Conclusión

El propósito de esta investigación es doble. Por un lado se concentra en analizar el impacto de una variable omitida relevante para la determinación del desempeño SIMCE. En este sentido concluimos que el absentismo docente tiene un impacto estadísticamente significativo y negativo, como cabría esperar, aunque su efecto marginal es relativamente pequeño. Sin embargo, si un docente acumula muchos días de ausencia el efecto total sobre el puntaje de sus alumnos no es despreciable.

Los resultados del modelo de efecto fijo bastante consistentes con los del modelo multinivel, los que indican una disminución aproximada de 0,15 puntos promedio en el SIMCE por día de ausencia del docente. Sin embargo al separar por asignatura obtuvimos resultados contradictorios. El primer modelo indica que la variable de interés tiene el mismo efecto para ambas asignaturas. Por otro lado, el modelo multinivel indica que si bien el impacto negativo del absentismo del profesor de lectura sigue siendo de 0,15 puntos por día, el del profesor de matemáticas es aproximadamente el doble, es decir, 0,3 puntos por día. Este resultado es de por sí interesante y creemos que puede ser de utilidad para colegios y otras instituciones relacionadas con la educación.

Por otro lado, el objetivo último de esta investigación es indagar en una arista de la educación que creemos ha sido dejada de lado en las discusiones, pero que podría ser relevante para determinar cuán exitoso resulta un sistema educativo. Los profesores son un requisito esencial para que se produzca la educación y se sabe que su calidad explica en cierta medida el desempeño de sus alumnos, por lo que se han realizado esfuerzos para medir dicha calidad de manera más exacta, permitiendo determinar más precisamente su impacto en los estudiantes. De la misma forma, creemos que el absentismo docente es un aspecto que debería comenzar a medirse de manera más rigurosa, de modo que futuras investigaciones puedan tomarlo en cuenta como otro determinante del desempeño académico.

Referencias

- Banerjee, A. y E. Duflo (2006). "Addressing Absence". *Journal of Economic Perspectives* 20.1, págs. 117-132.
- Banerjee, A. y E. Duflo (2011). *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty. PublicAffairs.
- Beavers, H. (1981). *The Relationship Between Selected Educational Variables and Student Achievement in a Selected School District*. East Texas State University.
- Benhabib, J. y M. M. Spiegel (1994). "The role of human capital in economic development evidence from aggregate cross-country data". *Journal of Monetary Economics* 34.2, págs. 143-173.
- Besley, T. J. (2013). "What's the Good of the Market? An Essay on Michael Sandel's What Money Can't Buy". *Journal of Economic Literature* 51.2, págs. 478-95.
- Cantrell, S. M. (2003). *Pay and performance: the utility of teacher experience, education, credentials, and attendance as predictors of student achievement at elementary schools in LAUSD*. Inf. téc. Los Angeles Program Evaluation y Research Branch.
- Cascio, W. (2006). *Managing Human Resources: Productivity, Quality of Work Life, Profits*. Team learning assistant. McGraw-Hill/Irwin.
- Chaudhury, N. et al. (2006). "Missing in Action: Teacher and Health Worker Absence in Developing Countries". *Journal of Economic Perspectives* 20.1, págs. 91-116.
- Clotfelter, C. T., H. F. Ladd y J. L. Vigdor (2009). "Are Teacher Absences Worth Worrying About in the United States?" *MIT Press Journals* 4.2, págs. 115-149.
- Cohn, E (1996). "Methods of teacher remuneration: merit pay and career ladders". *Assessing Educational Practices: The Contribution of Economics*.
- Contreras, D. y T. Rau (2012). "Tournament Incentives for Teachers: Evidence from a Scaled-Up Intervention in Chile". *Economic Development and Cultural Change* 61.1, págs. 219-246.
- Dionne, G. y B. Dostie (2005). *New Evidence on the Determinants of Absenteeism Using Linked Employer-Employee Data*. Inf. téc. CIRPEE.

