

Universidad de Los Andes
Facultad de Economía
Maestría en Economía (PEG)

La economía de la educación y el aprendizaje estudiantil:

Incorporando el “backwash effect” a la literatura

Asesor: Juny Montoya

Presentado por: Víctor Camilo Fernández Espinosa, (200714195)

21 de Mayo de 2013

Resumen

La literatura académica y en particular los estudios sobre economía de la educación han olvidado que los resultados obtenidos al interior de las aulas dependen en parte de la evaluación que se realiza a los estudiantes. Existe ya una extensa literatura en el campo de la psicología y educación que investiga lo conocido como el “backwash effect” y su influencia en la enseñanza, sin embargo, este análisis no ha sido realizado en el campo de la economía. En este trabajo se analiza si realmente existe el “backwash effect” por medio de un modelo de ecuaciones estructurales y se estima hasta qué punto tiene influencia en nuestras variables de interés utilizando una muestra global y a nivel de Colombia. Se confirma la existencia del “backwash effect” pero se discute la magnitud de este efecto. Finalmente, se discuten los resultados sobre acciones de política.

Palabras Clave: Economía, educación, backwash, washback, modelo de ecuaciones estructurales, PISA

1. Introducción

La educación ha sido un tema de extensa cobertura para distintas disciplinas. Investigaciones a través del tiempo se han dedicado a intentar descubrir cómo las personas aprenden y cómo este proceso se puede mejorar o enfocar para un fin u otro. Para la economía la educación también es un tema importante, no sólo por tener el deber de educar a futuros economistas, sino también porque la educación de futuros trabajadores influencia directamente nuestras variables de interés como producción, empleo y crecimiento. Sin embargo, los estudios en la economía de la educación, si bien han realizado diversos estudios sobre los efectos de muchas variables sobre el aprendizaje, no han tenido en cuenta que estos resultados se ven afectados por como son recogidos, específicamente, el método de evaluación que se le realiza a los estudiantes a la hora de medir este aprendizaje, conocido en los campos de la psicología y educación como el “backwash effect”.

1.1. El efecto de “backwash”

Se conoce en educación como efecto de backwash o washback al efecto que tiene el medio de evaluación sobre el método de enseñanza de los profesores o aprendizaje de los estudiantes. Este efecto ocurre pues al estudiante darse cuenta que su método de evaluación se convierte en un componente importante de su proceso de aprendizaje, este cambia sus estrategias de estudio de manera a enfrentarse al método de evaluación de forma específica (Harris & Bell, 1986).

La literatura del “backwash effect” tiene dos áreas principales de estudio: la que se dedica al análisis de diferentes tipos de pruebas observando sus efectos en la enseñanza y la manera de aprender de los estudiantes, y los estudios en los que se toma una prueba específica y se intenta mejorar sobre ella (Cheng & Curtis, 2004). Este trabajo se limita a analizar la primera área, dejando adicionalmente de lado el efecto que la evaluación puede tener en la enseñanza.

La principal manera en que la literatura analiza de manera cuantitativa el método de estudio del estudiante es por el cuestionario de procesos de estudio (SPQ, por su nombre en inglés, Study Process Questionnaire) (Biggs, 1987b) el cual consiste en una serie de preguntas acerca de métodos de estudio diferentes que están clasificados en profundo, superficial y de logro. La primera busca entender lo que se está aprendiendo, la segunda, solamente busca memorizar de manera rutinaria lo que se considera importante y la tercera se adapta a lo que se considere requerido por el método de evaluación (Marton & Säljö, 1976).

Thomas & Bain (1982) realizan un estudio en el cual 92 estudiantes de psicología son presentados con un cuestionario de estrategias de estudio inmediatamente después de toda prueba (superficial) y ensayo (profundo) que deban hacer durante su primer semestre de carrera. Mediante un análisis de factores, ellos encuentran que los estudiantes no cambian su método de estudio frente a un cambio del método de evaluación, concluyendo que estas decisiones son consistentes en un estudiante y no estratégicas.

Tang (1994) realiza un experimento en la sección de fisioterapia en el politécnico de Hong Kong, donde, después de realizar el cuestionario de SPQ (relativo a el proceso de estudio de los estudiantes) a sus 158 participantes, realiza diferentes tipos de evaluación, y conduce un “análisis de camino” (caso específico de modelo estructural) para llegar a la conclusión de que estudiar de la manera “equivocada” (es decir, de manera profunda en una prueba de memorización o viceversa) acaba siendo contraproducente en los resultados de las evaluaciones.

Scouller (1998) realiza un experimento con una muestra de 206 estudiantes de segundo año de educación australianos, quienes completaron dos cuestionarios relativos a su manera de prepararse y su percepción con respecto a dos métodos de evaluación: un ensayo o una prueba de opción múltiple. Scouller implementa un análisis de regresiones múltiples y concluye que los estudiantes consideran la evaluación de opción múltiple como algo más superficial y estudian de manera más superficial, lo opuesto ocurre para los ensayos. Además, los estudiantes que no siguen este patrón, es decir que utilizaran métodos de estudio profundo para la evaluación de opción múltiple o métodos superficiales para el ensayo no tendrían tan buenos resultados.

Smith & Miller (2005) realizan también un estudio con 248 alumnos de psicología y negocios. En este estudio no se realiza la evaluación en sí, pero sí se hace la encuesta de procesos de estudio a los estudiantes y se les pide que la llenen considerando que están prestes a hacer una evaluación de opción múltiple o un ensayo. En este estudio no se encuentra un efecto del tipo de evaluación sobre el proceso de estudio de los estudiantes, al contrario de los anteriores.

1.2. Objetivo

Podemos ver entonces que los resultados de las investigaciones son opuestos, y llegan a conclusiones diversas acerca de la existencia del efecto de backwash en los métodos de evaluación que se utilizan. Este trabajo busca agregar a la investigación que se lleva haciendo en este área de varias maneras, primeramente, nota que la literatura empírica descrita anteriormente está limitada

a muestras pequeñas de estudiantes, por lo cual el aumento a una muestra mundial puede dar luz sobre la medida en la que estos resultados pueden ser generalizables. En este caso, se utilizará el conocimiento de los medios de evaluación empleados por los colegios para analizar el efecto de estos en las estrategias comunes de estudio de los estudiantes y el impacto de ambos sobre la nota que reciben estos estudiantes en las diferentes preguntas de la prueba PISA, de esta manera, se puede hablar de un efecto persistente del método de evaluación en evaluaciones posteriores. Además nos permitimos analizar métodos de evaluación adicionales a los que muestra la literatura, la que se limita a comparar sólo dos tipos de métodos de evaluación, principalmente ensayos y preguntas de opción múltiple. Finalmente, añadimos controles adicionales conocidos en la literatura económica que sabemos también influyen el desempeño estudiantil en una evaluación (estos controles serán mencionados).

¿Pero por qué esto debería ser un tema relevante para la economía, que parece ya estar bien con su manera de medir el logro estudiantil por tantos años? Para responder a esa pregunta tenemos que mirar por fuera del área de investigación al día a día laboral, en él podemos encontrar diversas quejas de la aparente falta de preparación de los estudiantes para trabajar¹. Esto entonces muestra que hay un desfase entre lo que se quiere de los estudiantes al salir de la entidad educativa y lo que estos han sido preparados para hacer, y esto podría verse explicado, en parte, por el hecho de que estamos incentivando la acumulación de un conocimiento en las entidades educativas diferente al que será exigido posteriormente, además de acostumbrar al estudiante a priorizar la información de una manera específica.

El objetivo general de la tesis es analizar la presencia del “backwash effect” cuando se utiliza una base de datos mundial, de manera a observar si el método de evaluación de los estudiantes tiene un efecto sobre su método de estudio y sobre su desempeño cuando enfrentado a preguntas de diferentes tipos, teniendo como objetivos específicos analizar si la mayor presencia de un método evaluativo en la educación del colegio de un estudiante tiene efectos en su nota cuando se le hace una prueba posterior con diferentes tipos de preguntas y analizar si el método evaluativo escolar tiene algún efecto en la manera cómo los estudiantes optan por estudiar.

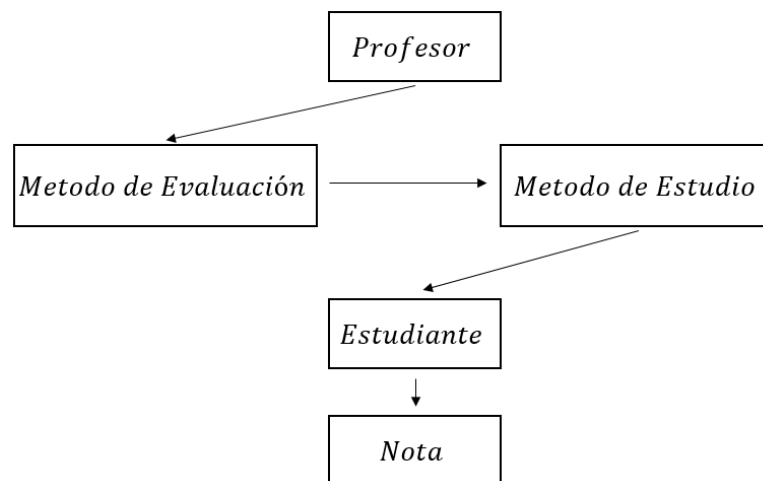
¹ Fuentes: http://www.p21.org/storage/documents/final_report_pdf09-29-06.pdf, <http://chronicle.com/article/Employers-Say-College/130013/>, <http://www.comercioyjusticia.com.ar/2012/05/03/los-jovenes-no-se-sienten-preparados-para-el-trabajo/>, entre otros. Recuperado el 15 de Mayo de 2012.

