

El efecto causal de la política de tiempo completo sobre los resultados educativos en la enseñanza media: aplicación de cuatro métodos no experimentales e identificación de posibles sesgos

María Cecilia Llambí¹

RESUMEN

El objetivo de este estudio es analizar, mediante la revisión y aplicación de distintos métodos de evaluación no experimental, los posibles impactos de la política de escuelas de tiempo completo sobre las competencias evaluadas a los 15 años mediante las pruebas PISA de matemática, lectura y ciencias; e identificar las dificultades y posibles fuentes de sesgo presentes en cada método de estimación. Se procura poner foco en algunas dificultades no despreciables que surgen cuando se utiliza esta información de las bases de PISA para evaluar políticas en el nivel primario y medio.

Se analizan y discuten cuatro métodos frecuentemente utilizados en la evaluación de programas: mínimos cuadrados ordinarios, *matching*, variables instrumentales y funciones de control. Se utiliza como restricción de exclusión la distancia del centro PISA a la ETC más cercana. En el caso de ciencias y matemática, se encontraron indicios de la presencia de factores inobservables correlacionados positivamente con haber asistido a ETC y negativamente con los resultados de dichas pruebas. En el caso de lectura no se encontró evidencia de la existencia de este tipo de sesgo. Al tomar en cuenta el sesgo de selección en inobservables, el impacto de haber asistido a ETC sobre las competencias en matemática y ciencias se vuelve positivo cuando la estimación se realiza por funciones de control, mientras que al realizarse por IV no puede rechazarse la hipótesis de que sea nulo. El tamaño del efecto por funciones de control es de 0.34 desvíos estándar en matemática y 0.64 desvíos estándar en ciencias. En el caso de lectura no puede rechazarse la hipótesis de que el efecto sobre esta área de aprendizaje sea negativo.

Palabras clave: educación, evaluación de impacto.

¹ CINVE- Centro de Investigaciones Económicas

I. Introducción

La calidad de los aprendizajes en Uruguay es hoy un tema de agenda pública. Si bien Uruguay se encuentra entre los países latinoamericanos con mejores resultados en las pruebas de aprendizaje evaluadas por PISA², aun se sitúa lejos del promedio alcanzado por los países de la OECD. Por otra parte, Uruguay es uno de los países con mayor desigualdad de resultados y con mayor incidencia del nivel socioeconómico del estudiante en el desempeño educativo, entre los países evaluados por PISA (véase OECD, 2010).

En ese marco, una de las políticas insignia del sistema educativo uruguayo en los últimos quince años fue extender el tiempo pedagógico en la educación primaria, a través del modelo de Escuelas de Tiempo Completo (ETC). Las ETC ampliaron el tiempo pedagógico de cuatro a siete horas y media diarias, además de implementar un nuevo modelo pedagógico, aumentar el componente nutricional de los niños asistentes, y agregar diversos talleres (plástica, música, educación física y/o inglés). La implementación de cada subcomponente de la política varió entre las escuelas. En particular, la combinación de los distintos talleres fue implementada en forma bastante disímil. No obstante, todas comparten la extensión del tiempo de clase, las tres comidas diarias, y prácticamente todas el fortalecimiento de la infraestructura edilicia y el programa general de capacitación a maestros. La política de tiempo completo fue extendiéndose gradualmente, pasando de unas 59 escuelas antes de la implementación del nuevo modelo pedagógico (en 1998) hasta unas 159 escuelas en 2011, con una cobertura de casi 29 mil niños en primaria. (10.5% de la matrícula pública de primaria). Inicialmente estas escuelas se focalizaron principalmente en la atención a niños en los contextos socioculturales más desfavorables, si bien también se extendieron a algunas escuelas de contextos favorables.

A través de la extensión del tiempo pedagógico y la implementación del nuevo modelo de atención se procuró promover mejores niveles de aprendizaje y de desarrollo socioemocional de los niños. En el corto plazo, mejores niveles de aprendizaje se traducen en menores tasas de repetición escolar y una mayor progresión educativa posterior (es decir, un mayor número de años de educación completados en la educación media o terciaria). En el largo plazo, un aumento del nivel de aprendizaje y una mayor progresión educativa se traduciría también en beneficios económicos a través de mayores ingresos, debido tanto al aumento de años de educación como al aumento de la calidad educativa por año completado.

En este estudio se procura estimar el efecto causal de haber asistido a una escuela de tiempo completo sobre las competencias en lectura, matemática y ciencias evaluadas a los 15 años de edad; y discutir e identificar las posibles fuentes de sesgo en la estimación de ese efecto a través de la base de datos de PISA. Para ello se analizan y discuten cuatro métodos frecuentemente utilizados en la evaluación de programas: mínimos cuadrados ordinarios, *matching*, variables instrumentales y funciones de control. El análisis de las ventajas y limitaciones de estos métodos aplicados concretamente a la evaluación de una

² Programme for International Student Assessment.

política utilizando la base de PISA procura también aportar a la discusión metodológica, teniendo en cuenta que dicha base de datos es la única disponible que reporta resultados de aprendizajes en el nivel de enseñanza media y constituye una base sobre la cual se realizan habitualmente análisis de efectos de diversas variables sobre los aprendizajes.

II. Marco conceptual y evidencia empírica

Partiendo del enfoque de Función de Producción Educativa (FPE). Dicho enfoque examina la relación de productividad entre inputs escolares y resultados educativos (en general, resultados de pruebas de aprendizaje) para individuos en edad escolar. Este enfoque provee un marco conceptual que procura guiar la elección de variables y proveer una interpretación coherente de sus efectos. (véase Glewwe, Hanushek, Humpage y Ravina, 2011). Glewwe y Kremer (2006) especifican la forma estructural de la función de producción educativa como sigue:

$$A = a(S, Q, C, H, I) \quad (1)$$

Donde A es el aprendizaje, S es la cantidad de años de educación, Q es un vector de características del centro educativo y de los docentes (inputs que incrementan la calidad educativa), C es un vector de características del estudiante, incluyendo su habilidad, H son las características del hogar (incluyen la educación de los padres y su valoración de la educación) e I es un vector de inputs escolares bajo el control de los padres (asistencia a clase, compras de libros y materiales, etc.). Se supone que los elementos de H y C son exógenos. Si se asume que los hogares pueden elegir entre más de una escuela para sus hijos, entonces Q es endógeno, incluso si es fijo para un centro dado. Por tanto, la elección de los padres para maximizar su función de utilidad es sobre cada posible escuela (Q). Condicional a la elección de Q , eligen S e I sujeto a las restricciones mencionadas. A su vez, S e I pueden expresarse en función de las siguientes variables exógenas:

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

$$I = f(Q, C, H) \quad (3)$$

Sustituyendo las expresiones anteriores en (1) se obtiene la siguiente forma reducida³:

$$A = h(Q, C, H) \quad (4)$$

El impacto que interesa en este caso es el de un elemento del factor Q (inputs escolares que incrementan la calidad educativa), que es el tiempo pedagógico (Q_i). La ecuación (1) provee una estimación de la derivada parcial de A con respecto a Q_i , es decir, permite estimar el efecto de un cambio en Q_i sobre A cuando el resto de los elementos permanecen constantes. La ecuación (4) en cambio, provee una estimación de la derivada total de A con respecto a Q_i , porque permite cambios en S y en I en respuesta a cambios en Q_i . Por ejemplo, los

³ Tal como lo señalan Glewwe et al (2011) la forma reducida contenida en la expresión (4) refleja una relación causal pero no es exactamente una función de producción, dado que refleja preferencias de los padres.

padres pueden responder a un aumento de la extensión del tiempo pedagógico reduciendo su tiempo de atención a los deberes escolares (caso de sustitución de inputs familiares por escolares) o, en otro sentido, aumentando la provisión de insumos escolares como libros (caso donde inputs familiares y escolares son complementarios). El hecho de que las decisiones de los padres pueden contrarrestar o reforzar las políticas a nivel del sistema educativo introduce dificultades al momento de estimar el impacto de los inputs escolares en los aprendizajes, lo que se discute en las secciones siguientes.

Con respecto al tema específico del tiempo de clase, en principio se espera un efecto positivo, dado que, en primer lugar, constituye un incremento de un input específico que incide sobre el aprendizaje. En particular, si el proceso de aprendizaje es secuencial, un mejor aprendizaje de los conceptos básicos al inicio es probable que reduzca las dificultades de los aprendizajes posteriores. En ese caso se esperaría que un incremento del tiempo de clases tenga efectos positivos que persistan en el tiempo. No obstante, que más tiempo de clase redunde en mejores resultados educativos depende del destino que se le de al tiempo extra de clase, y en qué hubieran destinado el tiempo los alumnos si ello no hubiera ocurrido. En particular, también depende de las decisiones intra hogar de asignación de inputs escolares bajo control de los padres, que pueden cambiar ante cambios en la política de extensión horaria.

En términos empíricos la evidencia internacional respecto al efecto de la extensión del tiempo de clase sobre los aprendizajes no es concluyente, si bien en general muestra resultados positivos. Walston y West (2004) comparan los resultados en matemática y lectura de alumnos en preescolar a tiempo completo y a tiempo parcial en la misma escuela, encontrando resultados positivos y significativos para los estudiantes en tiempo completo. Sin embargo, la probable autoselección de estudiantes entre tiempo completo y tiempo parcial en características inobservables hace difícil en este caso atribuir toda la diferencia a la política de tiempo completo.

DeCicca (2007) intenta paliar este problema utilizando datos longitudinales y estimadores de diferencias en diferencias para estimar el efecto de la extensión del preescolar a tiempo completo en Estados Unidos sobre los resultados en matemática y lectura medidos al final de la jardín y al final del primer año de primaria. En este caso se encuentra que la expansión a tiempo completo en el preescolar tiene efectos significativos sobre los resultados de corto plazo (al final del preescolar), pero el impacto decrece fuertemente en el transcurso de un año adicional (al final del primer año) lo cual introduce dudas acerca de su costo efectividad, considerando los efectos más allá del impacto inmediato.

Por otra parte, Cannon et al. (2011) se enfocan en un subgrupo poblacional dado por los estudiantes de habla no inglesa en Los Ángeles, y examinan el impacto de la extensión del preescolar sobre sus resultados académicos, repetición y fluidez en inglés en los primeros años de primaria, explotando la variación de la implementación del tiempo completo en la enseñanza preescolar y utilizando estimadores de diferencias en diferencias. Los autores no encuentran impactos significativos en la mayoría de los resultados académicos analizados hasta segundo grado, si bien encuentran un impacto positivo de reducción de

las tasas de repetición. También encuentran que el impacto varía según características del estudiante y de la escuela.

A nivel de la región latinoamericana, Valenzuela (2005) y analizan el impacto de la extensión del horario de clase en Chile aplicado desde 1997, que significó un 30% de incremento del tiempo de clase para los alumnos de entre tercer año de educación primaria y cuarto año de educación secundaria, explotando la expansión secuencial del programa en las escuelas.⁴ Encuentra un efecto positivo y robusto de dicha política, si bien el efecto encontrado es pequeño. También encuentra un efecto mayor en lengua que en matemática, y en estudiantes de centros privados subsidiados que en estudiantes asistentes a centros públicos. Bellei (2009), en su evaluación del mismo programa encuentra efectos positivos y significativos aunque pequeños, y un impacto mayor en estudiantes en centros rurales, en centros públicos y en estudiantes con mejores desempeños previos.