- Duflo, E. y R. Hanna (2005). *Monitoring Works: Getting Teachers to Come to School*. NBER Working Papers 11880. National Bureau of Economic Research, Inc.
- Duflo, E., P. Dupas y M. Kremer (2011). “Peer Effects, Teacher Incentives and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya”. *American Economic Review* 101.5, págs. 1739-1774.
- Duflo, E., R. Hanna y S. P. Ryan (2012). “Incentives Work: Getting Teachers to Come to School”. *American Economic Review* 102.4, págs. 1241-1278.
- Ehrenberg, R. G. et al. (1991). *School District Leave Policies, Teacher Absenteeism, and Student Achievement*. Working Paper 2874. National Bureau of Economic Research.
- Eisenberg, N. (2008). *The Performance of Teachers in Chilean Public Elementary Schools: Exploring Its Relationship with Teacher Backgrounds and Student Achievement, and Its Distribution Across Schools and Municipalities*. University of California, Los Angeles.
- Fehr, E. y L. Goette (2007). “Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment”. *American Economic Review* 97.1, págs. 298-317.
- Flórez Petour, M. T. (2013). *Análisis crítico de la validez del Sistema de Medición de la Calidad de la Educación (SIMCE)*. Inf. téc. 13. Oxford University Centre for Educational Assessment.
- Glewwe, P., N. Ilias y M. Kremer (2003). *Teacher Incentives*. NBER Working Papers 9671. National Bureau of Economic Research, Inc.
- Guerrero, G. et al. (2012). *What works to improve teacher attendance in developing countries? A systematic review*. Inf. téc. EPPI-Centre, Social Science Research unit, Institute of Education, University of London.
- Jensen, R. (2010). “The (Perceived) Returns to Education and the Demand for Schooling”. *The Quarterly Journal of Economics* 125.2, págs. 515-548.
- Kremer, M. (2003). “Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons”. *The American Economic Review* 93.2, págs. 102-106.
- Luhn, H. P. (1960). *Computer for verifying numbers*. US Patent 2,950,048.
- Meier, G. y J. E. Rauch (2005). *Leading Issues in Economic Development*. 8th. OUP Catalogue 9780195179606. Oxford University Press.

- Miller, R. (2012). *Teacher Absence as a Leading Indicator of Student Achievement*. Inf. téc. Center for American Progress.
- Miller, R. T., R. J. Murnane y J. B. Willett (2008). “Do Teacher Absences Impact Student Achievement? Longitudinal Evidence from One Urban School District”. *Educational Evaluation and Policy Analysis* 30.2, págs. 181-200.
- Mizala, A. y P. Romaguera (2002). “Evaluación del Desempeño e Incentivos en la Educación Chilena”. *Latin American Journal of Economics-formerly Cuadernos de Economía* 39.118, págs. 353-394.
- Paredes, V. (2014). “A teacher like me or a student like me? Role model versus teacher bias effect”. *Economics of Education Review* 39.C, págs. 38-49.
- Rabe-Hesketh, S. y A. Skrondal (2012). *Multilevel and Longitudinal Modeling Using Stata, 3rd Edition*. 3rd. StataCorp LP.
- Rivkin, S. G., E. A. Hanushek y J. F. Kain (2005). “Teachers, Schools and Academic Achievement”. *Econometrica* 73.2, págs. 417-458.
- Rockoff, J. E. (2004). “The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data”. *American Economic Review* 94.2, págs. 247-252.
- Sandel, M. (2012). *What Money Can't Buy: The Moral Limits of Markets*. Farrar, Straus y Giroux.
- Tirole, J. y R. Bénabou (2006). “Incentives and Prosocial Behavior”. *American Economic Review* 96.5, págs. 1652-1678.
- Tooley, J. (2012). “Big Questions and Poor Economics: Banerjee and Duflo on Schooling in Developing Countries”. *Econ Journal Watch* 9.3, págs. 170-185.
- Woods, R. C. (1990). *The effect of teacher attendance on student achievement in two selected school districts*. Inf. téc. Ball State University.

Apéndices

A. Tablas

Tabla A.1: Consolidación de la base de datos utilizada

	2010	2012	2013	Total
(1)	191.121	134.725	148.671	474.517
(2)	3.189	2.933	2.626	8.748
(3)	3.101	2.848	2.487	8.436

La fila (1) corresponde a las bases de 2M combinadas para los años 2010, 2012 y 2013. En la fila (2) se combinaron las observaciones SIMCE con la base de absentismo docente. En la fila (3) se eliminaron aquellas observaciones que no tenían puntaje de matemáticas ni de lectura.