El resto de este trabajo se organiza de la siguiente manera: la sección 2 desarrolla un simple modelo de decisión con base a lo recogido en la literatura de manera a analizar desde un punto de vista microeconómico de maximización de utilidad, la decisión óptima de un estudiante a la hora de estudiar para una prueba. La sección 3 presenta la metodología de estimación a ser utilizada. La sección 4 explica los datos a ser utilizados en la estimación. La sección 5 muestra los resultados de las estimaciones. La sección 6 plantea una discusión con base a los resultados de la sección anterior. Finalmente, la sección 7 concluye.

2. Modelo teórico

2.1. Asumiendo la importancia de la evaluación

De la literatura del backwash podemos entonces recoger una relación simple entre nuestra variable de interés. Si asumimos que son los profesores los que escogen el método de evaluación a ser usado en clase de manera a dejar quietos los posibles efectos de backwash que los métodos de evaluación pueden tener en los métodos de enseñanza, tenemos entonces:



Es decir, los profesores escogen el método de evaluación el cual tiene un efecto en la manera de estudiar del estudiante, lo que tiene un efecto junto con las otras variables endógenas al estudiante (que serán exógenas en este modelo) y finalmente esto tiene un efecto en la nota obtenida.

Adicionalmente, la literatura nos dice que el estudio de manera profunda (deep) y superficial (surface) son mutuamente excluyentes y al mismo tiempo, estos diferentes métodos de estudio tienen rendimientos diferentes dependiendo del tipo de evaluación (Tang, 1994), por lo cual, el

problema al que se enfrenta un estudiante a la hora de maximizar su nota puede ser expresado de la siguiente manera:

$$N_i = \theta_d(\alpha_{di} * D_i + \beta_{di} * S_i)^{\gamma_i} + \theta_s(\alpha_{si} * D_i + \beta_{si} * S_i)^{\gamma_i} \quad (1)$$

Donde N_i es la nota del estudiante i , θ_d es la probabilidad de que la prueba que se le presente al estudiante sea del tipo d , θ_s es la probabilidad de que la prueba que se le presente al estudiante sea del tipo s (de tal manera que $\theta_d + \theta_s = 1$), D es el tiempo que le dedica el estudiante a estudiar de la forma profunda, S es el tiempo que le dedica el estudiante a estudiar de la forma superficial ($D_i + S_i = 1 \forall i$), y α_{ni} , β_{ni} son índices del rendimiento del estudio de un estudiante i dado que el examen sea de una manera n . Para hacer específico que el método de estudio D es más útil para la evaluación d y S para s , suponemos adicionalmente que $\alpha_{di} > \beta_{di}$ y $\alpha_{si} < \beta_{si}$. Finalmente $\gamma_i \in (0,1)$ es el índice de rendimientos decrecientes del estudio que es el mismo sin importar el tipo de prueba a la que se enfrente el estudiante. Es importante añadir el hecho de que se sabe que hay muchos más efectos que estos que afectan al estudiante que son internos a este (tales como el género, el grado de escolaridad de sus padres, etc.) usamos el subíndice i para dejar clara la existencia de estos factores, pero suponemos que estas diferencias no son lo suficientemente grandes para romper los supuestos previamente establecidos.

Reemplazando las restricciones de probabilidad y tiempo en nuestra función objetivo, tenemos que:

$$N_i = \theta_d[\alpha_d * D_i + \beta_d * (1 - D_i)]^{\gamma_i} + (1 - \theta_d)[\alpha_s * D_i + \beta_s * (1 - D_i)]^{\gamma_i} \quad (2)$$

Maximizando con respecto a D_i y encontrando las condiciones de primer orden tenemos:

$$\frac{\partial N_i}{\partial D_i}: \theta_d * \gamma_i * (\alpha_d * D_i + \beta_d * (1 - D_i))^{\gamma_i - 1} * (\alpha_d - \beta_d) + (1 - \theta_d) * \gamma_i * (\alpha_s * D_i + \beta_s * (1 - D_i))^{\gamma_i - 1} * (\alpha_s - \beta_s) = 0 \quad (3)$$

Encontrando el valor máximo de D_i :

$$D_i^* = \frac{-\beta_{di} + \beta_{si} * \left[\frac{\theta_d * (\alpha_{di} - \beta_{di})}{(1 - \theta_d) * (\beta_{si} - \alpha_{si})} \right]^{\frac{1}{1 - \gamma_i}}}{(\alpha_{di} - \beta_{di}) + (\beta_{si} - \alpha_{si}) * \left[\frac{\theta_d * (\alpha_{di} - \beta_{di})}{(1 - \theta_d) * (\beta_{si} - \alpha_{si})} \right]^{\frac{1}{1 - \gamma_i}}} \quad (4)$$

Este resultado acaba siendo bastante intuitivo, siendo D_i^* afectado positivamente por un aumento en sus rendimientos en ambos tipos de evaluación (α_d y α_s), negativamente por el aumento de los rendimientos del otro método de estudio (β_d y β_s) y finalmente viéndose afectado positivamente por la probabilidad de que la prueba sea de tipo d (θ_d).

2.2. θ_d y la importancia de la probabilidad de la prueba

El modelo anterior presenta una función probabilística para especificar el tipo de prueba a la que se enfrenta el alumno, la razón detrás de esto es que en un momento dado, las escuelas donde están los estudiantes no se dedican a evaluar a los estudiantes de una sola manera. Así, θ_d nos ayuda a representar la combinación de evaluaciones a la que un estudiante se enfrenta en su colegio y al mismo tiempo deja claro la falta de certeza con respecto a la próxima evaluación generada por este.

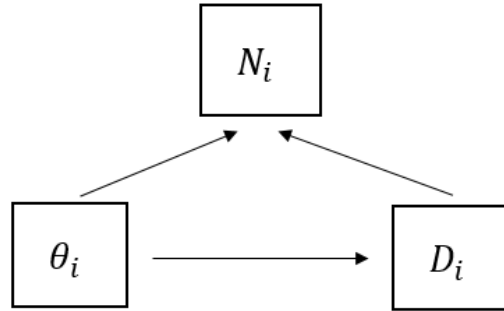
De una manera general, difícilmente en un ambiente educativo los estudiantes conocen el tipo de prueba antes de estudiar, teniendo así que predecir θ_d , siendo realmente entonces más cercano un análisis del valor esperado $E(N_i) = f(E(\theta_d))$. También es entonces fácil asumir que los estudiantes esperen que el tipo de prueba de hoy sea influenciado al menos parcialmente por el tipo de prueba que han tenido anteriormente, teniendo entonces $E(\theta_d) = g(\theta_{d,t-1})$, de esta manera ya estamos viendo que la persistencia de un método de evaluación tiene un efecto en la forma cómo los estudiantes se preparan, simplemente por el hecho de esperar que la prueba sea de esta manera.

Lo anterior entonces además deja claro que si un estudiante es evaluado de la manera por la cual está mejor preparado, le irá mejor que si esto no ocurriera.

3. Metodología

3.1. De la teoría a la estimación

Resolviendo el modelo teórico anterior notamos una relación interesante, partiendo de un análisis simple de maximización de una función $N_i = f(\alpha_{ni}, \beta_{ni}, \theta_d, D_i, \gamma_i)$ (con $n = \{d, s\}$), es decir, del análisis de la influencia del método de estudio y del método de evaluación en la nota, llegamos a una condición de optimalidad donde $D_i^* = f(\alpha_{ni}, \beta_{ni}, \theta_d, \gamma_i)$, lo que nos lleva a la siguiente relación entre nuestras variables de interés:



Donde las flechas señalan la dirección del efecto que existe entre las variables.