Finalmente, en la revisión de la literatura sobre recursos y resultados educativos desde 1990 hasta 2010 realizada por Glewwe et al. (2011), los resultados de los efectos de las horas de clase sobre los resultados educativos son relativamente no ambiguos y en la dirección esperada: seis de las ocho estimaciones reportan resultados positivos, y cuatro de ellas significativamente positivos. Cuando los autores restringen el análisis a los 43 estudios de “alta calidad” en términos de los métodos econométricos utilizados, encuentran que las cuatro estimaciones analizadas reportan resultados significativamente positivos.

Para el caso de Uruguay, a nivel de las evaluaciones nacionales de aprendizajes que realiza la ANEP, en general se ha encontrado un mejor desempeño de los alumnos en ETC respecto a sus pares en escuelas de similar contexto sociocultural (véase ANEP, 2000). También se ha encontrado una valoración positiva del modelo ETC por parte de docentes y padres, fundamentalmente expresada a través de un juicio favorable en cuanto al impacto del modelo sobre el desarrollo cognitivo y social de los niños (véase Equipos Mori, 2001). En términos de repetición escolar, las ETC presentan tasas de repetición algo más bajas a la media de las escuelas, si bien atienden a una población proveniente de contextos más desfavorables. No obstante, el Informe de ANEP (2007:a) encuentra una relación negativa entre asistir a una ETC durante la mayor parte de la Primaria, y el nivel de competencias científicas adquiridas a las 15 años, utilizando una estimación a través de un modelo de regresión multi nivel o modelo jerárquico lineal (HLM) sobre los datos de PISA 2006.

Los resultados basados en las evaluaciones nacionales de aprendizajes de ANEP y el informe de ANEP (2007:a), procuran identificar el efecto de asistir a ETC incluyendo un extenso set de variables de control en las estimaciones. No obstante, el sesgo de selección en variables inobservables de escuelas y niños al modelo ETC hace difícil atribuir las diferencias en el desempeño educativo a las diferencias entre asistir o no a una escuela con extensión del tiempo pedagógico. En particular, afecta la consistencia de los estimadores obtenidos a través de regresiones de mínimos cuadrados ordinarios (MCO) o modelos jerárquicos (HLM).

⁴ La reforma implicó que los alumnos pasaran a tener unas 1.140 horas anuales de clase, lo que significó un 35% más que el promedio de la OCDE.

Dos estudios han procurado identificar el efecto ETC sobre el desempeño educativo a través de técnicas de evaluación de impacto que procuran controlar el sesgo por variables inobservables de asistencia a una ETC. Cerdan Infantes y Vermeesch (2007) utilizan un pool de bases de datos de pruebas nacionales de aprendizajes para evaluar el efecto de ETC sobre el desempeño escolar a nivel de sexto año de Primaria, utilizando técnicas de diferencias en diferencias y *propensity score matching*, en base a una muestra de escuelas que observan en 1996 (antes de convertirse a tiempo completo) y 2002. Encuentran un impacto significativo de las ETC sobre el aprendizaje en lengua y matemática, en particular en escuelas de contextos más desfavorables. Estiman una mejora de 0.26 desvíos estándar del puntaje de Lengua y de 0.38 de Matemática, para niños que realizan los seis años de Primaria en este tipo de escuelas.

Por otra parte, Perera y Llambí (2009) estiman el efecto ETC sobre las competencias evaluadas a los 15 años usando la información de PISA 2006. Los autores procuran corregir dos tipos de sesgo. Por un lado, el sesgo por no observación de estudiantes desertores en las pruebas PISA, lo que podría causar sesgo en la estimación del efecto ETC, si se presume que uno de los resultados de dicha política es la menor repetición en Primaria y consecuentemente, la menor deserción en Secundaria. Por otro lado, utilizan variables instrumentales y el método de función de control para procurar corregir los posibles sesgos en la estimación del efecto de ETC debido a la endogeneidad de dicha variable. Los instrumentos utilizados, que se encuentra que están altamente correlacionados con asistir a ETC son: variable indicadora de que la ocupación de la madre es de baja calificación, variable indicadora de que la madre del alumno está ocupada y *dummies* regionales. Encuentran que la magnitud y la significación del efecto ETC es sensible al método utilizado para tratar el sesgo provocado por la endogeneidad de esta variable, pero no obstante se constata un cambio de signo en el coeficiente respecto a las estimaciones que no tratan este problema (que encuentran un coeficiente negativo). Si bien este estudio avanza en la discusión de los efectos de ETC, es discutible la idoneidad de los instrumentos utilizados, en particular el requisito indispensable de no estar correlacionado con la variable de resultado.

En definitiva, a nivel teórico el signo esperado del efecto de un aumento del tiempo de clase sobre los aprendizajes es, en principio, incierto, y depende de en qué se destina el tiempo adicional y cómo cambia la asignación de inputs educativos de las familias ante este cambio. En términos empíricos en general los resultados indican una relación positiva entre tiempo de clases y resultados académicos. No obstante, el tamaño del efecto puede variar significativamente dependiendo de las características de las escuelas, los profesores o los estudiantes; de área del conocimiento sobre la que se extiende el tiempo de clase y sus complementariedades con el área evaluada; del resultado educativo analizado; y del momento en el cual se evalúan los impactos.

III. Datos

Se utiliza la base de datos de PISA (*Programme for International Student Assessment*) del año 2006. PISA es un programa de evaluación de competencias de estudiantes a los 15 años de edad, sobre una muestra representativa de la población asistente a la educación media, que se aplica en países de la OECD y también en otros países del resto del mundo (57 países en total, en 2006). PISA provee información sobre resultados en pruebas de Matemática, Lectura y Ciencias, grado que cursa el estudiante, si repitió algún año escolar, así como también abundante información sobre características del estudiante y su familia, características del centro y características institucionales. En 2006 se incluyó además una pregunta sobre a qué tipo de escuela asistió el estudiante la mayor parte de la primaria, donde una de las opciones es la escuela de tiempo completo.

En términos de la evaluación académica PISA procura evaluar cada área, no basándose en el currículum (que varía entre países), sino en términos de conocimientos y habilidades requeridas para la vida adulta (OECD, 2000). Adicionalmente a los *tests* en las áreas evaluadas, los estudiantes y los directores de los centros educativos responden cuestionarios, de donde surge la información de *background* de los estudiantes y sus familias, recursos de los centros educativos y características de los profesores, así como también el set up organizacional. Combinando todos los sets de información, se construye una base de datos que contiene, para 2006, información de 4.839 estudiantes donde se combina la información de los *tests* de matemática, lectura y ciencias con las características personales y familiares del estudiante y las del centro al que asiste.

Además de las preguntas originales de los cuestionarios a los directores de los centros y a los estudiantes, la base de PISA incluye índices derivados de las preguntas originales que resumen las respuestas de los estudiantes y de los directores a un set de preguntas. Estos índices están diseñados sobre la base de consideraciones teóricas e investigaciones empíricas previas, mediante técnicas estadísticas de análisis factorial (OCDE, 2003). La ventaja de utilizar estos índices radica en que han sido construidos con rigor y su validez ha sido testeada a través de varios estudios a nivel internacional (OCDE, 2003).

En particular, se utiliza el índice de estatus económico, social y cultural, elaborado en base a las siguientes variables: el índice de status ocupacional; el nivel educativo más alto alcanzado por los padres, convertido a años de educación; el índice PISA de riqueza familiar; el índice PISA de recursos educacionales del hogar; y el índice PISA de recursos culturales en el hogar. En relación con los factores escolares, algunos ejemplos de índices que procuran captar características de los recursos del centro son el de calidad de los recursos educativos⁵, el ratio estudiantes por profesor o el porcentaje de profesores con título docente o grado universitario. También se dispone de índices que captan aspectos institucionales y de gobernanza. Por ejemplo, a los efectos de capturar el grado de autonomía de los centros se construyen dos índices: el índice de autonomía de recursos,

⁵ Derivado de un conjunto de ítems que relevan la percepción de los directores de los centros sobre los factores que dificultan el desarrollo de la actividad educativa. Valores positivos del índice indican mejor calidad de los recursos educativos.

que se construye en base al número de decisiones relativas a los recursos escolares que son responsabilidad del propio centro; y el índice de autonomía curricular, elaborado de acuerdo al número de decisiones sobre currículum y evaluación que son responsabilidad del centro.

La otra fuente de información utilizada es la base de distancias entre el centro educativo relevado en PISA y la escuela de tiempo completo más cercana, disponible al momento en que el estudiante cursaba educación primaria. La base fue elaborada por CINVE, y fue construida de la siguiente forma. Si bien no se dispuso de la dirección de residencia de cada estudiante evaluado en PISA, sí fue posible identificar a los centros educativos evaluados en la base de PISA 2006, únicamente a los efectos de obtener este vector de distancias. Así, se obtuvieron dos bases de datos: una con la localización geográfica (dirección) de cada centro de educación media de Montevideo y del Interior del país evaluado en PISA 2006 (en adelante, centro PISA); y otra con los datos de localización de las escuelas de tiempo completo y el año de conversión al programa. A partir de ambas fuentes de información, se identificó para cada centro PISA cuál era la escuela de tiempo completo más cercana para cada año comprendido entre 1999 y 2003 (lapso de tiempo en el que la mayor parte de los estudiantes evaluados en PISA 2006 realizaron sus estudios de primaria). En base a dicha información se calculó la distancia (en metros) de cada centro PISA a la escuela de tiempo completo identificada como más cercana en cada uno de los años del período considerado. Las distancias se obtuvieron utilizando un software de geo-referenciación de datos.

El vector de variables de distancias fue adicionado a la base de datos de PISA, asignando a cada estudiante de cada centro educativo los correspondientes vectores de distancias del centro a la ETC más cercana para cada año entre 1999 y 2003. Finalmente, se asignó a cada estudiante el vector de distancia de alguno de los años considerados, en función del grado que estuviera cursando al momento de ser evaluado por PISA. La regla utilizada fue asignar el vector de distancias a las ETC disponibles en el año en que el estudiante estuviera cursando la mitad (tercer año) de los estudios de primaria.⁶ De todas maneras, en la gran mayoría de los casos la escuela de tiempo completo más cercana al centro PISA (y por consiguiente, la distancia del centro a la misma) no cambia en el período considerado (1999-2003). En el caso del Interior, el 84% de los centros tiene en todo el período una misma ETC clasificada como “más cercana”. En el caso de Montevideo dicho porcentaje es de 69%.

IV. Estrategia empírica

La estimación del efecto de haber asistido a una escuela de tiempo completo sobre las competencias en lectura, matemática y ciencias evaluadas por PISA a los 15 años de edad se realiza a través de cuatro métodos frecuentemente utilizados en la evaluación de programas: mínimos cuadrados ordinarios, *matching*, variables instrumentales y funciones

⁶ Por ejemplo, a los estudiantes de primer año de educación media se les asignó el vector de distancias correspondientes al año 2003; a los de segundo año, el de 2002; y así sucesivamente.

de control. Se analizan comparativamente, tanto desde el punto de vista metodológico como en la aplicación concreta a la evaluación de ETC utilizando la información de PISA y evaluando la sensibilidad de los resultados. A partir del contraste de la magnitud y el signo de las diferentes estimaciones se intenta inferir qué tipo de selección se encuentra por detrás de los datos observados.