Tabla A.2: Inasistencias de profesores por tramos y grupo socioeconómico

Días ausente	Grupo socioeconómico								Total	
	1		2		3		4			
	No.	%	No.	%	No.	%	No.	%	No.	%
0	57	27	16	12	12	26	7	29	92	22
1 a 5	48	22	57	44	13	28	8	33	126	30
6 a 10	42	20	21	16	4	9	3	12	70	17
11 a 15	31	14	15	11	8	17	1	4	55	13
16 a 20	10	5	4	3	3	6	2	8	19	5
21 o más	26	12	18	14	7	15	3	12	54	13
Total	214	100	131	100	47	100	24	100	416	100

La base SIMCE incluye una variable que indica el “Grupo Socioeconómico” (GSE) del colegio. Nunca se explicita en las glosas o en algún otro documento si corresponde al “quintil”, pero cabe pensar que sí, ya que se establece que puede tomar valores de 1 a 5. En todo caso, faltan detalles acerca del criterio para establecer que un colegio pertenece a uno u otro GSE. Finalmente, ningún colegio de nuestros datos tiene un valor GSE = 5, por lo que se excluye esa columna de la tabla.

Tabla A.3: Resumen del panel, variables seleccionadas

Variable	Variación	Promedio	D.E.	Min.	Max.	Obs.
Puntaje	Total	242	54.7	91.2	422.2	N= 22507
	Inter		49.7	120	398.8	n= 11381
	Intra		22.9	138.2	345.8	T= 2
Días	Total	10.4	14.7	0	154	N= 21621
	Inter		12.7	0	92	n= 13815
	Intra		7.7	-51.6	72.4	T= 1.6
Año	Total	2011.6	1.2	2010	2013	N= 27642
	Inter		1.2	2010	2013	n= 13821
	Intra		.1	2010.6	2012.6	T= 2
Profesor	Total	94.3	53.2	1	181	N= 21621
	Inter		45.5	1	181	n= 13815
	Intra		32.6	5.8	182.8	T= 1.6

Para las variables seleccionadas los datos muestran que existe mayor variación inter-grupo que intra-grupo, lo que implicará una pérdida de eficiencia al hacer estimaciones con efecto fijo. La variable días tiene un mínimo negativo de variación intra-grupo porque el mínimo es $x_{kj} - \bar{x}_k + \bar{x}$.

Tabla A.4: Resultados del modelo multinivel mixto con interceptos aleatorios por año y colegio.

	Matemática			Lectura		
	(1)	(2)	(3)	(4)	(5)	(6)
Parte fija						
Días	-0,48*** (0,072)	-0,49*** (0,071)	-0,49*** (0,071)	-0,061 (0,061)	-0,055 (0,060)	-0,054 (0,060)
Mujer		-13,3*** (1,18)	-13,3*** (1,18)		5,59*** (1,12)	5,59*** (1,12)
Año 2012			6,01 (6,99)			0,63 (6,43)
Año 2013			8,17 (6,96)			-0,38 (6,63)
Constante	237,5*** (2,88)	257,7*** (3,41)	252,8*** (5,40)	239,2*** (2,75)	230,7*** (3,21)	230,6*** (4,96)
IA N3						
Constante	0,000000056 (0,00000079)	0,0000012 (0,000021)	0,000067 (0,00097)	0,000068 (0,00092)	0,000067 (0,00091)	0,000068 (0,00092)
IA N2						
Constante	30,5*** (2,04)	30,8*** (2,30)	30,6*** (2,05)	27,9*** (1,97)	27,6*** (1,96)	27,6*** (1,95)
Error N1						
Constante	46,6*** (0,41)	46,1*** (0,43)	46,1*** (0,40)	43,2*** (0,39)	43,1*** (0,39)	43,1*** (0,38)
<i>N</i>	6687	6687	6687	6413	6413	6413

Errores robustos en paréntesis.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Tabla A.5: Resultados del modelo multinivel mixto con interceptos aleatorios por colegio y curso. Modelos separados por asignatura.