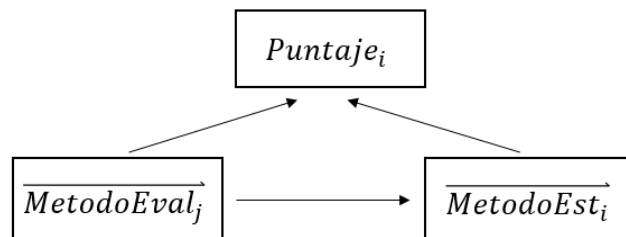
Expresemos ahora las funciones encontradas anteriormente en términos de regresiones lineales y tenemos:

$$Puntaje_i = \beta_0 + \beta_1 * \overline{MetodoEval}_j + \beta_2 * \overline{MetodoEst}_i + \varepsilon_i \quad (5)$$

$$\overline{MetodoEst}_i = \beta_0 + \beta_1 * \overline{MetodoEval}_j + \varepsilon_i \quad (6)$$

Donde $Puntaje_i$ es el puntaje que obtiene un estudiante, $\overline{MetodoEval}_j$ es el método de evaluación implementado por el colegio y su intensidad y $\overline{MetodoEst}_i$ son los métodos de estudio de los estudiantes.

Y la consecuente relación planteada anteriormente se transforma en:



Así, podemos ver claramente la relación entre ambas ecuaciones de interés, por lo cual se plantea un modelo de ecuaciones estructurales (SEM) para estimarlo. Este ejercicio se plantea adicionalmente como un modelo de ecuaciones estructurales, pues como vemos en la literatura (Tang, 1994), ha sido un método ya utilizado para el análisis del “backwash effect”.

Este modelo se planteará tanto en un nivel mundial como en uno sólo para Colombia.

4. Base de datos

Para realizar la estimación, se utiliza la base de datos PISA de 2009. Esta base de datos consta de encuestas a los colegios, a los estudiantes y la evaluación de los estudiantes de 73 países. Dado que PISA es una evaluación que se les hace a estudiantes y que además, no poseemos datos de las notas de los estudiantes en sus respectivos colegios, utilizamos como N_i que planeamos analizar, la nota obtenida en cada una de las preguntas de PISA.

Los datos que se utilizan son entonces:

Las pruebas que se realizan a los estudiantes de los colegios para determinar $Puntaje_i$.

La encuesta que se hacen a los colegios para determinar $\overline{MetodoEval}_j$.

La encuesta que se hacen a los estudiantes para determinar $\overline{MetodoEst}_l$.

Diferentes controles provenientes de ambas encuestas.

4.1. $Puntaje_i$

La prueba de los estudiantes consiste en 219 preguntas, de las cuales cada estudiante responde 60, que están divididas en 3 áreas del conocimiento: Matemáticas, Ciencia y Lectura. Estas preguntas son evaluadas con respecto a unos criterios fijos, manteniendo la objetividad y siendo evaluadas 0 (incorrecto) o 1 (correcto). Además de esto, cada pregunta diferente que se les hace a los estudiantes está agrupada en una de 6 categorías: Pregunta Abierta (Open Response), Pregunta Corta (Short Response), Pregunta abierta construida (Open Constructed Response), Pregunta cerrada construida (Closed constructed response), Opción múltiple compleja (Complex Multiple Choice) y Opción múltiple (Multiple Choice). La diferenciación sobre el tipo de pregunta que se hace en la prueba de PISA es esencial para poder evaluar el efecto de los diferentes métodos evaluativos de los colegios en los puntajes de los tipos diferentes de preguntas.

Para fines de nuestro análisis, realizaremos las estimaciones planteadas en 3 diferentes bases de datos: una con todos los tipos de preguntas, la segunda sólo con las preguntas de opción múltiple, y la última con las preguntas abiertas. Como mencionado anteriormente, son 219 preguntas de todos los tipos, siendo 113 de opción múltiple y 106 abiertas en la base de datos completa, sin embargo, cada estudiante responde aproximadamente la mitad de cada tipo en su respectiva evaluación. Así, nuestra variable de interés $Puntaje_i$ no es el total que obtiene el estudiante en su prueba PISA, sino el puntaje individual que cada estudiante obtiene en cada una

de las preguntas que responde. Esto nos permitirá analizar si hay un cambio en los efectos de los tipos de evaluación a los que los diferentes estudiantes están acostumbrados cuando evaluados contra los diferentes tipos de preguntas de la prueba PISA 2009. En general, como podemos ver en la tabla 1 de estadísticas descriptivas, las medias y desviaciones estándar de los 3 tipos de puntajes son muy similares, pero significativamente diferentes al 1% entre ellos.

4.2. Vector de métodos de evaluación: MetodoEval

La encuesta de los colegios se utiliza para obtener la frecuencia con la cual un colegio específico utiliza un método de evaluación en sus estudiantes, asumiendo que este sistema de evaluación es persistente a lo largo de la educación en un mismo colegio. Esta pregunta nos sirve para analizar no sólo el efecto de una evaluación en un momento del tiempo, sino también la persistencia de la misma a lo largo de la educación de un estudiante, pues asumimos que la intensidad de los diferentes tipos de evaluaciones para los estudiantes no cambia significativamente en el tiempo. Para esto, utilizamos la pregunta 15 de la encuesta a colegios de PISA 2009: “¿Generalmente, en su colegio, qué tan seguido son los estudiantes en el <grado de estudiantes de 15 años> evaluados usando los siguientes métodos?”. Las categorías en nuestra pregunta de interés son: Pruebas estandarizadas, Pruebas hechas por los profesores, Juicio del profesor, Portafolio estudiantil, Tareas/Proyectos de los estudiantes. En esta pregunta, cada ítem se responde de acuerdo a su intensidad de 1 a 5 siendo 1 nunca y 5 más de una vez al mes. Para nuestro análisis incluiremos todos los métodos de evaluación excepto los portafolios estudiantiles, pues no tenemos a priori un comportamiento esperado de esta variable. Adicionalmente, tomamos la libertad de recodificar los valores que toman estas variables para facilitar la interpretación de los estimadores posteriormente. Como se puede observar en el Anexo B, Q15, esta variable agrupa en números de 1 a 5 cantidades de evaluaciones anuales que no corresponden realmente a esos números, por lo tanto, recategorizamos la variable para que se comporte 0, 1.5, 4, 12, 20, y así poder tener más claro el impacto de una evaluación adicional al año. Como era de esperarse, se puede observar con estas variables, que el uso más frecuente de evaluación son los proyectos y tareas (13.91 en promedio y 17.49 para Colombia), pero es interesante observar que las pruebas estandarizadas se encuentran muy por debajo de los otros métodos de evaluación (aproximadamente 3 evaluaciones de este tipo al año para ambas muestras) (Tabla 1).

4.3. Vector de métodos de estudio: *MetodoEst_i*

La encuesta PISA a los estudiantes cuenta con muchas variables que podemos utilizar para lograr un acercamiento al método de estudio. Para nuestro estudio utilizamos las preguntas 27 “Cuando está estudiando, ¿qué tan seguido hace lo siguiente?”, 41 “¿Qué tan útil considera las siguientes estrategias para entender y recordar un texto?” y 42 “¿Qué tan útil considera las siguientes estrategias para escribir un resumen?” de la encuesta. Consideramos que estas preguntas cumplen su papel como capaces de analizar el método de estudio de un estudiante, pues se le pide al estudiante que mida la intensidad con la que implementa un método de análisis a la hora de enfrentarse a estas tareas. Estas preguntas son tan diversas que permiten ver la intensidad con la que un estudiante utiliza métodos de estudio superficiales – como “Leo el texto rápidamente dos veces” a la hora de intentar recordar un texto o “Intento copiar exactamente tantas oraciones como sea posible” a la hora de hacer un resumen – frente a los que requieren más tiempo de análisis - como “después de leer el texto, discuto su contenido con otras personas” o “Cuando estudio intento descubrir qué conceptos aún no he entendido” – junto con algunos que requieren aún más análisis que no necesariamente es incentivado por el colegio pero que sería interesante desarrollar en los estudiantes – como “Cuando estudio descubro cómo la información puede ser útil fuera del colegio” y “Cuando estudio descubro cómo la información del texto es acorde con lo que pasa en la vida real”. Estas preguntas son categorizadas con respecto al tipo de estudio en el cual encajan, siendo las 3 opciones: profundo, superficial o extra. Si bien no se ha hablado de un método de estudio “Extra” en esta categoría se incluyen aquellos métodos de estudio que reflejan un interés en aprender más allá del colegio, por lo que se considera pertinente analizar en adición a los otros dos tipos de estudio que planteamos en el modelo. Una vez categorizadas las preguntas de interés, se generan las variables Profundo, Superficial y Extra que son el promedio de las variables incluidas en ellas.

Así, el método de estudio profundo está conformado por las preguntas correspondientes a: Discutir el contenido de un texto con otras personas, resumir un texto en sus propias palabras, leer un texto tantas veces sea posible antes de escribir un resumen, revisar cuidadosamente que los hechos más importantes de un texto estén representados en un resumen, revisar si entiende bien lo que ha leído a la hora de estudiar, intentar descubrir que conceptos no ha entendido a la hora de estudiar, cuando no entiende algo busca información adicional para aclararlo.

El método de estudio superficial está conformado por las preguntas correspondientes a: concentrarse en las partes de un texto que son fáciles de entender, leer un texto rápidamente dos veces, copiar oraciones exactas al hacer un resumen, subrayar oraciones importantes y escribirlas en sus propias palabras a la hora de hacer un resumen, memorizarse todo lo que cubre un texto para estudiar, memorizar tantos detalles como sea posible, y leer un texto tantas veces que puede recitarlo.