Se parte de una especificación lineal de la función de producción educativa en su forma reducida (ecuación (4)), A efectos de facilitar la exposición, de aquí en más se denomina X a la matriz de características observables del estudiante, del hogar y de los centros educativos. Por lo tanto la ecuación de interés se reduce a la siguiente expresión:

$$A_{ij} = X_{ij}\beta + ETC_{ij}\theta + u_{ij} \quad (5)$$

El parámetro θ de la ecuación (5) corresponde a la forma reducida, por lo cual mide el efecto total ETC sobre los resultados: tanto el efecto directo como el indirecto a través de los años completados. El parámetro de la ecuación (5) también incorpora los efectos indirectos a través de posibles cambios en los insumos educativos que son decisión de los hogares (por ejemplo, el tiempo dedicado a apoyar el estudio de los hijos, etc.). El término de error u , que captura variables inobservables, errores de medida de las variables de resultado y de inputs, o relaciones no lineales no capturadas por la forma lineal de especificación.

Finalmente, para analizar si existen efectos diferenciales a nivel de los aprendizajes adquiridos en la educación media se introducen interacciones de la variable de interés con las características que se presumen que se asocian a impactos diferenciales. Ello permite identificar *heterogeneidad en observables*. La ecuación a estimar en ese caso será la siguiente:

$$A_{ij} = X_{ij}\beta + ETC_{ij}\theta + ETC_{ij}X_{ij}\psi + u_{ij} \quad (5')$$

Cabe notar que se consideran tres variables de resultado, correspondientes a los puntajes de las pruebas de aprendizaje en lectura, matemática y ciencias, respectivamente.

El interés es estimar el ATT (impacto promedio del programa sobre los tratados, *Average Treatment effect on the Treated*) de la política de ETC, sobre las competencias adquiridas a los 15 años. El ATT corresponde a la diferencia entre el valor esperado de la variable de resultado (por ejemplo, el resultado de la prueba de matemática de PISA) en el grupo de los tratados (los que fueron a ETC la mayor parte de la primaria) y el que hubieran obtenido los tratados de no haber asistido a ETC. El punto crucial es que la muestra de individuos que eligió asistir a ETC no es aleatoria, sino que sufre de sesgo de selección por varios posibles factores que se discuten más adelante. Si esto es ignorado, la comparación simple entre individuos que asistieron y no asistieron a ETC estará afectada por este sesgo.

Heckman, Ichimura, Smith y Todd (1998) descomponen el sesgo en tres componentes: a) el sesgo debido a la ausencia de suficiente solapamiento en las distribuciones de las variables observables (X) en los dos grupos de comparación; b) el sesgo debido a una errónea ponderación de las variables observables dentro del soporte común o solapamiento, si la distribución empírica de las variables observables no es la misma entre tratados y no tratados, incluso dentro del soporte común; y c) el sesgo de selección en variables inobservables. Las primeras dos fuentes de sesgo se originan en diferencias en la distribución de las características observables (X) entre tratados y no tratados, mientras que la tercera fuente refiere al sesgo econométrico que surge de la existencia de variables inobservables que afectan simultáneamente la participación en el programa y los resultados.

Dentro de los métodos no experimentales que buscan identificar el ATT, se encuentran aquellos que procuran controlar por la correlación entre los factores individuales y la participación en la política evaluada por la vía de una variable instrumental o restricción de exclusión; y aquellos que intentan considerar a todos los factores observables que pueden causar simultáneamente la participación y los resultados para luego hacer un *matching* en estas variables. La implementación de cada método y sus propiedades difieren de acuerdo a si existe selección en inobservables; y si el modelo subyacente es uno de respuestas homogéneas o heterogéneas a la política.

Como ya fue mencionado, en el presente análisis se aplican cuatro métodos de estimación no experimentales que buscan estimar el parámetro de interés (ATT) bajo distintas condiciones: Mínimos Cuadrados Ordinarios (MCO), *matching*, funciones de control y variables instrumentales. La función de control permite evaluar la importancia del sesgo de selección debido a variables inobservables que afectan simultáneamente la participación en ETC y los resultados en las pruebas de aprendizaje. Por otra parte, la comparación de la regresión MCO y la técnica de *matching* permite vislumbrar posibles problemas de especificación de la forma funcional, así como también la presencia de efectos heterogéneos, (coeficientes de respuesta diferentes a la política de tiempo completo de acuerdo a algunas características observables de los individuos). Se analiza la existencia de heterogeneidad del efecto en observables en la aplicación MCO y de funciones de control y se compara con el estimador de *matching*. Finalmente, el método de variables instrumentales permite obtener un estimador consistente en presencia de selección en inobservables, si bien el efecto estimado es local, como se detalla más adelante.

Volviendo a las especificaciones de las ecuaciones (5) y (5'), existen varias fuentes potenciales de sesgo en una regresión de los resultados de las pruebas de aprendizaje sobre la asistencia a ETC para estimar el ATT (véase Blundell, Dearden y Sianesi, 2004):

a. Sesgo debido a observables

La especificación estándar MCO generalmente controla linealmente por el set de variables observables X , lo que puede inducir a dos potenciales fuentes de sesgo en observables: a) especificación errónea de la influencia de las X , independientemente de la variable de tratamiento. Por ejemplo, si el verdadero modelo contiene términos cuadráticos o de mayor orden de las X , o si contiene interacciones entre las X ; y b) impactos heterogéneos. Si el efecto de ETC varía de acuerdo a alguna característica contenida en X y ello no es incluido en la regresión, el estimador MCO de θ no recupera el ATT. Estos aspectos de errónea especificación están ligados a las dos primeras fuentes de sesgo señaladas por Heckman, Ichimura Smith y Todd (1998).

La aproximación MCO de la función sobre las X en la región donde no se solapan las densidades entre tratados y no tratados se basa enteramente en la forma funcional elegida. Debido a ello, la identificación del contrafactual a través de MCO para los individuos tratados que se encuentran fuera de la región de solapamiento depende crucialmente de que la forma funcional especificada sea la “correcta” (véase Blundell, Dearden y Sianesi, 2004; e Imbens y Wooldridge, 2008).

b. Sesgo debido a inobservables

El efecto causal de haber asistido a ETC sobre los resultados evaluados en PISA puede ser consistentemente estimado por MCO sobre una función correctamente especificada solamente si el término de error u_{ij} está incorrelacionado con la variable de participación en ETC (supuesto de independencia condicional).

Hay una serie de factores que hacen difícil que la condición de exogeneidad (incorrelación de la variable de interés con el error) se cumpla. Dada la fuente de datos utilizada, pueden identificarse dos potenciales fuentes de sesgo: i) el sesgo de selección de niños y de escuelas en el programa ETC; e ii) el sesgo de selección muestral ocasionado por la no observación de los estudiantes que abandonaron los estudios a los 15 años, no evaluados en la prueba PISA.

i) Sesgo de selección de niños y escuelas en ETC

Dejando momentáneamente de lado el tema de la no observación de alumnos que abandonaron los estudios a los 15 años, algunos de los factores que pueden ocasionar que se viole el supuesto de independencia condicional para el caso de una variable de interés dentro de una función de producción educativa son (Glewwe et al, 2011):

a) Variables omitidas

El conjunto de información sobre los inputs educativos es siempre incompleta, y tampoco se dispone en general de información sobre la habilidad. Si las variables omitidas estuvieran correlacionadas con la variable de interés, la estimación de efecto tendrá un

sesgo proporcional a la importancia del factor omitido y su correlación con la variable de interés. El factor de variables omitidas puede ser un potencial problema para estimar consistentemente el efecto ETC, en particular en lo que refiere a las características de los inputs o factores escolares que reciben los estudiantes a lo largo de la trayectoria por la educación media. Dado que este trabajo procura identificar si existen efectos *persistentes* de asistir a ETC sobre los resultados en la educación media, importa el efecto de los inputs educativos recibidos durante este nivel. Si existen inputs *diferenciales* entre asistentes y no asistentes a ETC y ello no es incluido en la regresión, no se cumple el supuesto de independencia condicional, y el efecto de ETC sobre los resultados en educación media estaría sesgado. La dirección del sesgo depende del signo del efecto de la variable omitida sobre el resultado educativo, y de la correlación de la variable de interés (haber asistido a ETC) con la variable omitida. Por ejemplo, se observan diferencias significativas en algunos recursos educativos de los centros a los que concurren asistentes y no asistentes a ETC, así como también diferencias significativas en cuanto al nivel socioeconómico promedio de los estudiantes en el centro. Si dichas características (de los recursos y de los pares) tienen efectos sobre los resultados y no se incluyen en la regresión, el efecto de haber asistido a ETC estará presumiblemente sesgado a la baja.⁷

b) Sesgo de selección de niños en escuelas.

La decisión de los padres sobre el centro educativo al que asiste al hijo está afectada por las características del centro y de los inputs escolares, lo que se conoce como sesgo de selección de niños en escuelas. Efectivamente, la introducción de escuelas de tiempo completo puede haber afectado la elección de algunos padres sobre el centro educativo donde enviar a sus hijos, en un signo que depende de la valoración del programa que realizan los padres y su capacidad de agencia. En tal caso, la selección de niños en estas escuelas no será aleatoria, sino que estará correlacionada con características inobservables de los padres.

c) Políticas educativas endógenas

El set de inputs escolares puede estar correlacionado con características (observables e inobservables) de los centros educativos. Por ejemplo, podría existir una política que destinara más recursos a escuelas con problemas, o en sentido contrario, que se asignen más recursos a escuelas con más capacidad de gestionarlos. En ambos casos, la estimación de los efectos de los inputs escolares sobre los aprendizajes estaría sesgada. En el caso del programa de Tiempo Completo, el mismo no fue asignado aleatoriamente en escuelas, sino que tuvo un propósito específico de atender a las escuelas de contextos más desfavorables. Aun así, no hubo una regla de decisión explícita para incorporar a escuelas al programa, observándose que también fue introducido en escuelas de los contextos más favorables. Este hecho indicaría que la asignación de escuelas al programa Tiempo Completo estuvo

⁷ Dado que la calidad de los recursos y el nivel socioeconómico promedio de los centros está correlacionado positivamente con los resultados educativos obtenidos por los estudiantes, y negativamente con haber asistido a ETC.

influido por una serie de factores inobservables que podrían estar correlacionados con los resultados educativos.

d) Errores de medida.

La imperfecta medición de variables de resultado y covariables (lo cual es habitual en funciones de producción educativa) pueda sesgar las estimaciones de los efectos.