	Matemática				Lectura			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Parte fija								
Días	-0,25** (0,11)	-0,25** (0,11)	-0,26** (0,11)	-0,27** (0,11)	-0,16* (0,094)	-0,15 (0,094)	-0,15 (0,094)	-0,14 (0,093)
Año 2012		1,07 (3,16)	0,93 (3,19)	0,82 (3,16)		0,20 (2,90)	0,13 (2,89)	0,17 (2,88)
Año 2013		2,55 (3,26)	2,37 (3,28)	2,73 (3,26)		-4,50 (3,15)	-4,52 (3,13)	-4,59 (3,11)
Mujer			-13,7*** (1,12)	-13,2*** (1,14)			4,50*** (1,07)	4,65*** (1,09)
Educ. madre				1,38*** (0,19)				1,07*** (0,18)
Constante	234,1*** (4,18)	232,9*** (4,57)	253,6*** (4,92)	239,7*** (5,16)	239,2*** (3,84)	240,3*** (4,23)	233,5*** (4,50)	222,7*** (4,79)
IA N3								
Constante	27,3*** (2,98)	27,2*** (2,97)	27,5*** (3,00)	26,2*** (2,89)	23,7*** (2,79)	23,9*** (2,80)	23,7*** (2,78)	22,8*** (2,71)
IA N2								
Constante	18,3*** (1,03)	18,2*** (1,03)	18,5*** (1,04)	18,2*** (1,04)	17,1*** (1,01)	17,0*** (1,00)	16,9*** (1,00)	16,7*** (1,01)
Error N1								
Constante	43,8*** (0,39)	43,8*** (0,39)	43,3*** (0,38)	43,2*** (0,39)	40,9*** (0,37)	40,9*** (0,37)	40,9*** (0,37)	40,8*** (0,38)
<i>N</i>	6687	6687	6687	6443	6413	6413	6413	6166

Errores robustos en paréntesis.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B. Test de hipótesis para para coeficientes de los regresores

En un entorno multinivel los tests de hipótesis usualmente se realizan aplicando máxima verosimilitud o test de Wald con distribución asintótica $\chi^2(q)$ bajo la nula. Esta última es la que utiliza Stata, testendo la hipótesis

$$H_0 : \beta = 0$$

contra la alternativa de dos colas:

$$H_a : \beta \neq 0$$

El estadístico de Wald para la hipótesis nula es

$$w = \left(\frac{\hat{\beta}}{\widehat{DE}(\hat{\beta})} \right)^2$$

el que tiene una distribución asintótica $\chi^2(1)$ bajo la nula. Ahora, el estadístico utilizado para computar la significancia es

$$z = \frac{\hat{\beta}}{\widehat{DE}(\hat{\beta})}$$

el que tiene, a su vez, una distribución asintótica normal bajo la nula.

C. Test de hipótesis para varianza intra-grupos

Consideramos testear una hipótesis de la varianza intra-grupos ψ

$$H_0 : \psi = 0$$

$$H_a : \psi > 0$$

Como expusimos en el texto, esta nula es equivalente a la hipótesis de que $\zeta_k = 0$, es decir, que no hay intercepto aleatorio en el modelo. De no rechazarse la nula, el modelo de intercepto aleatorio no sería necesario.

El test más común para verificar esta hipótesis es el de máxima verosimilitud, el que se implementa estimando el modelo con y sin intercepto. El estadístico es

$$L = 2(l_1 - l_0)$$

donde l_1 es el logaritmo de verosimilitud maximizada del modelo de intercepto aleatorio (que incluye ζ_k) y l_0 es el logaritmo de de verosimilitud maximizada de un modelo de regresión común (sin ζ_k). Como advierten Rabe-Hesketh y Skrondal (2012), la distribución de L bajo H_0 no es χ^2 con 1 grado de libertad, como es usual. Esto ocurre porque la hipótesis nula se encuentra en el borde del espacio de los parámetros, ya que $\psi \geq 0$, lo que viola las condiciones de regularidad de la teoría estadística estándar.

Para datos simulados bajo la hipótesis nula se esperaría que sin el intercepto aleatorio hubiese una correlación intra-grupos positiva en alrededor de la mitad de los datos, mientras que sería negativa para la otra mitad. De esta forma, ψ sería estimado como mayor que cero la mitad de las veces y como igual a 0 la otra mitad de las veces, ya que ψ está restringido a ser no-negativo.

La distribución muestral asintótica correcta bajo la hipótesis nula toma la forma

$$0,5\chi^2(0) + 0,5\chi^2(1)$$

donde $\chi^2(0)$ es una “punta” de altura 1 en 0. Esto porque es una combinación 50:50 de una “punta” en 0 y una χ^2 con 1 grado de libertad.