Finalmente, el método de estudio que relaciona lo aprendido con el mundo está conformado por las preguntas correspondientes a: relacionar la información nueva con conocimiento adquirido en otras materias, descubrir cómo la información puede ser útil fuera del colegio, entender el material mejor relacionándolo con sus propias experiencias y descubrir cómo la información del texto se adapta a lo que pasa en la vida real.

Las estadísticas descriptivas de estas variables también se pueden ver en la tabla 1, estando distribuidas de 0 a 5 para los promedios de profundo y superficial, y de 0 a 3 la variable extra.

4.4 Controles

La base de datos de PISA contiene diversos controles a nivel de colegio y de estudiante. Consideramos los más importantes, y por lo tanto aquellos que se utilizaran para controlar las regresiones: género del estudiante, educación de la madre, educación del padre, si el colegio es público o privado, expectativas de los profesores, absentismo estudiantil, malas relaciones profesor-alumno, interrupción de clases por los estudiantes, profesores no satisfaciendo las necesidades individuales de los estudiantes, falta de respeto por los profesores, uso de alcohol o drogas por los estudiantes, profesores siendo muy estrictos, bullying y grado de presión de los padres por poner altos estándares académicos. Estas variables se usarán para controlar las ecuaciones estructurales pero no se considera pertinente a este trabajo el análisis específico de sus coeficientes.

5. Estimación y Resultados

5.1. Efecto del método de evaluación y estudio sobre la nota obtenida dependiendo del tipo de pregunta

Las tablas 1 y 2 nos muestran los resultados de la ecuación (5) para los tres diferentes conjuntos de preguntas de la evaluación PISA - preguntas abiertas, preguntas cerradas y ambas – para la

muestra mundial y para la muestra de Colombia. Las primeras tres columnas muestran las estimaciones con la base de datos completa, sólo con las preguntas abiertas y sólo con las preguntas de opción múltiple, respectivamente, las siguientes tres columnas repiten el ejercicio, esta vez incluyendo los controles mencionados anteriormente. Como explicado anteriormente, el modelo se estima para estos diferentes tipos de preguntas pues se quiere ver si, como encontrado en la literatura, diferentes tipos de evaluación recibidos en su colegio y su manera de estudiar predisponen al estudiante, de manera a generar aptitudes diferentes en él a la hora de enfrentarse a preguntas de diferentes tipos en la prueba PISA.

En respuesta a esto, los resultados nos dicen que no. Analizando la tabla 1, podemos ver que los métodos de estudio profundos tienen un efecto positivo significativo al 1% en el puntaje de las preguntas de PISA de 0.0871 por cada unidad de estudio adicional para la base de datos con todos los tipos de pregunta, a su vez, las preguntas abiertas de PISA también se ven afectadas de manera positiva significativa al 1%, esta vez con una magnitud de 0.0938 por cada unidad de estudio adicional, finalmente, las preguntas cerradas de PISA también se ven afectadas de manera positiva significativa al 1% con un coeficiente de 0.0809 por cada unidad de estudio adicional.

Esta persistencia de la dirección del efecto se mantiene para el resto de nuestras variables de análisis. Los métodos de estudio superficiales tienen efectos negativos significativos con coeficientes de -0.0781, -0.0814, -0.0749 para todas las preguntas, las preguntas abiertas y las preguntas cerradas, respectivamente. Analizando los métodos de estudio que buscan relacionar información con el mundo se encuentran los coeficientes -0.0158, -0.0168 y -0.0148.

Este mismo fenómeno también ocurre para los métodos de evaluación utilizados por el colegio. En general, podemos ver que el efecto que el método de estudio y de evaluación tiene en el puntaje PISA de la base de datos con todas las preguntas se mantiene en dirección cuando analizamos las bases de datos parciales. Con respecto a la magnitud, también es similar entre las muestras, siendo mayor para las preguntas abiertas, y menor para las preguntas de opción múltiple.

La tabla 2 nos muestra el mismo ejercicio pero esta vez para la base de datos de Colombia. Aquí podemos encontrar el mismo fenómeno que observamos en la tabla 1, es decir, el efecto de los métodos de evaluación y estudio sobre la nota son los mismos si analizamos todas las preguntas, sólo las preguntas abiertas, o sólo las preguntas cerradas y el efecto tiende a ser mayor en las preguntas abiertas que en las cerradas. Se encuentran algunas excepciones, las pruebas

estandarizadas tienen un efecto más negativo en la nota de las preguntas cerradas (-0.00211) que en las abiertas (-0.00175) y encontramos que las pruebas desarrolladas por los profesores no tienen un efecto significativo diferente de 0 en la nota de las preguntas abiertas pero sí lo tienen positivo y significativo al 10% en las preguntas de opción múltiple (0.00039). Lo opuesto a esto ocurre en los proyectos, los cuales tienen un efecto positivo significativo al 5% en las preguntas abiertas (0.000601), pero no tienen ningún efecto en las preguntas de opción múltiple.

De lo anterior podemos ver que diferentes métodos de estudio y evaluación no tienen diferentes efectos si analizamos los diferentes tipos de preguntas de la prueba PISA, y aún en los casos extremos encontrados, los efectos de un tipo de evaluación nunca son opuestos, por lo que podemos ver que hay una manera óptima de estudiar sin importar el tipo de pregunta, y no agrega nada a nuestro estudio seguir analizando las tres diferentes bases, por lo que seguiremos interpretando los resultados sólo con base en la muestra que posee todas las preguntas de la evaluación de PISA.

Las tablas 4 y 5 muestran las estimaciones completas del modelo estructural planteado anteriormente para la base de datos mundial y para Colombia, respectivamente. En ambas tablas, las columnas 1-4 muestran la estimación del modelo estructural, siendo 1 lo correspondiente a la ecuación (5) y 2-4 lo correspondiente a la ecuación (6), las columnas 5-8 repiten el ejercicio con los controles establecidos anteriormente.

5.2. Efecto del método de evaluación y estudio sobre la nota

Para la muestra mundial podemos ver que un aumento en el método de estudio profundo tiene un efecto positivo significativo en la nota. Específicamente, un aumento de una desviación estándar del estudio profundo tiene un efecto de 0.0706 en la nota. Para Colombia este efecto sigue siendo significativo positivo, pero cambia su magnitud a 0.0591 cuando aumenta en una desviación estándar el estudio profundo. Este es un resultado positivo pues esto incentiva los métodos de estudios profundos en los estudiantes que busquen maximizar su nota.

Por otro lado, el estudio superficial tiene un efecto negativo significativo, tanto en la muestra mundial como para Colombia. Específicamente, un aumento de una desviación estándar tiene un efecto de -0.0559 en la nota para la muestra mundial y de -0.0474 para Colombia. Este resultado

nos muestra que hay un desincentivo a estudiar por métodos superficiales si se busca maximizar la nota.

Infelizmente los resultados también nos muestran que los métodos de estudio que buscan relacionar el conocimiento con el mundo real tienen un efecto negativo significativo para ambas muestras, un aumento de una desviación estándar tiene un efecto de -0.0112 en la nota para la muestra mundial y -0.0055 para Colombia. Si bien este efecto es mucho menor que los asociados con los otros dos métodos de evaluación, sigue siendo importante que se esté desincentivando este tipo de estudio.

Con respecto a los métodos de evaluación, podemos ver que las pruebas estandarizadas tienen un efecto negativo de -0.0034 por evaluación al año que se realice en la nota en la muestra mundial, y de -0.00194 para Colombia. Con relación a un aumento en la desviación estándar, esto implica un efecto negativo de 0.0157 y 0.0083 , respectivamente.

Las pruebas desarrolladas por los profesores tienen un efecto positivo significativo en la nota, siendo este de 0.0011 para la muestra mundial y de 0.0003 para Colombia por evaluación al año. Esto implica que un aumento de una desviación estándar tiene un efecto de 0.0081 y de 0.0023 en la nota para la muestra mundial y Colombia, respectivamente.

Las evaluaciones que se realizan por medio de juicios de los profesores tienen un efecto negativo significativo en la nota de los estudiantes, para la muestra mundial este es de -0.0002 por evaluación anual (-0.0016 al aumentar una desviación estándar). Para Colombia este efecto es de -0.0006 por evaluación anual (-0.0048 al aumentar una desviación estándar).

Finalmente, los proyectos tienen el efecto positivo significativo esperado sobre la nota, teniendo un aumento de 0.0003 y 0.0005 en la nota por evaluación anual en la muestra mundial y de Colombia, respectivamente. Esto se traduce en un efecto de un aumento de una desviación estándar de 0.0023 para la muestra mundial y 0.0028 para Colombia.