En definitiva, existen al menos las siguientes fuentes de endogeneidad que dificultan la correcta identificación del parámetro de interés: el posible sesgo de selección de niños que asistieron a tiempo completo, el posible sesgo de selección de escuelas en tiempo completo y el posible sesgo en la elección de los centros de educación media a los que asisten los estudiantes luego de finalizada la educación primaria. Ello hace que sea probable que la asistencia a Tiempo Completo esté correlacionada con características observables e inobservables que afectan los resultados educativos, y por lo tanto, su omisión en la regresión estimada resulta en una estimación sesgada del efecto del programa ETC.

ii) Sesgo de no observación de estudiantes desertores

Una potencial limitación de los datos de PISA es el problema ocasionado por la no observación de los estudiantes que abandonaron el sistema educativo entre los 15 y 16 años (al momento de realizar las pruebas PISA), lo que causa una selección muestral no aleatoria. Este aspecto es particularmente relevante en Uruguay dadas las elevadas tasas de deserción durante el pasaje por la enseñanza media. En 2006, el 19,6% de los jóvenes entre 15 y 16 años no asistían al sistema educativo (véase Perera y Llambí, 2009). Si dicha selección muestral está correlacionada con asistir a una ETC, el efecto estimado de esta variable puede estar sesgado.

Perera y Llambí (2009) y también Méndez y Zerpa (2010) recurren al modelo de dos etapas propuesto por Heckman (1979) para intentar corregir este problema. No obstante, se deben utilizar dos conjuntos de bases distintas para la estimación de cada ecuación (la primera y la segunda etapa). En definitiva, utilizan para la estimación de la primera etapa a las Encuestas Continuas de Hogares, que permiten disponer de una muestra de jóvenes de 15 años asistentes y no asistentes al sistema educativo. Luego estiman la segunda ecuación (la FPE) en la base de PISA, computando el inverso del ratio de Mills a partir de las variables en dicha base de datos. No obstante, la utilización de dos conjuntos de información independientes no es el procedimiento estándar de estimación de este tipo de modelos. Ello exige como condición que el conjunto de variables explicativas de la ecuación de selección esté conformado por variables que sean observadas en ambas bases de datos (ECH y PISA). En la práctica, si bien existen variables incluidas en ambas bases de datos, en general las mismas están diferentemente medidas, lo que causa que la distribución en ambas muestras puede diferir.⁸ Por otra parte, también es difícil encontrar una restricción de exclusión válida y creíble. Debido a ello, en este trabajo no se realiza el intento de corrección por dicho sesgo. No obstante, debe tenerse en cuenta que existe selectividad

⁸ Por ejemplo, la variable de educación de los padres presenta una distribución bastante diferente en ambas muestras.

(presumiblemente no aleatoria) en la asistencia a educación media a los 15 años, y que es posible que dicha selectividad se correlacione con la asistencia a una ETC, lo que provocaría un sesgo en la estimación del impacto de la variable de interés. Este aspecto se discute a continuación.

iii) ¿Qué tipo de sesgo en inobservables sugieren primariamente los datos?

Existen algunos indicios que podrían dar cuenta de selección muestral correlacionada con asistir a ETC, y/o selección de niños en ETC, en base a los datos de PISA y los provenientes de los registros administrativos oficiales.

En primer lugar, el porcentaje de alumnos asistentes a ETC en la muestra de estudiantes evaluados por PISA es bastante superior al porcentaje de matriculados en ETC sobre el total de matriculados en enseñanza primaria, en todo el período anterior a 2006. De acuerdo a la base de PISA 2006, los asistentes a ETC representaron el 9,1% del total de estudiantes de 15 años asistentes a la educación media. En el mismo año, la matrícula de ETC representó el 6,2% de estudiantes de primaria. Si se consideran años anteriores, que son los relevantes porque corresponden al período cuando cursó la primaria la cohorte evaluada por PISA 2006, los porcentajes son incluso inferiores (5,4% en el promedio 2002-2005). Ello podría sugerir que la permanencia en el sistema educativo a los 15 años es mayor entre los asistentes a ETC en relación a los no asistentes.

Por otra parte, si bien de acuerdo a los registros administrativos la tasa de repetición en ETC es similar o inferior (dependiendo del año considerado) a la tasa promedio de los asistentes a escuelas públicas, los datos de PISA evidencian que los asistentes a ETC presentan tasas de repetición sustancialmente mayores a la media de asistentes a escuelas públicas. Efectivamente, el 32% de los estudiantes que asistieron a ETC declara haber repetido algún año en primaria. El mismo porcentaje entre los no asistentes a ETC se ubica en 17%. Si se restringe la comparación únicamente entre asistentes a escuelas públicas urbanas la diferencia es igualmente significativa: 32% versus 20% en ETC y escuelas de tiempo parcial, respectivamente.

Una posible explicación a estas discrepancias entre los datos de PISA y los datos de los registros administrativos es que exista una mayor tasa de abandono de los estudios medios entre estudiantes repetidores *que no asistieron* a ETC. Si el hecho de abandonar los estudios estuviera negativamente correlacionado con los aprendizajes, ello podría sesgar a la baja la estimación del impacto de ETC sobre las competencias a los 15 años. No obstante, el tamaño de la selectividad debería ser muy alto para que solamente este factor por sí solo explique estas discrepancias.

Otra posibilidad (que puede ser adicional a la primera) es que existiera un sesgo de selección de niños con más experiencias de repetición *previas* a entrar en ETC. Recuérdese que la identificación en PISA de los estudiantes que asistieron a ETC refiere a aquellos que asistieron *la mayor parte* de primaria. En general, cuando una escuela pasa al programa de Tiempo Completo, todos sus alumnos pasan al programa, independientemente del grado que les corresponde cursar. Es posible que algunos alumnos hayan pasado a tiempo

completo luego de cursar los primeros grados en un régimen de tiempo parcial. Considérese además que entre 1998 y 2004 el número de escuelas de tiempo completo prácticamente se duplicó, por lo que este tipo de situaciones puede ser no despreciable en la muestra analizada.

Si las ETC tienden a instalarse en escuelas de contextos más desfavorables y con mayores índices de repetición, es posible que los alumnos asistentes a ETC observados en PISA presenten tasas de repetición mayores a las que los registros administrativos muestran (dado que estos últimos computan la repetición en ETC solo cuando la escuela pasa a ese régimen). Lamentablemente los registros no permiten identificar si la repetición en primaria ocurrió antes o después de que el alumno entrara en ETC, por lo que en realidad no es posible asociarla ni a una condición previa ni a un efecto causal de ETC. Pero si la entrada de escuelas a ETC estuviera asociada a contextos de mayores problemas de aprendizajes *previos* ello podría sesgar a la baja el impacto estimado de la política.

En definitiva, la sobre representación de alumnos asistentes a ETC en la muestra de PISA, unido a la diferencia en la tasa media de repetición observada entre asistentes y no asistentes a ETC (sustancialmente mayor a la que se derivaría de los registros administrativos) parecen evidenciar problemas de sesgo de selección muestral y/o sesgo de selección en inobservables que podrían sesgar a la baja la estimación del impacto de ETC sobre los aprendizajes en la educación media.

iv) Los diversos estimadores: MCO, *matching*, variables instrumentales y funciones de control

El método de *matching* trata el tema del sesgo en observables procurando controlar en forma directa y *flexible* por las variables que pueden causar dicho sesgo. (véase Heckman, Ichimura y Todd (1998); Caliendo y Kopeinig (2005); Bernal y Peña (2011)). El supuesto explícito es que todas las diferencias relevantes entre tratados y no tratados que inciden en el resultado están capturadas en las variables observables, siendo el hecho de haber sido tratado (en este caso, de haber asistido a ETC) la única diferencia entre los dos grupos. El método de *matching* se basa en el supuesto fundamental de independencia condicional, que implica que la participación en ETC no está determinada por variables no observadas que también determinan los resultados potenciales. Bajo estos supuestos, se forma un subconjunto de observaciones comparables de la muestra original y es posible obtener un estimador del ATT (dentro del soporte común) como la diferencia de las medias de los resultados sobre el soporte común, *ponderadas apropiadamente por la distribución de las características X en el grupo de los tratados*.

Una alternativa que procura corregir por el sesgo debido a inobservables es el uso de variables instrumentales (IV) (véase Imbens y Angrist, 1994; Heckman, 1997; y Angrist y Krueger, 2001). El método se basa en la utilización de una variable (instrumento) que esté correlacionada con la variable de interés (haber asistido a una ETC) y no esté correlacionada con los inobservables de la ecuación de resultado (el error (u) en la

ecuación 5); Si se cuenta con un instrumento de estas características, es posible obtener un estimador consistente del ATT, en el caso de que los efectos sean homogéneos.⁹

En este trabajo se utiliza como instrumento la *distancia* a la escuela de tiempo completo más cercana, disponible al momento en que el estudiante cursaba la primaria. Este tipo de variables geográficas suelen ser utilizadas como instrumentos válidos, debido a que constituyen elementos exógenos al resultado final, pero afectan a la variable de interés (véase por ejemplo Card, 1993). En particular, la distancia a la escuela de tiempo completo más cercana afecta la decisión (o posibilidad) de asistir a una escuela de estas características, pero, condicional al vector de características relevantes X , no debería afectar el resultado final en términos de aprendizaje, excepto a través de la propia elección de asistir o no a una ETC.

Un aspecto importante del método de variables instrumentales es que el efecto del programa evaluado se identifica a partir de un *subconjunto* de individuos: aquel que *cambia su decisión de participar* (de asistir a ETC) debido a la cercanía de una ETC (o disponibilidad de ETC en su barrio o localidad, individuos cooperativos). Debido a ello, el efecto estimado es un efecto local (*LATE: local average treatment effects*) y no promedio (véase Imbens y Angrist, 1994). Es decir, con el método de variables instrumentales se no se recupera el ATT, excepto bajo el supuesto que los efectos fueran homogéneos.

En este caso, debe notarse que la distancia del centro PISA a la ETC más cercana sería un instrumento aceptable si los estudiantes asisten a centros de educación media cercanos a sus hogares de residencia, y si no ocurren traslados de zonas entre el momento en que el estudiante cursó la primaria y el momento en que es evaluado por PISA. Respecto al segundo aspecto, se procuró controlar por la movilidad de zona excluyendo del análisis a aquellos estudiantes que cursaron la primaria en una zona distinta a su actual zona de residencia.¹⁰ El primer aspecto no es verificable, pero de no ocurrir afectaría la relevancia del instrumento, lo cual se testea mediante los *tests* habituales que se exponen más adelante.

Finalmente, debe notarse que la distancia de un centro PISA a la ETC más cercana no sería un buen instrumento si la instalación de ETC estuviera *sistemáticamente* ligada a factores inobservables de las *escuelas* que presumiblemente afecten los resultados de aprendizaje (por ejemplo, si la instalación de ETC estuviera ligada sistemáticamente a escuelas con directores más proactivos; o contrariamente, a escuelas con menor capacidad de gestión de los recursos). En tal caso, si bien la distancia corregiría el sesgo de selección *de niños* en las escuelas, no corregiría un posible sesgo de endogeneidad de la propia política educativa.

Por otra parte, el método de funciones de control procura estimar el efecto controlando directamente por la correlación del error en la ecuación de resultado con la variable de

⁹ También si los efectos son heterogéneos en observables, pero se requerirán tantos instrumentos como variables endógenas.