5.3. El efecto del método de evaluación sobre el método de estudio

Siguiendo nuestro análisis del modelo estructural, las tablas 4 y 5 (columnas 2-4 y 6-8 como descrito anteriormente) muestran lo correspondiente a la ecuación (6) de nuestro modelo.

Las pruebas estandarizadas tienen un efecto positivo significativo en el estudio por medio de métodos profundos teniendo un aumento de 0.002 y 0.004 por evaluación anual para la muestra mundial y de Colombia, respectivamente. Un aumento de una desviación estándar tiene un efecto de 0.0092 y 0.0175 puntos de estudio para el mundo y Colombia. Así mismo, estas pruebas tienen un efecto positivo para el estudio por medio de métodos superficiales de 0.0098 para la muestra mundial y 0.008 para Colombia, esto se traduce en un cambio de 0.0453 y 0.0352 puntos de estudio cuando aumenta en una desviación estándar. Finalmente, con respecto a los estudios que relacionan la información con el mundo real, este método de evaluación tiene un efecto positivo significativo de 0.0072 para la muestra mundial y 0.0051 para Colombia. Esto quiere decir que un aumento de una desviación estándar en el método de evaluación tiene un efecto de 0.0333 y 0.0224 puntos de estudio, respectivamente.

Las pruebas desarrolladas por los profesores tienen un efecto positivo significativo en el método de estudio profundo de los estudiantes. Este efecto es de 0.0004 y 0.0014 por evaluación anual para la muestra mundial y Colombia, respectivamente, lo que se traduce en que un aumento de una desviación estándar del método de evaluación tiene un efecto de 0.0029 y 0.0109 puntos de estudio. Su efecto sobre los métodos de estudio superficiales es el opuesto, teniendo un efecto de -0.0015 para la muestra mundial y -0.0025 para Colombia, un aumento de una desviación estándar disminuye el estudio por este método en -0.0107 y 0.0193, respectivamente. Finalmente, un aumento del número de evaluaciones anuales por este método tiene un efecto de -0.0022 puntos del método de estudio que relaciona la información con el mundo real para la muestra mundial, un aumento de una desviación estándar del método de evaluación tiene un efecto de -0.0157 puntos en este método de estudio. Este método de evaluación no tiene ningún efecto en este método de estudio para Colombia.

Los juicios de los profesores tienen un efecto negativo significativo en los métodos de estudio profundos teniendo un efecto de -0.0001 y -0.0027 puntos de estudio por evaluación anual para la muestra mundial y de Colombia, respectivamente. Esto se traduce en un cambio de -0.0008 y -0.0217 puntos cuando hay un aumento de una desviación estándar. Los juicios también tienen un efecto negativo en los métodos de estudio superficiales teniendo un efecto de -0.0015 y -0.0004 para la muestra mundial y de Colombia, respectivamente. Esto quiere decir que un aumento de una desviación estándar en el método de evaluación tiene un efecto de -0.0123 y -0.0032 en los

métodos de estudio superficiales. Finalmente, este método de evaluación tiene un efecto positivo significativo sobre relacionar lo aprendido con el mundo real, con un efecto de 0.0003 y 0.0004 puntos de estudio por evaluación anual para la muestra mundial y de Colombia, respectivamente. Un aumento de una desviación estándar tiene un efecto de 0.0025 y 0.0032.

Evaluar a los estudiantes por medio de proyectos tiene un efecto negativo en los métodos de estudio profundos, teniendo un efecto de -0.001 para la muestra mundial y -0.002 para la muestra de Colombia, esto quiere decir que un aumento de una desviación estándar tiene un efecto de -0.0077 y -0.0114 puntos de estudio, respectivamente. Por otro lado, los proyectos tienen un efecto positivo significativo en los métodos de estudio superficiales tanto para la muestra mundial como de Colombia, siendo estos de 0.0007 para ambos casos. Esto resulta en un efecto de 0.0054 y 0.004 al aumentar una desviación estándar, respectivamente. Por último, este método de evaluación tiene un efecto negativo significativo en los métodos de estudio que buscan relacionar la información con el mundo, teniendo un efecto de -0.0004 puntos de estudio para la muestra mundial y -0.003 para Colombia, lo que quiere decir que un aumento de una desviación estándar tiene un efecto de -0.0031 y -0.017 en el método de estudio.

5.4. Estimaciones adicionales

Las conclusiones preliminares a las que llegamos en la sección 5.1 nos permiten adicionalmente realizar el mismo análisis de ecuaciones estructurales para la nota total que cada estudiante obtiene en su prueba. Realizamos esta estimación adicional entonces para observar si hay efectos adicionales de nuestras variables de interés que no son capturados en nuestro análisis de pregunta por pregunta y si lo son en la nota total – pues esta nota total nos permite entonces acercarnos de mejor manera a lo que sería un concepto de “aprendizaje”. Analizando nuestras variables de interés (Tabla 6), podemos ver que nuestras variables de métodos de estudio y de métodos de evaluación se comportan de manera similar en ambas muestras tanto para Colombia como para la muestra mundial. Igualmente, se puede observar que la variable de puntaje es consistente con la variable de puntaje total, siendo ambas menores en la muestra de Colombia que en la mundial.

Las tablas 7 y 8 muestran los resultados de la estimación teniendo como variable dependiente el puntaje total. Con respecto a la ecuación (5), esta nueva estimación no muestra diferencias de signo o significancia de las variables independientes con respecto al puntaje total para ninguna de

las muestras. Es claro que el cambio de la magnitud de los estimadores cambiará dado que el comportamiento de la variable dependiente es diferente, sin embargo, para hacer estos resultados comparables debemos analizar su tamaño relativo a sus respectivas desviaciones estándar.

Un aumento de una desviación estándar del método de estudio profundo tiene un efecto de 4.05 y 3.32 en el puntaje total de la muestra mundial y de Colombia, lo que equivale a 0.31 y 0.32 desviaciones estándar del puntaje total en sus respectivas muestras, lo que presenta un gran aumento del efecto que se encontraba en nuestra estimación anterior, donde se encontraba un efecto de 0.14 y 0.12 desviaciones estándar del puntaje.

Un aumento de una desviación estándar del método de estudio superficial tiene un efecto de -3.25 y -2.65 en el puntaje mundial de nuestras muestras, lo que equivale a 0.26 y 0.25 desviaciones estándar para la muestra mundial y Colombia, respectivamente. Esto también presenta un efecto de la magnitud que vimos en nuestra estimación de las secciones anteriores, donde el efecto era de 0.11 y 0.1 desviaciones estándar del puntaje.

Un aumento de una desviación estándar del método de estudio que busca relacionar la información con el mundo real tiene un efecto de -0.71 y -0.42 para nuestras muestras, lo que es equivalente a 0.6 y 0.4 desviaciones estándar del puntaje total en las muestras mundial y de Colombia, respectivamente. Esto cambio muestra un aumento en la magnitud con respecto a nuestra estimación anterior, donde se veía un efecto de 0.2 y 0.1 desviaciones estándar del puntaje.

Un aumento de una desviación estándar en el uso de pruebas estandarizadas tiene un efecto de -1.03 y -0.6 para nuestras muestras, lo que equivale a 0.08 y 0.05 desviaciones estándar del puntaje total. Esto es un aumento con respecto a nuestra estimación anterior, donde encontrábamos un efecto de 0.03 y 0.02 para el mundo y Colombia, respectivamente.

Con respecto a las pruebas desarrolladas por los profesores, un aumento de una desviación estándar en estas tiene un efecto de 0.47 y 0.12, lo que equivale a 0.04 y 0.1 desviaciones estándar de nuestra variable de interés, para la muestra mundial y Colombia. Esto es un aumento con respecto a nuestra estimación anterior, donde el efecto encontrado era de 0.02 y 0.005 desviaciones estándar.

Para los juicios de los profesores, un aumento de una desviación estándar de estos tiene un efecto de -0.04 y -0.32 sobre el puntaje total para las muestras mundial y de Colombia,

respectivamente, lo que equivale a 0.003 y 0.03 desviaciones estándar de nuestra variable de interés. Esto es un aumento con respecto a nuestra estimación anterior, donde estas variables tenían un efecto de 0.0027 y 0.01 desviaciones estándar.

Finalmente, una desviación estándar de las pruebas realizadas por medio de tareas y proyectos tiene un efecto de 0.1 y 0.28 sobre el puntaje total, lo que corresponde a 0.008 y 0.27 desviaciones estándar para la muestra mundial y de Colombia, respectivamente. Esto también presenta un aumento con respecto a nuestra estimación anterior, donde el efecto encontrado era de 0.004 y 0.005 desviaciones estándar.

Analizando lo correspondiente a la ecuación (6), dado que nuestra estimación no cambia las variables de interés para este caso, no deberíamos encontrar cambios en esta estimación para nuestra base de datos, y para el caso de la muestra mundial este es el caso, teniendo la misma significancia y magnitud en todos los casos. Sin embargo, al analizar la muestra de Colombia, podemos ver que pocos casos se mantienen significativamente diferentes de 0, siendo estos los juicios de los profesores para el estudio profundo, las pruebas estandarizadas para el estudio superficial, y las pruebas estandarizadas y los proyectos para los estudios que buscan relacionar la información con el mundo real. El tamaño o la dirección de estos efectos tampoco cambian.

6. Discusión

La literatura económica no ha estudiado de manera importante el “backwash effect”, sin embargo, este trabajo muestra evidencia de su existencia. Si analizamos los resultados obtenidos podemos ver que los métodos de evaluación tienen un efecto importante sobre el puntaje obtenido en las preguntas de la evaluación PISA, y este efecto solo aumenta cuando pasamos a considerar como variable de interés, el puntaje total que un estudiante obtiene en la prueba PISA, frente a los resultados de las preguntas individuales. Si queremos ir un poco más allá, podemos ver que en el ejercicio realizado con la prueba PISA no se encuentra el efecto esperado bajo el cual estudiantes más acostumbrados a un tipo de evaluación les iría mejor cuando encontrados con una pregunta del mismo tipo, lo que muestra que hay una manera óptima de estudiar y se podría hablar del puntaje obtenido en PISA como una buena medida de aprendizaje. Si consideramos esto cierto, entonces es claro que hay métodos de estudio que hacen que el estudiante aprenda de mejor manera con respecto a unos estándares. Específicamente, aquí se ve que un método de estudio profundo es la mejor opción de estudio para una persona en este caso. Sin embargo, también es importante

ver que a la luz de estos resultados, estudiar de una manera a relacionar lo aprendido con el mundo real es malo, si bien el efecto de este estudio en la nota es menor que los otros dos métodos, de manera a ser en su punto más alto 1% de la variable de interés, siento que más se debería hacer en este área para que su efecto fuera positivo, pasando de tener un efecto despreciable negativo a realmente incentivar a los estudiantes a comportarse de esa manera.

Adicionalmente podemos ver que, aunque la mayoría de los efectos sean significativos, al analizar las magnitudes se puede observar que estas no alteran de manera importante las variables de interés, entonces, aunque queda clara la existencia del “backwash effect”, no es claro si su efecto es suficientemente positivo o negativo para incentivar un cambio en la estructura de evaluación. Sin embargo, algo que queda claro por la comparación de la muestra mundial con Colombia es que los efectos de las evaluaciones pueden cambiar en magnitud, lo que abre las puertas a la investigación de, no tanto el efecto de una mayor intensidad de una evaluación, sino la manera de hacer que un tipo de evaluación, cualquiera que este sea, traiga con él los incentivos de los métodos de estudio que queremos incentivar, pues podemos ver que la evaluación de PISA mostraba sólo un comportamiento optimo desde un punto de vista maximizador racional, teniendo en sí, diferentes tipos de evaluación.

7. Conclusión

Este trabajo refuerza la literatura por detrás de la existencia del “backwash effect”, sin embargo, muestra problemas en el tamaño de sus efectos en nuestras variables de interés, las notas de los estudiantes y sus métodos de estudio, siendo en su mayor parte muy pequeños para tener un efecto importante, además de encontrar resultados contra intuitivos y no deseados. La evaluación de PISA nos da una guía inicial del trabajo a seguir, pues muestra que diferentes tipos de preguntas pueden igualmente incentivar un mismo tipo de estudio, por lo tanto señala que el camino a seguir puede no ser intentar poner un método de evaluación sobre el otro, sino encontrar una manera de que todos los diferentes tipos de evaluación lleven al aprendizaje deseado en los estudiantes.

Bibliografia

Alderson C., & Wall, D. (1992). Does Washback exist?. En Oxford Journals. Volume 14, No. 2. Pp. 115-129.

Akerhielm, K. (1995). Does Class size matter?. En Economics of Education Review. Volume 14, No. 3. Pp. 229-241.

Akerlof, G. & Kranton, R. (2002). Identity and Schooling: Some lessons for the Economics of Education. En Journal of Economic Literature. Volume XL. Pp. 1167-1201.

Beleche, T., Fairris, D., Marks, M. (2010). Do Course Evaluations Truly Reflect Student Learning?: Evidence from an Objectively Graded Post-test.

Biggs, J. (1987). Student approaches to learning and studying.

Biggs, J. (1999). Teaching for Quality Learning at University.

Biggs, J. (1996). Assessing learning quality: reconciling institutional, staff and educational demands. En Assessment & Evaluation in higher Education. Volume 21, Issue 1.

Biggs, J. (1979). Individual differences in study processes and the quality of learning outcomes. En Higher Education. Volume 8. Pp. 381-394.

Boud, D., Falchikov, N. (2006). Aligning assessment with long-term learning. En Assessment & Evaluation in Higher Education. Volume 31, Issue 4.

Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. En Language and Education, Volume 37, No. 17.

Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. *Washback in language testing: Research contexts and methods*, 3-17.

Coleman, J. & Lister, R. (2004). The backwash effect on SQL skills grading. En ITiCSE '04 Proceedings of the 9th annual SIGCSE conference on Innovation and technology in computer science education. Pp. 32-36.

Dee, T. (2004). Teachers, race and student achievement in a randomized experiment. En Review of Economics and Statistics. Volume 86, No. 1. Pp. 195-210.

- Dee, T. (1998). Competition and the Quality of Public Schools. En *Economics of Education Review*, Volume 17, No. 4, pp. 419-427.
- Entwistle, N. & Tait, H. (1990). Approaches to learning, evaluations of teaching, and preferences for contrasting academic environments. En *Higher Education*. Volume 19. Pp. 169-194.
- Feldman, K., Paulsen, M. (1998). Teaching and learning in the college classroom.
- Fujimoto, H. (1999). The Examination Backwash effect on English Language Education in Japan. En *The Japanese Learner*.
- Fry, H., Ketteridge, S., Marshall, S. (2003). A handbook for teaching & learning in higher education: enhancing academic practice.
- Hanushek, E. (2007). Education Production Functions. En *Palgrave Encyclopedia*.
- Havnes, A. (2004). Examination and learning: an activity-theoretical analysis of the relationship between assessment and educational practice. En *Assessment & Evaluation in Higher Education*. Volume 29, Issue 4.
- Hill, H., Rowan, B. & Loewenberg, D. (2005). Effects of teachers' Mathematical knowledge for teaching on student achievement. En *American Educational Research Journal*. Volume 42, No. 2. Pp. 371-406.
- Kember, D. et al. (2000). Development of a questionnaire to measure the level of reflective thinking. En *Assessment & Evaluation in Higher Education*. Volume 24, Issue 4.
- Klapp Lekholm, A. (2008). Grades and grade assignment: Effects of student and school characteristics.
- Li, M., Shavelson, R.J. (2003). Validating the links between knowledge and test items from a protocol analysis.
- McMillan, J. (2001). Classroom assessment: Principles and practices for effective instruction.
- Monk, D. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. En *Economics of Education Review*. Volume 13, No. 2. Pp. 125-145.

- Nye, B., Hedges, L.V., Konstantopoulos, S. (2000). The effect of small classes on academic achievement: The Results of the Tennessee Class Size Experiment. En *American Educational Research Journal*, 37: 123-151
- Price, M. et al. (2011). If I was going there I wouldn't start from here: a critical commentary on current assessment practice. En *Assessment & Evaluation in Higher Education*. Volume 36, Issue 4.
- Prior, J. C., & Lister, R. (2004, June). The backwash effect on SQL skills grading. In *ACM SIGCSE Bulletin* (Vol. 36, No. 3, pp. 32-36). ACM.
- Reimann, N. (2011). To risk it or not to risk it: student (non-)engagement with seen examination questions. En *Assessment & Evaluation in Higher Education*. Volume 36, Issue 3.
- Reynolds, C., Livingston, R. Willson, V. (2006). *Measurement and assessment in education*.
- Savalei, V., & Bentler, P. M. (2010). Structural equation modeling. *Corsini Encyclopedia of Psychology*.
- Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. En *Higher Education*. Volume 35. Pp. 453-472.
- Smith, S. & Miller, R. (2005). Learning approaches: Examination type, discipline of study, and gender. En *Educational Psychology*. Vol. 25, No. 1. Pp. 43-53.
- Sternberg, R. J. (2000). *Handbook of Intelligence*.
- Sternberg, R.J. (2007). *Teaching for successful intelligence: to increase student learning and achievement*.
- Tang, C. (1994). Assessment and student learning: Effects of modes of assessment on students' preparation strategies. En *Improving Student Learning: Theory and Practice*.
- Thomas, P. & Bain, J. (1982). Consistency in learning strategies. En *Higher Education*. Volume 11. Pp. 249-259.
- Walstad, W. B., & Becker, W. E. (1994). Achievement differences on multiple-choice and essay tests in economics. *The American Economic Review*, 84(2), 193-196.

Watkins, D. (1983). Depth of processing and the quality of learning outcomes. En *Instructional Science*. Volume 12. Pp. 49-58.