¹⁰ Para ello se utilizó la pregunta al estudiante: *¿en qué departamento del país cursaste la mayor parte de primaria?*

interés. Para ello, se requiere modelar explícitamente la ecuación de participación. En este caso, la ecuación de participación relevante es la decisión de asistencia a una ETC.

Al igual que en el método de variables instrumentales, para identificar el modelo debe existir una restricción de exclusión, es decir, al menos una variable que explique la decisión de participación y no afecte directamente la variable de resultado. En este caso, se utiliza la misma restricción de exclusión que en el método IV, la distancia del centro PISA a la ETC más cercana. Sin embargo, la estructura impuesta por el método de función de control tiene algunas ventajas respecto a IV. En primer lugar, permite obtener el ATT, mientras que el método IV estima solamente el efecto local para el instrumento específico relacionado a la subpoblación de los cooperativos. A su vez, con el método IV en el caso de efectos heterogéneos observables, los instrumentos disponibles pueden ser muy débiles para predecir todas las interacciones apropiadamente.

En este caso, el efecto heterogéneo que se evalúa es el posible impacto diferencial asociado al *background* sociocultural del estudiante. La distancia interactuada con la variable de educación de los padres es un instrumento muy débil para predecir la interacción de la variable endógena con la educación de los padres. O sea, que el método de IV permite solamente recuperar un efecto local, mientras que la función de control permite recuperar el ATT en forma más eficiente. La contracara es que este método es mucho menos robusto que IV porque descansa en una especificación precisa de la regla de participación en la política y supone normalidad de la distribución conjunta de los residuos.

En definitiva, en el caso de que sólo hubiera sesgo de selección en observables, si no hay suficiente solapamiento de las características X entre los grupos comparados, o si hay errores de especificación funcional, el método de *matching* siempre será preferible al MCO. Pero incluso si se asume que los aspectos anteriores no son un problema, el método de *matching* provee una mejor estimación del ATT en presencia de efectos heterogéneos.¹¹

Si hay selección en inobservables, tanto *matching* como MCO producirán estimadores sesgados. En ese caso, si IV y funciones de control están correctamente especificadas ambas producirían estimadores consistentes del ATT; pero en el caso de IV solamente bajo el supuesto de efectos homogéneos. En el caso de efectos heterogéneos IV estima el efecto local ligado al instrumento utilizado, que puede diferir del ATT. La función de control permite además contrastar la hipótesis de ausencia de selección en inobservables.

V. Resultados

V.1 MCO

En primer lugar, se presentan los estimadores MCO bajo distintas especificaciones en cuanto al condicionamiento de las variables X incluidas en la regresión. Recuérdese que

¹¹ Si bien puede aproximarse si no hay efectos muy disímiles de acuerdo a características X o si los valores del *propensity score* son menores a 0.5 (véase Blundell, Dearden y Sianesi, 2004).

todas las variables que tienen influencia *simultánea* sobre haber asistido a ETC y sobre los resultados deben ser incluidas como regresores para asegurar independencia condicional si no hubiera selección en inobservables. Por otra parte, las X deberían ser características no afectadas por la variable a evaluar (es decir, no deberían incluirse características que puedan ser resultado de haber asistido a ETC). Finalmente, pueden incluirse características que afecten el resultado final pero no a la variable de interés. Ello no afecta el sesgo de la estimación del efecto de ETC, pero sí contribuye a reducir la varianza del estimador. El cuadro 1 (panel superior) presenta los estimadores MCO del coeficiente asociado a haber asistido a ETC sobre los puntajes de las pruebas PISA de ciencias, matemática y lectura, en cuatro especificaciones que difieren en cuanto al set de covariables utilizadas como controles en la regresión. El en panel inferior se presentan los resultados admitiendo la posibilidad de efectos diferenciales para alumnos provenientes de hogares de bajo nivel educativo, para la especificación preferida.

Las dos primeras columnas muestran los estimadores MCO del efecto de haber asistido a ETC condicionando por un amplio set de variables sobre características del estudiante y su hogar, sus pares y el centro de educación media al que asistía al momento de ser evaluado por PISA.¹² Solamente difieren en que la primera columna incluye como variable de zona a una *dummy* indicadora de residir en Montevideo, mientras que la columna 2 incluye un set de *dummies* por departamento. Las columnas 3 y 4 solamente incluyen las variables que tienen un efecto estadísticamente significativo sobre las variables de resultado, y también difieren en que la columna 3 incluye la *dummy* de Montevideo, y la 4 el conjunto de *dummies* regionales.

Puede observarse que los efectos estimados de ETC son muy similares entre las columnas (1) y (3) y entre la (2) y la (4). Es decir, no se pierde información al quedarse con un número relativamente reducido de variables que son significativas: género, *background* socioeconómico y cultural del estudiante; *background* socioeconómico y cultural de sus pares; proporción de docentes titulados, si el centro de de educación técnica; si el centro es público y las *dummies* regionales. El resto de las variables evaluadas no afecta simultáneamente a ETC y a los resultados de las competencias evaluadas.

¹² Se incluyen como variables de control: género, nivel socioeconómico y cultural del estudiante (índice PISA) nivel socioeconómico promedio de los pares (promedio del índice PISA de los estudiantes del centro), ratio de estudiantes por profesor y su cuadrado; índice PISA de calidad de los recursos educativos; índice PISA de escasez de docentes; índice PISA de autonomía en asignación de recursos; índice PISA de autonomía curricular, porcentaje de docentes titulados en el centro; porcentaje de docentes con título universitario en el centro; el centro es de educación técnica; tamaño del centro en alumnos; y *dummies* regionales.

Cuadro 1: Estimadores MCO del efecto de haber asistido a ETC sobre los puntajes de las pruebas de Ciencias, Matemática y Lectura (en desvíos estándar). PISA Uruguay 2006

	MCO (1)	MCO (2)	MCO (3)	MCO (4)
Efectos homogéneos: Asistió a ETC				
Ciencias	-0.301***	-0.267***	-0.293***	-0.258***
Matemática	-0.279***	-0.236***	-0.270***	-0.226***
Lectura	-0.226***	-0.178***	-0.222***	-0.179***
Efectos heterogéneos: Asistió a ETC y padres con hasta educación primaria				
Ciencias				
(1) Asistió a ETC				-0.325***
(2) Asistió a ETC y padres con hasta primaria				0.259**
P>t test (1) + (2) =0				0.525
Matemática				
(1) Asistió a ETC				-0.276***
(2) Asistió a ETC y padres con hasta primaria				0.191**
P>t test (1) + (2) =0				0.285
Lectura				
(1) Asistió a ETC				-0.223***
(2) Asistió a ETC y padres con hasta primaria				0.172
P>t test (1) + (2) =0				0.592

Notas: MCO (1): incluye como variables de control: género, nivel socioeconómico y cultural del estudiante (índice PISA) nivel socioeconómico promedio de los pares (promedio del índice PISA de los estudiantes del centro), ratio de estudiantes por profesor y su cuadrado; índice PISA de calidad de los recursos educativos; índice PISA de escasez de docentes; índice PISA de autonomía en asignación de recursos; índice PISA de autonomía curricular, porcentaje de docentes titulados en el centro; porcentaje de docentes con título universitario en el centro ; el centro es de educación técnica; el centro es público; tamaño del centro en alumnos; y dummy indicadora de si el estudiante reside en Montevideo.

MCO(2): incluye las mismas variables que MCO(1) excepto que sustituye dummy de Montevideo por un conjunto de dummies indicadoras del departamento de residencia.

MCO(3): incluye sólo las variables significativas de la especificación MCO(1): género, nivel socioeconómico y cultural del estudiante, nivel socioeconómico promedio de los pares, porcentaje de docentes titulados en el centro; el centro es de educación técnica; el centro es público y dummy indicadora de si el estudiante reside en Montevideo.

MCO(4): incluye las mismas variables que MCO(3) excepto que sustituye dummy de Montevideo por un conjunto de dummies indicadoras del departamento de residencia.

Las variables índices se encuentran estandarizadas (media 0 y varianza 1).

P>t indica la probabilidad de que la suma de los coeficientes (1) y (2) (el efecto de ETC sobre los que provienen de hogares de bajo nivel educativo) sea igual a cero.

Número de casos: 4016 en MCO(1) y (2); 4171 en MCO(3) y (4).

*(**)[***] indica significación al 10%(5%)[1%]. El cálculo del desvío estándar de cada estimador se realiza mediante el método de Replicación Repetido Balanceado (BRR por sus siglas en inglés). Se utilizan 80 replicaciones.

Fuente: Elaborado en base a PISA Uruguay, 2006.

De las características incluidas en los modelos (3) y (4), la proporción de docentes titulados, si el centro es de educación técnica y si el centro es público no están correlacionados con haber asistido a ETC, condicional a las otras variables

estadísticamente significativas incluidas en la regresión. No obstante, tienen incidencia sobre los resultados finales de las pruebas, por lo que incluirlas en las estimaciones MCO reduce la varianza del estimador del impacto de ETC.¹³

La inclusión de *dummies* de departamento afecta levemente la estimación puntual del efecto de ETC sobre las competencias evaluadas (compárese columna 3 con 4), lo que indica que existen factores regionales que están correlacionados simultáneamente con haber asistido a ETC y con los resultados obtenidos. Dado todo el análisis anterior, se selecciona la especificación de la columna (4), la que se utiliza para evaluar efectos heterogéneos por MCO y también para el resto de las metodologías que se exponen más adelante. De todas maneras, puede observarse que el efecto estimado de haber asistido a ETC sobre las competencias de las tres áreas evaluadas es negativo y significativo en todas las especificaciones MCO, condicional al set de observables estadísticamente relevante.

A efectos de identificar posibles impactos heterogéneos se incluyó una especificación con una interacción de haber asistido a una ETC y que los padres del estudiante tengan hasta primaria como máximo nivel educativo completado.¹⁴ Puede observarse que el coeficiente asociado a la interacción es positivo y significativo en todos los casos, excepto en Lectura, lo que indicaría efectos diferenciales de asistir a ETC en la población proveniente de hogares de bajo nivel educativo (véase panel inferior el cuadro 1). En este caso, no es posible en ningún caso rechazar la hipótesis de que el impacto total de haber asistido a ETC para los asistentes a ETC con padres de bajo nivel educativo sea nulo.

V.2 Matching

En primer lugar se estima el *propensity score*, a partir de variables que se correlacionan simultáneamente con la asistencia a Tiempo Completo y con los resultados educativos. Como ya fue mencionado antes, el set de covariables encontrado es reducido: se compone del índice de nivel sociocultural del estudiante, del promedio de sus pares (como indicador del entorno del estudiante y/o de la selección de estudiantes en centros de educación media de distintas características) y el departamento de residencia (que recoge características inobservables específicas de las distintas zonas del país). El cuadro 2 presenta un conjunto de indicadores del balance de las covariables mencionadas, resultado de diferentes métodos de emparejamiento.

Todos los métodos de emparejamiento analizados balancean en forma muy aceptable las covariables en los dos grupos de comparación. Las columnas (2) y (4) muestran que luego del emparejamiento, la asistencia a ETC no se explica por el conjunto de características consideradas. Asimismo, la media y mediana del sesgo en las características se reduce considerablemente después del *matching*, en todos los métodos utilizados. Finalmente, el

¹³ Cabe resaltar que el efecto de estas variables sobre el resultado de las pruebas no se interpreta como un efecto causal, dado que también pueden estar afectadas por otros problemas de endogeneidad. Su inclusión en las regresiones solamente tiene el objetivo de mejorar la precisión del efecto estimado de ETC.

¹⁴ Recuérdese que tal como se citó antes, el estudio de Cerdan Infantes y Vermeesch (2007) encuentra un efecto positivo de ETC sobre los aprendizajes en primaria, mayor en las escuelas de contexto más desfavorable.

porcentaje de individuos que queda fuera del soporte común es prácticamente despreciable en todos los casos.¹⁵ A partir de los indicadores expuestos en el cuadro 2 se elige el *matching* basado en el método de distancia máxima de radio 0.005, dado que es el que presenta los mejores indicadores de balance de las covariables utilizadas, sin perder casi observaciones de los individuos tratados.¹⁶

Cuadro 2. Indicadores de balance de las covariables antes y después del matching, según método.

Método de emparejamiento	Pseudo R2		p>chi2		Media del sesgo		Mediana del sesgo		% perdido p soporte común
	Antes del matching	Después del matching	Antes del matching	Después del matching	Antes del matching	Después del matching	Antes del matching	Después del matching	
5 vecinos más cercanos	0.076	0.003	0.000	1.000	12.7	1.9	6.3	2.1	0.02%
10 vecinos más cercanos	0.076	0.002	0.000	1.000	12.7	1.6	6.3	1.5	0.02%
Distancia max (rc 0.001)	0.076	0.001	0.000	1.000	12.7	1.1	6.3	0.7	0.26%
Distancia max (rc 0.005)	0.076	0.000	0.000	1.000	12.7	0.7	6.3	0.6	0.02%
Kernel	0.076	0.004	0.000	1.000	12.7	2.3	6.3	1.3	0.02%

Notas: Las covariables utilizadas son: nivel socioeconómico y cultural del estudiante (índice PISA) nivel socioeconómico promedio de los pares (promedio del índice PISA de los estudiantes del centro), y *dummies* indicadoras del departamento de residencia.

Pseudo R2 (antes): es el Pseudo R2 de la estimación *probit* de la probabilidad de asistir a ETC condicional a las covariables utilizadas.

Pseudo R2 (después): es el Pseudo R2 de la estimación *probit* de asistir a ETC en la muestra emparejada en $p(X)$.

$p>chi2$: es el p-valor del test de razón de verosimilitud, testando la hipótesis de que los regresores NO son significativos en conjunto para explicar la asistencia a ETC; o sea están bien balanceados en los dos grupos.

Media del sesgo: Sesgo absoluto medio estandarizado antes y después del matching.

Mediana del sesgo: Idem al anterior, sobre la mediana.

Los estimadores de distancia máxima consideran a todos los vecinos dentro del radio especificado.

El tipo de Kernel utilizado es el Epanechnikov

Fuente: Elaborado en base a PISA Uruguay, 2006.

El cuadro 3 muestra la estimación del ATT de haber asistido a ETC sobre los resultados de las pruebas PISA para el método elegido. Al igual que las estimaciones por MCO, el ATT

¹⁵ El método para identificar el soporte común es excluir a los tratados cuyo PS es mayor al máximo PS de los controles.

¹⁶ Este resultado es casi siempre el esperable, ya que el método de distancia máxima es de mejor calidad que el vecino más cercano, pero como contracara, puede perder más observaciones de tratados. Sin embargo, en el caso del método de distancia máxima de radio 0.005 caso solamente se pierde una observación de los tratados.

estimado por el método de *matching* es negativo y significativo para las tres variables de resultado analizadas.

Cuadro 3: Estimadores de PSM del efecto de haber asistido a ETC sobre los puntajes de las pruebas de Ciencias, Matemática y Lectura (en desvíos estándar). Método de emparejamiento por distancia máxima, radio 0.005. PISA Uruguay 2006

	Bootstrap			P>z	Normal-based	
	ATT	Std. Err.	z		[95% Conf. Interval]	
Ciencias	-0.294	0.0509	-5.78	0.000	-0.394	-0.195
Matemática	-0.273	0.0412	-6.63	0.000	-0.354	-0.193
Lectura	-0.239	0.0467	-5.13	0.000	-0.331	-0.148

Nota: Las covariables utilizadas son: nivel socioeconómico y cultural del estudiante (índice PISA) nivel socioeconómico promedio de los pares (promedio del índice PISA de los estudiantes del centro), y *dummies* indicadoras del departamento de residencia.

Fuente: Elaborado en base a PISA Uruguay, 2006.

La estimación puntual del ATT es en valor absoluto mayor a la obtenida por MCO (comparando con la especificación elegida, véase columna 4 del cuadro 7), si bien ambas estimaciones no son estadísticamente diferentes al 95% de confianza.

Como se discutió en la sección anterior, no se esperan diferencias importantes en la estimación del ATT entre MCO y *matching* si no hay problemas de soporte común o solapamiento; o si los *propensity scores* están por debajo de 0.5.¹⁷ En particular, en los datos utilizados el soporte común no es una restricción; solamente 0.2% de los casos quedan fuera. Además, todos los *propensity scores* de la muestra están por debajo de 0.5

En definitiva, de la comparación de los estimadores de MCO y *matching* surge que no habría importantes fuentes de sesgo *en observables* que podrían derivarse de insuficiente solapamiento de las características observables que influyen simultáneamente en la asistencia a ETC y en el resultado entre ambos grupos; y/o la forma funcional impuesta para la estimación MCO.

V.3 Variables instrumentales

El cuadro 4 presenta los estadísticos y p-valores asociados a distintas pruebas de subidentificación y a la prueba de instrumentos débiles de Stock y Yogo.¹⁸ Se presentan las pruebas para dos tipos de modelos: para el modelo lineal en dos etapas (ajustando en la primera etapa un modelo de probabilidad lineal de la asistencia a ETC sobre la distancia del centro PISA a la ETC más cercana) y para un modelo de variables instrumentales no lineal, donde se estima en una etapa anterior un modelo *probit* de asistencia a ETC sobre la distancia y el resto de las variables, y el *valor predicho* se utiliza como instrumento de asistir a ETC (a través de un MCO en dos etapas).

¹⁷ En este caso el esquema de ponderación de MCO es proporcional al de *matching*. Véase Blundell, Dearden y Dianesi, 2004.

¹⁸ Las pruebas se computan a través del comando de Stata `ivreg2`.

Puede observarse que en ambos casos los instrumentos superan las pruebas de subidentificación de Anderson y de Cragg Donald¹⁹ (en ambos casos se rechaza la hipótesis nula de subidentificación), lo que indica que existe una relación significativa entre el instrumento (la distancia a la ETC más cercana) y la variable endógena (haber asistido a una ETC la mayor parte de la primaria).

Cuadro 4: Pruebas de subidentificación y de instrumentos débiles asociado a distancia en metros (en logs.) y al valor predicho de ETC según modelo probit

	Distancia en metros (en logs)	Valor predicho de ETC (1)
Pruebas de subidentificación		
Prueba canónica de Anderson (Chi2(1))	13.99	49.64
P valor	0.000	0.000
Prueba de Cragg Donald	14.04	50.24
P valor	0.000	0.000
Prueba de instrumentos débiles		
Prueba de Cragg Donald Wald F statistic	13.948	49.92
Stock-Yogo weak ID test critical values: 10% maximal IV size	16.38	
15% maximal IV size	8.96	
20% maximal IV size	6.66	
25% maximal IV size	5.53	

Notas: Se incluyen las covariables: género, nivel socioeconómico y cultural del estudiante, nivel socioeconómico promedio de los pares, porcentaje de docentes titulados en el centro; el centro es de educación técnica; el centro es público y dummies indicadoras del departamento de residencia. El valor predicho de ETC surge del modelo de asistencia a ETC sobre el instrumento distancia y el resto de variables exógenas que se presenta en el cuadro A.10 del Anexo.

Fuente: Elaborado en base a PISA Uruguay, 2006 y base de distancias elaborada por CINVE.

Si bien la condición de identificación es necesaria, es deseable que el instrumento tenga un buen poder de predicción de la variable endógena, condicional al resto de las variables exógenas. El contraste de este aspecto se realiza mediante la prueba de instrumentos débiles de Stock y Yogo.²⁰ La primera columna del cuadro 4 muestra el estadístico de Cragg Donald asociado a la variable distancia del centro PISA a la ETC más cercana. Puede observarse que el estadístico supera al valor crítico a partir de 15% máximo de sesgo “deseado”. Es decir, la distancia del centro PISA a la escuela de tiempo completo más cercana no es lo suficientemente fuerte como para no generar un sesgo mayor a 10%, si bien podría asegurarse que el sesgo no sería mayor a 15%.

¹⁹ Ambas pruebas contrastan si la matriz de coeficientes de la primera etapa (el modelo de asistir a ETC) tiene rango completo. Si ese es el caso, los instrumentos son relevantes. Los tests se diferencian en que construyen un estadístico distinto.

²⁰ El estadístico se construye en base a la relación de los coeficientes estimados por MCO y MCO en dos etapas, y un tamaño máximo de sesgo deseado.

En base a estos resultados se optó por utilizar el valor predicho de ETC como variable instrumental, en lugar de la distancia en forma lineal. De todas maneras, el test de instrumentos débiles sobre la variable de distancia incorporada linealmente en la primera etapa es un aspecto delicado que no debe ser soslayado, dado que los instrumentos débiles tienden a aproximarse al estimador (sesgado) MCO o incluso amplificar el sesgo, inclusive en muestras grandes (Staiger y Stock (1994)). La debilidad del instrumento utilizado puede provenir por varias razones. En primer lugar, recuérdese que se utiliza la distancia del centro PISA a la escuela de tiempo completo más cercana, disponible en el momento en que el estudiante cursó la primaria, buscando con ello una fuente de variación exógena en la asistencia a Tiempo Completo (la proximidad a una escuela de estas características, condicional al resto de los observables que influyen sobre la participación). Sin embargo, el indicador no sería relevante para aquellos estudiantes que asisten a un centro de educación media en una zona alejada de su lugar de residencia, o para aquellos que cambiaron de residencia hacia otra zona entre el momento que cursaron la primaria y la educación media, si bien como fue mencionado antes, parte de esta movilidad entre grandes zonas se tiene en cuenta. Además, al utilizarse como unidad de referencia al centro, todos los estudiantes que asisten al mismo centro PISA (y están en el mismo grado) tienen asignado el mismo valor del instrumento, lo que limita la variabilidad exógena que justamente se quiere captar con cualquier instrumento.

En definitiva, presumiblemente el instrumento es válido (en el sentido de que, condicional a las variables observables no estaría correlacionado con el error de la ecuación de resultado, o dicho de otra manera, no es esperable que la variable distancia influya sobre el resultado final sino solo a través de su influencia sobre la variable endógena); y también existe una relación significativa con la variable endógena, de los *tests* expuestos no puede rechazarse que el instrumento sea débil, es decir, que su poder de predicción (al incorporarse en forma lineal) sea bajo.