Watkins, D., Dahlin, B., & Ekholm, M. (2005). Awareness of the backwash effect of assessment: A phenomenographic study of the views of Hong Kong and Swedish lecturers. *Instructional Science*, 33(4), 283-309.

Wesdorp, H. (1981). Backwash Effects of Language-Testing in Primary and Secondary Education.

Wiggins, G. (1999). Assessing Student performance: Exploring the purpose and limits of testing.

Willingham, W. W., Pollack, J. M. and Lewis, C. (2002), Grades and Test Scores: Accounting for Observed Differences. *Journal of Educational Measurement*, 39: 1–37.

Zimmerman, D.J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. En *The Review of Economics and Statistics*, 85: 9-23

Anexo A

Tabla 1 - Estadísticas descriptivas

	Mundial		Colombia		Min.	Max.
	Media	Desv. Estandar	Media	Desv. Estandar		
Puntaje	0.5082	0.4999	0.4472	0.4972	0	1
Puntaje (A)	0.4758	0.4994	0.4142	0.4926	0	1
Puntaje (OM)	0.5386	0.4985	0.4802	0.4996	0	1
Estudio Profundo	2.5951	0.8115	2.6092	0.8419	0	5
Estudio Superficial	2.1406	0.7154	2.1854	0.7355	0	5
Estudio Extracurricular	1.3881	0.7090	1.6450	0.6978	0	3
Pruebas Estandarizadas	3.0399	4.6210	3.1516	4.3949	0	20
Pruebas desarrolladas por profesores	11.3492	7.1402	13.5752	7.7511	0	20
Juicio de Profesores	10.3516	8.2075	11.0565	8.0363	0	20
Proyectos	13.9122	7.7164	17.4920	5.6753	0	20

Tabla 2 - Estimación SEM - Ecuacion (5) sobre diferentes tipos de pregunta - Mundial

	(1) Todas	(2) Opcion Multiple	(3) Abierta	(4) Todas	(5) Opcion Multiple	(6) Abierta
Profundo	0.111*** (0.000139)	0.121*** (0.000199)	0.100*** (0.000193)	0.0871*** (0.000162)	0.0938*** (0.000232)	0.0809*** (0.000225)
Superficial	-0.0923*** (0.000153)	-0.0966*** (0.000219)	-0.0881*** (0.000213)	-0.0781*** (0.000170)	-0.0814*** (0.000243)	-0.0749*** (0.000236)
Extra	-0.0264*** (0.000145)	-0.0288*** (0.000208)	-0.0241*** (0.000202)	-0.0158*** (0.000161)	-0.0168*** (0.000230)	-0.0148*** (0.000223)
Pruebas Estandarizadas	-0.00392*** (2.06e-05)	-0.00426*** (2.94e-05)	-0.00358*** (2.86e-05)	-0.00344*** (2.26e-05)	-0.00378*** (3.24e-05)	-0.00309*** (3.15e-05)
Pruebas desarrolladas por profesores	0.00195*** (1.46e-05)	0.00223*** (2.10e-05)	0.00168*** (2.03e-05)	0.00114*** (1.60e-05)	0.00135*** (2.30e-05)	0.000931*** (2.22e-05)
Juicios de los profesores	-0.000109*** (1.22e-05)	-0.000248*** (1.75e-05)	1.03e-05 (1.70e-05)	-0.000165*** (1.34e-05)	-0.000293*** (1.92e-05)	-5.50e-05*** (1.86e-05)
Proyectos	0.000551*** (1.31e-05)	0.000733*** (1.88e-05)	0.000385*** (1.83e-05)	0.000290*** (1.45e-05)	0.000438*** (2.08e-05)	0.000156*** (2.02e-05)
Controles	No	No	No	Si	Si	Si
Observaciones	28,310,910	13,690,263	14,620,647	23,423,008	11,324,075	12,098,933

Errores estandar en parentesis *** p<0.01, ** p<0.05, * p<0.1

Tabla 3 - Estimación SEM - Ecuacion (5) sobre diferentes tipos de pregunta - Colombia

	(1)	(2)	(3)	(4)	(5)	(6)
	Todas	Opcion Multiple	Abierta	Todas	Opcion Multiple	Abierta
Profundo	0.0916*** (0.00113)	0.101*** (0.00158)	0.0823*** (0.00160)	0.0702*** (0.00134)	0.0768*** (0.00187)	0.0639*** (0.00190)
Superficial	-0.0762*** (0.00124)	-0.0812*** (0.00173)	-0.0716*** (0.00176)	-0.0645*** (0.00139)	-0.0682*** (0.00195)	-0.0610*** (0.00199)
Extra	-0.0123*** (0.00122)	-0.0135*** (0.00171)	-0.0114*** (0.00174)	-0.00794*** (0.00138)	-0.00943*** (0.00193)	-0.00662*** (0.00196)
Pruebas Estandarizadas	0.000212 (0.000178)	0.000607** (0.000249)	-0.000170 (0.000253)	-0.00194*** (0.000208)	-0.00175*** (0.000292)	-0.00211*** (0.000296)
Pruebas desarrolladas por profesores	5.59e-06 (0.000112)	-0.000101 (0.000157)	0.000108 (0.000159)	0.000290** (0.000132)	0.000240 (0.000184)	0.000339* (0.000188)
Juicios de los profesores	-0.000525*** (0.000101)	-0.000523*** (0.000141)	-0.000532*** (0.000143)	-0.000608*** (0.000118)	-0.000728*** (0.000165)	-0.000505*** (0.000167)
Proyectos	0.00105*** (0.000151)	0.00117*** (0.000211)	0.000923*** (0.000215)	0.000462** (0.000181)	0.000601** (0.000253)	0.000318 (0.000258)
Controles	No	No	No	Si	Si	Si
Observaciones	407,238	203,332	203,906	327,634	163,551	164,083

Errores estandar en parentesis *** p<0.01, ** p<0.05, * p<0.1

Tabla 4 - Estimación SEM - Ecuaciones (5) y (6) - Mundial

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Puntaje	Profundo	Superficial	Extra	Puntaje	Profundo	Superficial	Extra
Profundo	0.111*** (0.000139)				0.0871*** (0.000162)			
Superficial	-0.0923*** (0.000153)				-0.0781*** (0.000170)			
Extra	-0.0264*** (0.000145)				-0.0158*** (0.000161)			
Pruebas Estandarizadas	-0.00392*** (2.06e-05)	0.00350*** (3.35e-05)	0.0114*** (2.94e-05)	0.00896*** (2.94e-05)	-0.00344*** (2.26e-05)	0.00204*** (3.39e-05)	0.00975*** (3.11e-05)	0.00719*** (3.16e-05)
Pruebas desarrolladas por profesores	0.00195*** (1.46e-05)	0.000839*** (2.39e-05)	-0.00185*** (2.10e-05)	-0.00277*** (2.10e-05)	0.00114*** (1.60e-05)	0.000393*** (2.42e-05)	-0.00147*** (2.21e-05)	-0.00221*** (2.25e-05)
Juicios de los profesores	-0.000109*** (1.22e-05)	-0.000920*** (2.00e-05)	-0.00171*** (1.76e-05)	3.04e-05* (1.76e-05)	-0.000165*** (1.34e-05)	-0.000144*** (2.01e-05)	-0.00151*** (1.84e-05)	0.000270*** (1.87e-05)
Proyectos	0.000551*** (1.31e-05)	-0.000153*** (2.15e-05)	0.00125*** (1.89e-05)	0.000225*** (1.89e-05)	0.000290*** (1.45e-05)	-0.00101*** (2.19e-05)	0.000746*** (2.00e-05)	-0.000397*** (2.04e-05)
Controles	No	No	No	No	Si	Si	Si	Si
Observaciones	28,310,910	28,310,910	28,310,910	28,310,910	23,423,008	23,423,008	23,423,008	23,423,008

Errores estandar en parentesis

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

Tabla 5 - Estimación SEM - Ecuaciones (5) y (6) - Colombia

	(1) Puntaje	(2) Profundo	(3) Superficial	(4) Extra	(5) Puntaje	(6) Profundo	(7) Superficial	(8) Extra
Profundo	0.0916*** (0.00113)				0.0702*** (0.00134)			
Superficial	-0.0762*** (0.00124)				-0.0645*** (0.00139)			
Extra	-0.0123*** (0.00122)				-0.00794*** (0.00138)			
Pruebas Estandarizadas	0.000212 (0.000178)	0.00744*** (0.000302)	0.00358*** (0.000263)	0.00404*** (0.000252)	-0.00194*** (0.000208)	0.00401*** (0.000321)	0.00795*** (0.000292)	0.00511*** (0.000280)
Pruebas desarrolladas por profesores	5.59e-06 (0.000112)	0.000725*** (0.000190)	-0.00207*** (0.000165)	-0.00109*** (0.000158)	0.000290** (0.000132)	0.00142*** (0.000206)	-0.00247*** (0.000188)	8.77e-05 (0.000179)
Juicios de los profesores	-0.000525*** (0.000101)	-0.00388*** (0.000170)	-0.000132 (0.000148)	0.000481*** (0.000142)	-0.000608*** (0.000118)	-0.00274*** (0.000181)	-0.000412** (0.000165)	0.000356** (0.000158)
Proyectos	0.00105*** (0.000151)	-1.46e-05 (0.000256)	0.000295 (0.000223)	-0.00162*** (0.000214)	0.000462** (0.000181)	-0.00200*** (0.000281)	0.000658*** (0.000256)	-0.00304*** (0.000245)
Controles	No	No	No	No	Si	Si	Si	Si
Observaciones	407,238	407,238	407,238	407,238	327,634	327,634	327,634	327,634