El cuadro 5 presenta los estimadores IV del efecto de haber asistido a ETC sobre los resultados de ciencias, matemática y lectura. En el caso de ciencias y matemática, puede observarse que si bien el signo del coeficiente cambia respecto a los estimadores por MCO y *matching*, el tamaño del desvío estándar no permite rechazar la hipótesis nula de que los efectos sean nulos. En tanto, en lectura el efecto estimado es negativo y significativo. En este último caso, si bien la estimación puntual es muy superior en valor absoluto del estimador por MCO y *matching*, el desvío estándar es también muy superior, por lo que los intervalos de ambas estimaciones se solapan.

Cuadro 5. Estimadores IV del efecto de haber asistido a ETC sobre los puntajes de las pruebas de Ciencias, Matemática y Lectura (en desvíos estándar). PISA Uruguay 2006

	IV
Ciencias	0.113 [0.471]
Matemática	0.495 [0.711]
Lectura	-1.535** [0.749]

Notas:

Se incluyen las covariables: género, nivel socioeconómico y cultural del estudiante (índice PISA), nivel socioeconómico promedio de los pares; porcentaje de docentes titulados en el centro; el centro es de educación técnica; el centro es público; y conjunto de dummies indicadoras del departamento de residencia .

Instrumento para haber asistido a ETC: Se ajusta un modelo Probit para asistencia a ETC sobre la distancia del centro PISA a la ETC mas cercana (en logaritmos) y covariables exógenas. Se utiliza la predicción como instrumento para la segunda etapa.

Número de casos: 4171

*(**)[***] indica significación al 10%(5%)[1%].

Los datos entre paréntesis [] corresponden al desvío estándar de cada estimador, que se estima mediante el método de Replicación Repetido Balanceado (BRR por sus siglas en inglés). Se utilizan 80 replicaciones.

Fuente: Elaborado en base a PISA Uruguay, 2006 y base de distancias elaborada por CINVE.

V.4 Función de control

Finalmente, en este apartado se presentan estimaciones basadas en el método de función de control. Como ya fue mencionado, este método permite estimar un efecto promedio de la política evaluada. Asimismo, la evaluación de efectos heterogéneos con la metodología de funciones de control es más eficiente que la de IV; se incluye la interacción ZX en la ecuación de participación, y la interacción $ETCX$ en la ecuación de resultado, pero en la primera etapa solamente se requiere predecir el inverso del ratio de mills (λ), al igual que en el caso de efectos homogéneos. Si el modelo de decisión está bien especificado, al condicionar por la función de control el término de error de la ecuación de resultado se vuelve ortogonal a los regresores de interés. Obviamente, esta estimación requiere imponer una mayor estructura y supuestos sobre la distribución conjunta de los inobservables que afectan la participación y el resultado.

El cuadro 6 (panel superior) muestra las estimaciones de función de control del efecto de haber asistido a ETC sobre los resultados de las pruebas evaluadas en PISA, sin considerar la posibilidad de efectos heterogéneos.

Cuadro 6: Estimadores de Función de Control del efecto de haber asistido a ETC sobre los puntajes de las pruebas de Ciencias, Matemática y Lectura (en desvíos estándar). PISA Uruguay 2006

	FC (MV)
Efectos homogéneos: Asistió a ETC	
Ciencias	
Asistió a ETC	0.640***
Athrho	-0.643***
Matemática	
Asistió a ETC	0.337*
Athrho	-0.384***
Lectura	
Asistió a ETC	-1.072***
Athrho	0.268
Efectos heterogéneos: Asistió a ETC y padres con hasta educación primaria	
Ciencias	
(1) Asistió a ETC	0.337
(2) Asistió a ETC y padres con hasta primaria	0.223**
P>t test (1) + (2) =0	0.054
Athrho	-0.460**
Matemática	
(1) Asistió a ETC	0.129
(2) Asistió a ETC y padres con hasta primaria	0.182
P>t test (1) + (2) =0	0.152
Athrho	-0.274*
Lectura	
(1) Asistió a ETC	-1.077***
(2) Asistió a ETC y padres con hasta primaria	0.250**
P>t test (1) + (2) =0	0.000
Athrho	0.341

Notas: Se incluyen las covariables: género, nivel socioeconómico y cultural del estudiante, nivel socioeconómico promedio de los pares, porcentaje de docentes titulados en el centro; el centro es de educación técnica; el centro es público y dummies indicadoras del departamento de residencia. Número de casos: 4171

La función de control: distancia del centro PISA a la ETC más cercana como restricción de exclusión; esta variable interactuada con padres de nivel educativo hasta primaria completa en la estimación de efectos heterogéneos.

*(**)[***] indica significación al 10%(5%)[1%].

Se estima por máxima verosimilitud.

Athrho es el inverso de la tangente hiperbólica de rho, $atanh\rho = \frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right)$ que se estima por razones de estabilidad numérica en el procedimiento de estimación por máxima verosimilitud. Su significancia estadística, signo y magnitud indican presencia de selección en inobservables.

Fuente: Elaborado en base a PISA Uruguay 2006, y base de distancias elaborada por CINVE.

El coeficiente ligado a ρ^{21} (la correlación entre los inobservables de la ecuación de resultado y de asistencia a ETC) muestra un signo negativo y significativo en los modelos de ciencias y matemática, y no significativamente distinto de cero en lectura. Ello indicaría la presencia de factores inobservables asociados positivamente (negativamente) con haber asistido a tiempo completo y negativamente (positivamente) con los resultados educativos posteriores (en ciencias y matemática, estrictamente).

Al corregir por ese eventual sesgo, el efecto de haber asistido a tiempo completo se vuelve positivo y significativo en ambas pruebas (en 0.64 y 0.34 desvíos estándar, respectivamente). Sin embargo, el efecto asociado a lectura continúa estimándose negativo y de hecho no es posible rechazar la hipótesis de que no hay sesgo de selección en inobservables, lo que validaría en este caso una estimación por MCO o *matching*.

Al incorporar la posibilidad de efectos heterogéneos el signo y significación del coeficiente indicador de la existencia de sesgo de selección en inobservables no cambia en ningún caso. No obstante, las estimaciones respecto a la existencia de efectos diferenciales arrojan resultados diversos. En el caso de matemática, el coeficiente asociado a la variable interactuada no es significativamente distinto de cero, lo que no permite avalar la hipótesis de efectos diferenciales. Las estimaciones referidas al impacto en la prueba de lectura indican un impacto diferencial, pero dado que se rechaza la hipótesis de selección en inobservables, se debería remitir el análisis del estimador por MCO. Finalmente, las estimaciones referidas a ciencias parecen indicar un efecto positivo y significativo para estudiantes provenientes de hogares de bajo nivel educativo, mientras que el efecto para el resto de los estudiantes, si bien puntualmente se estima positivo, pierde significación estadística.

En definitiva, considerando la evaluación del efecto de haber asistido a ETC sobre las competencias en matemática y ciencias, las estimaciones por función de control no permiten rechazar la hipótesis de existencia de sesgo de selección en inobservables y esbozan indicios de un efecto positivo de la política sobre los aprendizajes en dichas áreas, al controlar por este tipo de sesgo. La estimación incorporando efectos diferenciales no arroja resultados robustos, si bien debe tenerse en cuenta que el poder de predicción del instrumento interactuado con la educación de los padres es bajo y puede limitar la validez de estas últimas conclusiones. En cuanto al efecto de la política sobre las competencias en lectura, los resultados indican que se rechaza la hipótesis de selección en inobservables, por lo que las estimaciones por MCO o *matching* no adolecerían de este tipo de sesgo. Si ello es cierto, no puede refutarse la hipótesis de que el efecto de haber asistido a ETC sobre las competencias en lectura sea negativo.

Los diferentes resultados ligados al tipo de competencia sobre la que se evalúa el impacto de la política de tiempo completo están parcialmente en línea con los resultados encontrados por Cerdan Infantes y Vermeesch (2007). Cuando estos autores utilizan toda la muestra de escuelas de tiempo completo, no encuentran resultados estadísticamente significativos en la prueba de lengua, aunque sí en matemática, evaluadas al final de la

²¹ Se presenta la transformación reportada de ρ .

primaria. El impacto que estiman sobre los aprendizajes en matemática es de 0.07 de desvío estándar por año y de 0.38 desvíos estándar en seis años de participación en el programa. En este trabajo, la estimación por funciones de control arroja un impacto total estimado para esta área de 0.34 desvíos estándar.

En cuanto a la prueba de lengua, Cerdan Infantes y Vermeesch encuentran un efecto positivo de ETC solamente cuando excluyen a las escuelas de tiempo completo de contextos más favorables de la muestra evaluada, mientras que sería nulo para el total de la muestra. Ello podría indicar que la política tiene impactos en los aprendizajes de lengua solamente para niños de contextos más desfavorables, pero no para toda la población tratada. En este trabajo, el efecto principal evaluado corresponde al promedio de los tratados, por lo que si se hubieran encontrado impactos nulos sobre la prueba de lectura, sería un resultado bastante consistente con los encontrados en el estudio anteriormente citado²². No obstante, llama la atención que el impacto estimado en la prueba de lectura sea negativo, resultado muy difícil de explicar.

Finalmente, es posible que existan factores inobservables que incidan en los resultados evaluados que no hayan podido recogerse con el tipo de datos y el instrumento utilizado. En particular, la existencia de efectos diferenciales está limitadamente evaluada, debido a que los datos disponibles no permiten realizar mayores desagregaciones; tampoco está tratada la selectividad en la asistencia al sistema educativo a los 15 años, que de estar asociada a asistir a ETC podría introducir sesgos en los impactos estimados. Finalmente, debe tenerse en cuenta que el instrumento utilizado también adolece de debilidades, en particular de un relativamente bajo poder de predicción de la variable instrumentada. Ello limita su alcance para incorporar estimaciones de efectos diferenciales, e incrementa la varianza de los estimadores IV, lo que reduce la precisión de las estimaciones.

VI. Conclusiones

En este trabajo se exploró la utilización de cuatro métodos no experimentales que se basan en distintos supuestos para evaluar el efecto de las escuelas de tiempo completo sobre las competencias adquiridas a los 15 años, en el nivel de la educación media. A partir del análisis realizado pueden resaltarse las siguientes conclusiones.

No se encontró evidencia de que exista un sesgo en observables, derivado de insuficiente solapamiento en las distribuciones de las características observables entre asistentes y no asistentes a ETC; ni debido a una ponderación errónea de las variables observables dentro del solapamiento. En definitiva, si bien la distribución de las características observables entre asistentes y no asistentes a ETC difiere significativamente, la casi totalidad de individuos de la muestra PISA con las mismas características observables relevantes tienen una probabilidad positiva tanto de asistir como de no asistir a ETC; lo que está ligado a que la política no se focalizó exclusivamente en los contextos más desfavorables, ni en regiones

²² Si bien las pruebas no son estrictamente comparables, corresponden a una misma área del conocimiento.

específicas. Esta condición de amplio soporte común (y que los *propensity scores* estimados están por debajo de 0.5) hace que las estimaciones por MCO y *matching* no difieran significativamente.

Una gran parte de la varianza en la asistencia a ETC no puede ser explicada por las variables utilizadas, y más en general, por las variables contenidas en la base de datos de PISA; es decir, que está determinada por factores inobservables.

Las estimaciones por funciones de control (y en menor medida, por IV) para ciencias y matemática sugieren la presencia de factores inobservables correlacionados positivamente con haber asistido a ETC y negativamente con los resultados de dichas pruebas. En el caso de lectura no se encontró evidencia de la existencia de este tipo de sesgo.

Un indicio de posible fuente de sesgo es la mayor tasa de repetición en primaria que se observa en los estudiantes que asistieron a tiempo completo la mayor parte de primaria; y que no se condice los registros oficiales de la ANEP. Como se discutió, ello podría indicar la existencia de un sesgo de selección de niños con experiencias de repetición *previas* a entrar en ETC. En el caso de que hubiera una parte exógena en la variable de repetición en primaria, y dado que está negativamente correlacionada con los aprendizajes, no incluirla sesga a la baja la estimación del efecto de ETC por *MCO o matching*. Dada la influencia de esta variable sobre los resultados y la posibilidad de que estuviera indicando un sesgo de entrada a tiempo completo, sería deseable contar con información más precisa de cada estudiante, que permitiera identificar más adecuadamente su historial previo a la entrada a la política de ETC; y más en general, a toda política de alcance significativo que se implementa en el sistema educativo.

Al tomar en cuenta el posible sesgo de selección en inobservables, el impacto de haber asistido a ETC se vuelve positivo en ciencias y en matemática. Cuando las estimaciones se realizan por IV, si bien la estimación puntual es positiva, la varianza del estimador no permite rechazar la hipótesis de que los efectos sean nulos. Cuando la estimación se realiza por funciones de control (sin considerar efectos heterogéneos), se encuentra un impacto positivo y significativo de la política sobre los aprendizajes de ambas áreas. El tamaño del efecto en matemática (0.34 desvíos estándar) es consistente con el encontrado por Cerdan Infantes y Vermeesch, en la evaluación realizada sobre los impactos de ETC al final de la primaria. El impacto en ciencias es mayor (0.64 desvíos estándar). En el caso de lectura no se encontró evidencia de sesgo de selección en inobservables, y no puede rechazarse la hipótesis de que el efecto sobre esta área de aprendizaje sea negativo.

Si bien existen algunos indicios de que los efectos podrían ser diferenciales para estudiantes de hogares con bajo nivel educativo no fue posible encontrar evidencia robusta sobre ellos. La metodología de variables instrumentales sobre exige al instrumento, y no es posible encontrar dos instrumentos relevantes para las dos variables endógenas. En el caso de funciones de control, si bien la exigencia que se le impone al instrumento es menor, los supuestos y la estructura impuesta son más fuertes: en ese caso la evidencia no es muy robusta.

Las estimaciones aquí presentadas no están exentas de problemas. En particular, el bajo poder predictivo del instrumento utilizado genera una varianza considerable de los estimadores IV, lo que reduce su precisión. El bajo poder predictivo del instrumento se asocia en parte a que refiere a la distancia del centro PISA a la ETC más cercana, y no a la distancia del hogar de residencia del estudiante. Lamentablemente no se dispone de esta última información, que presumiblemente mejoraría el poder predictivo de la variable instrumental.

Otro tipo de problema es el asociado al sesgo de selección muestral ocasionado por la no observación de alumnos que abandonaron los estudios a los 15 años. Si la política evaluada tiene un efecto positivo sobre la continuación de los estudios, el hecho de no observar a los que abandonaron el sistema educativo puede sesgar (presumiblemente a la baja) las estimaciones del impacto de la política. En este caso, los alumnos asistentes a ETC parecen estar sobre representados en la muestra de PISA, lo que indica que este tipo de sesgo puede estar presente. No obstante, los datos disponibles no permiten tratar este problema sin introducir nuevos supuestos y sin recurrir a otras fuentes de datos que pueden introducir incluso más distorsiones.

Otra dificultad que surge cuando se quiere evaluar la política de tiempo completo es la propia diversidad de la implementación de la política. Si bien todas las ETC comparten la extensión del tiempo de clase (además de la extensión de la alimentación y la capacitación de los maestros), difieren de *a qué* se destina el tiempo adicional: la implementación de talleres de expresión, de inglés u otras actividades varió significativamente entre escuelas. Como el efecto de la extensión del tiempo de clase depende efectivamente de a qué se dedique ese tiempo extra, el impacto sobre los aprendizajes en distintas áreas puede diferir entre alumnos que asistieron a distintas escuelas. En definitiva, los efectos encontrados probablemente escondan diferencias asociadas a la heterogeneidad de la implementación de la política. Para poder evaluar más precisamente los efectos asociados a distintos componentes o formas de implementación se precisa información mucho más detallada por escuela y en el tiempo.

Finalmente, cabe notar que se ha procurado evaluar la política de tiempo completo en *un* set de resultados sobre los cuales se espera que tenga impacto: la adquisición de competencias en algunas áreas básicas de conocimiento. No obstante, existen otros resultados buscados por la política, ligados básicamente al desarrollo socioemocional de los niños, que no han sido evaluados y sobre los que también sería deseable conocer sus efectos.

En suma, las estimaciones realizadas procuraron poner énfasis en los distintos tipos de sesgo que pueden estar presentes cuando se quiere estimar el impacto de una política sobre los resultados en la educación media, utilizando la base de datos disponible (las pruebas PISA) y datos complementarios. Se identificaron varias dificultades para estimar impactos y se discutieron las posibles fuentes de sesgo. No obstante, dada la importancia de la política específica evaluada, sería deseable contar con otro tipo de información (por ejemplo, datos longitudinales) que permita estimar los impactos en forma más robusta.

VI. Bibliografía

- ANEP (2007: a): "Informe Nacional PISA 2006 Uruguay"
- ANEP (2007: b): "Relevamiento de Características Socioculturales de las escuelas públicas del Consejo de Educación Primaria, 2005", Área de Investigación y Estadística Educativa, Dirección de Investigación, Evaluación y Estadística de CODICEN.
- ANEP (2000): "Evaluación Nacional de Aprendizajes en Lengua y Matemática. Sexto año de Primaria, 1999"
- Angrist, J. y Krueger, A. (2001): "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments", *Journal of Economic Perspectives* 15, 4: 69-85.
- Bellei, C. (2009): "Does Lengthening the School Day Increase Students' Academic Achievement? Results from a Natural Experiment in Chile", *Economics of Education Review* 28: 629-640.
- Bernal, R. y Peña, X. (2011): "Guía Práctica para la Evaluación de Impacto". Universidad de los Andes, Facultad de Economía. Ediciones Uniandes, 2011.
- Blundell, R., Dearden L. y Sianesi, B. (2004): "Evaluating the Impact of Education on Earnings in the UK: Models, Methods and Results from the NCDS", *The Institute for Fiscal Studies*, WP03/20.
- Brown, B. y Saks, D. (1986): "Measuring the Effects of Instructional Time on Student Learning: Evidence from the Beginning Teacher Evaluation Study." *American Journal of Education* 94: 480-500.
- Caliendo, M. y Kopeinig, S. (2005): "Some Practical Guidance for the Implementation of Propensity Score Matching." IZA Discussion Paper No.1588.
- Cannon, J., Jackowitz, A. y Painter, G. (2011): "The Effect of Attending Full-Day Kindergarten on English Learner Students." *Journal of Policy Analysis and Management* 30: 287-309.
- Card, D. (1993): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling". NBER WP 4483.
- Carneiro, P. y Heckman, J. (2003): "Human Capital Policy", in J. Heckman and A. Krueger, eds., *Inequality in America: What role for human capital policies?* Boston, MIT Press.
- Cerdan Infantes, P. y Vermeesch, C. (2007): "More Time is Better: An Evaluation of the Full Time School Program in Uruguay". World Bank Policy Research WP 4167, March.
- Coates, D. (2003): "Education Production Functions Using Instructional Time as an Input." *Education Economics* 11: 273-292.
- DeCicca, P. (2007): "Does Full-Day Kindergarten Matter? Evidence from the First Two Years of Schooling." *Economics of Education Review* 26: 67-82.
- Equipos Mori, (2001): "Estudio de Valoración Social de las Escuelas de Tiempo Completo", noviembre.
- Glewwe, P. Hanushek, E. Humpage, S. Ravina, R. (2011): "School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010". NBER WP 17554, October.
- Glewwe, P. y Kremer (2006): "Schools, Teachers and Educational Outcomes in Developing Countries: In *Handbook of Economics of Education*, edited by Eric A. Hanushek y Finis Welch. Amsterdam, North Holland: 943-1017.

- Heckman, J. (2008): "Schools, Skills, and Synapses". UCD Geary Institute. Discussion Paper Series, WP/33/2008.
- Heckman, J. (1997): "Instrumental Variables: A Study of Implicit Behavioural Assumptions Used in Making Program Evaluations", *Journal of Human Resources* 32 (3): 441-462.
- Heckman, J. (1979). "Sample Selection Bias as a Specification Error", *Econometrica* 47: 153-161.
- Heckman, J., Ichimura, H., y Todd, P. (1997): "Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies* 64: 605-654.
- Heckman, J., Ichimura, H., Smith, J. y Todd, P. (1998): "Characterizing selection bias using experimental data", *Econometrica* 66, 1017-98.
- Imbens, G. y Angrist, J. (1994): "Identification and Estimation of Local Average Treatment Effects". *Econometrica* 62 (2): 669-738.
- Imbens, G. y Wooldridge, J. (2008): "Recent Developments in the Econometrics of Program Evaluation". NBER WP 14251, August.
- Link, C. y Mulligan, J. (1986): "The Merits of a Longer School Day." *Economics of Education Review* 5: 373-381.
- Mendez, N. y Zerpa, M. (2010): "Desigualdad en las capacidades educativas. Los casos de Uruguay y Chile" Tesis para el título de grado de la Licenciatura en Economía de la Facultad de Ciencias Económicas y Administración, Universidad de la República
- OECD (2010): "PISA 2009 Results: What Students Know and Can Do- Student Performance in Reading, Mathematics and Science (Volume I)".
- OCDE (2003) "Pisa 2003: Manual de análisis de datos".
- Perera, M y Llambí, C. (2009): "La Función de Producción Educativa: el posible sesgo en la estimación de efectos "institucionales" con los datos PISA. El caso de las escuelas de Tiempo Completo", Documento de Trabajo DT 03/2009, CINVE.
- Rosenbaum, P., y Rubin, D. (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-50.
- Staiger, D. y Stock, J.H. (1994): "Instrumental variables regressions with weak instruments", NBER WP 151.
- Walston J. y West, J (2004): "Full-Day and Half-Day Kindergarten in the United States. Findings from the Early Childhood Longitudinal Study, Kindergarten Class of 1998.99", *The Education Statistics Quarterly*, Vol. 6, Issues 1 & 2
- Valenzuela, JP. (2005): "Partial Evaluation of a Big Reform in the Chilean Education System: From a Half Day to a Full Day Schooling". University of Michigan.