Errores estandar en parentesis *** p<0.01, ** p<0.05, * p<0.1

Tabla 6 - Estadísticas descriptivas - Puntaje Total

	Mundial		Colombia		Min.	Max.
	Media	Desv. Estandar	Media	Desv. Estandar		
Puntaje Total	28.3770	12.6880	23.3425	10.7298	0	60
Estudio Profundo	2.5947	0.8148	2.6012	0.8489	0	5
Estudio Superficial	2.1436	0.7193	2.1875	0.7403	0	5
Estudio Extracurricular	1.3914	0.7095	1.6488	0.6999	0	3
Pruebas Estandarizadas	3.0647	4.6488	3.1704	4.4337	0	20
Pruebas desarrolladas por profesores	11.3354	7.1401	13.5966	7.7612	0	20
Juicio de Profesores	10.3339	8.2066	11.0487	8.0427	0	20
Proyectos	13.9087	7.7161	17.4816	5.6825	0	20

Tabla 7 - Estimación SEM - Ecuaciones (5) y (6) con puntaje total - Mundial

	(1) Puntaje Total	(2) Profundo	(3) Superficial	(4) Extra	(5) Puntaje	(6) Profundo	(7) Superficial	(8) Extra
Profundo	6.329*** (0.0245)				4.969*** (0.0265)			
Superficial	-5.350*** (0.0269)				-4.521*** (0.0277)			
Extra	-1.657*** (0.0256)				-1.000*** (0.0263)			
Pruebas Estandarizadas	-0.251*** (0.00361)	0.00356*** (0.000250)	0.0115*** (0.000220)	0.00898*** (0.000219)	-0.222*** (0.00369)	0.00204*** (0.000253)	0.00987*** (0.000232)	0.00718*** (0.000236)
Pruebas desarrolladas por profesores	0.116*** (0.00258)	0.000932*** (0.000179)	-0.00176*** (0.000158)	-0.00270*** (0.000157)	0.0656*** (0.00262)	0.000469*** (0.000182)	-0.00139*** (0.000167)	-0.00212*** (0.000169)
Juicios de los profesores	0.00165 (0.00216)	-0.000909*** (0.000150)	-0.00173*** (0.000132)	-5.16e-06 (0.000131)	-0.00524** (0.00219)	-0.000162 (0.000151)	-0.00152*** (0.000139)	0.000239* (0.000141)
Proyectos	0.0280*** (0.00232)	-0.000160 (0.000161)	0.00120*** (0.000142)	0.000244* (0.000141)	0.0135*** (0.00238)	-0.000975*** (0.000165)	0.000708*** (0.000151)	-0.000384** (0.000153)
Controles	No	No	No	No	Si	Si	Si	Si
Observaciones	505,900	505,900	505,900	505,900	417,487	417,487	417,487	417,487
Errores estandar en parentesis	*** p<0.01, ** p<0.05, * p<0.1							

Tabla 8 - Estimación SEM - Ecuaciones (5) y (6) con puntaje total - Colombia

	(1) Puntaje Total	(2) Profundo	(3) Superficial	(4) Extra	(5) Puntaje	(6) Profundo	(7) Superficial	(8) Extra
Profundo	5.278*** (0.164)				3.906*** (0.179)			
Superficial	-4.328*** (0.181)				-3.583*** (0.187)			
Extra	-0.940*** (0.178)				-0.599*** (0.185)			
Pruebas Estandarizadas	-0.00577 (0.0258)	0.00682*** (0.00218)	0.00402** (0.00189)	0.00438** (0.00181)	-0.131*** (0.0277)	0.00348 (0.00232)	0.00830*** (0.00211)	0.00543*** (0.00201)
Pruebas desarrolladas por profesores	-0.00528 (0.0163)	0.000392 (0.00138)	-0.00167 (0.00120)	-0.000996 (0.00115)	0.0156 (0.0177)	0.00132 (0.00150)	-0.00198 (0.00136)	0.000341 (0.00130)
Juicios de los profesores	-0.0211 (0.0147)	-0.00354*** (0.00124)	-0.000218 (0.00108)	0.000274 (0.00103)	-0.0397** (0.0158)	-0.00242* (0.00132)	-0.000592 (0.00120)	9.26e-05 (0.00115)
Proyectos	0.0640*** (0.0221)	-0.000190 (0.00186)	0.000149 (0.00162)	-0.00167 (0.00155)	0.0492** (0.0243)	-0.00240 (0.00204)	0.000293 (0.00186)	-0.00337* (0.00177)
Controles	No	No	No	No	Si	Si	Si	Si
Observaciones	7,789	7,789	7,789	7,789	6,240	6,240	6,240	6,240
Errores estandar en parentesis	*** p<0.01, ** p<0.05, * p<0.1							

Anexo B

Q15 ¿Generalmente, en su colegio, que tan seguido son los estudiantes en el <grado de estudiantes de 15 años> evaluados usando los siguientes métodos?

	<i>Nunca</i>	<i>1-2 veces al año</i>	<i>3-5 veces al año</i>	<i>Mensual</i>	<i>Más de una vez al mes</i>
a) Prueba estandarizada	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
b) Prueba desarrollada por los profesores	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
c) Juicio de los profesores	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
d) Portafolio estudiantil	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅
e) Proyectos de los estudiantes	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅

Q27 ¿Cuándo está estudiando, que tan frecuentemente hace lo siguiente?

	<i>Casi nunca</i>		<i>Casi siempre</i>	
a) Cuando estudio, Intento memorizar todo lo que cubre el texto	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
b) Cuando estudio comienzo por descubrir exactamente lo que tengo que aprender	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
c) Cuando estudio, intento memorizar tantos detalles como sea posible	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
d) Cuando estudio intento relacionar información nueva con conocimiento adquirido anteriormente en otras materias	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
e) Cuando estudio, leo el texto tantas veces que puedo recitarlo	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
f) Cuando estudio, reviso si entiendo lo que he leído	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
g) Cuando estudio leo el texto una y otra vez	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
h) Cuando estudio descubro como la información puede ser útil fuera del colegio	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
i) Cuando estudio, intento descubrir que conceptos aún no he entendido	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
j) Cuando estudio intento entender el material mejor relacionándolo con mis propias experiencias	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
k) Cuando estudio, me aseguro de entender los puntos más importantes del texto	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
l) Cuando estudio descubro como la información del texto se adapta a lo que pasa en la vida real	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
m) Cuando estudio y no entiendo algo, busco información adicional para aclararlo	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄

Q41 Debes entender y recordar la información en un texto

¿Que tan útil considera las siguientes estrategias para entender y memorizar el texto?

Posible estrategia	Para nada útil					Muy útil
	(1)	(2)	(3)	(4)	(5)	
a) Me concentro en las partes del texto que son fáciles de entender	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	<input type="checkbox"/> ₆
b) Leo rápidamente el texto dos veces	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	<input type="checkbox"/> ₆
c) Después de leer el texto, discuto su contenido con otras personas	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	<input type="checkbox"/> ₆
d) Subrayo los hechos importantes del texto	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	<input type="checkbox"/> ₆
e) Resumo el texto en mis propias palabras	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	<input type="checkbox"/> ₆
f) Leo el texto en voz alta a otra persona	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	<input type="checkbox"/> ₆

Q42 Acabas de leer un largo y difícil text de dos páginas. Debes escribir un resumen

¿Que tan util considera las siguientes estrategias para escribir un resumen de este texto de dos páginas?

Posible estrategia	<i>Para nada útil</i>					<i>Muy útil</i>
	<i>(1)</i>	<i>(2)</i>	<i>(3)</i>	<i>(4)</i>	<i>(5)</i>	<i>(6)</i>
a) Escribo un resumen, despues reviso que cada párrafo esté incluido pues el contenido de cada parrafo deberia estar incluido	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	<input type="checkbox"/> ₆
b) Intento copiar exactamente tantas oraciones como sea posible	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	<input type="checkbox"/> ₆
c) Antes de escribir el resumen leo el texto tantas veces como posible	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	<input type="checkbox"/> ₆
d) Reviso cuidadosamente si los hechos mas importantes del texto están representados en el resumen	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	<input type="checkbox"/> ₆
e) Leo el texto subrayando las oraciones mas importantes. Despues las escribo en mis propias palabras como resumen	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅	<input type="checkbox"/> ₆