



Memorias
III Seminario del ICFES Sobre Investigación en Calidad de Educación
Bogotá, D.C., Noviembre 1 y 2 de 2012



Contenido

1. El Seminario Anual del ICFES	3
2. El Seminario anual del ICFES edición 2012.....	4
3. Comparación de contenidos de los seminarios 2010 a 2012.....	8
4. Las presentaciones académicas	10
5. Conferencias alternas	28
6. Estudios presentados en formato de poster	35
7. Los talleres de capacitación sobre metodologías de investigación	37
8. Los participantes y su evaluación sobre el seminario	42
9. Conclusiones	44

Tabla de ilustraciones

Ilustración 1 – Conferencias Magistrales de los Seminarios ICFES 2010 a 2012, por línea de investigación y disciplina	9
Ilustración 2 – Muestreo para construir una prueba de logro académico (Koretz)	12
Ilustración 3 – Impacto de la publicación de resultados en colegios públicos y privados	15
Ilustración 4 – Predicciones teóricas del efecto Pez-grande en Laguna Pequeña	17
Ilustración 5 –Intuición Geométrica Básica sobre el Valor Agregado.....	19
Ilustración 7 – Brechas de ingreso/logro en Estados Unidos por cohorte de nacimiento de los estudiantes. Calculadas con 13 estudios de logro.....	26
Ilustración 8 – Visiones Cambiantes sobre la Crianza 1900-1985 (Wrigley, 1989).....	27
Ilustración 9 –Me aportó nuevos elementos útiles para mi desempeño profesional	42
Ilustración 10 – Me gustaría participar en siguientes seminarios.....	43
Ilustración 11 –Me motiva a investigar sobre calidad de educación.....	43
Ilustración 12 –Las temáticas son pertinentes sobre calidad de educación.....	43
Ilustración 13 – Percepción sobre los posters	44

Memorias

III Seminario del ICFES Sobre Investigación en Calidad de Educación Bogotá, D.C. Noviembre 1 y 2 de 2012

Este informe contiene las memorias del *III Seminario Internacional de Investigación sobre Calidad de la Educación*, organizado por el Instituto Colombiano para la Evaluación de la Calidad (ICFES) y realizado en la ciudad de Bogotá, los días 1 y 2 de Noviembre de 2012.

El primer capítulo del documento contiene los antecedentes del Seminario. A continuación se encuentra un resumen ejecutivo de las principales ideas desarrolladas en cada conferencia, tanto en las sesiones principales como alternas. Sus detalles pueden encontrarse en diversos estudios, cuyas referencias se indican en el documento. Varios de ellos se han publicado en la página web www.icfes.gov.co/investigacion.

Luego se describen las estadísticas sobre el perfil de los participantes (incluyendo una comparación con los Seminarios anteriores), y los resultados de la evaluación realizada por los asistentes al Seminario. Finalmente se presentan algunas conclusiones.

1. El Seminario Anual del ICFES

La misión del Instituto Colombiano para la Evaluación de la Educación, ICFES, tiene dos frentes: por una parte, ofrecer servicios de evaluación de la educación en todos sus niveles y, por otra, fomentar la investigación sobre calidad educativa para contribuir a su mejoramiento, a partir de los datos de las pruebas aplicadas.

Para promover la investigación de la calidad educativa, el ICFES formuló un *Programa de Investigación sobre Calidad de la Educación*. Éste se ejecuta a través de dos convocatorias a investigadores, adelantadas mediante un proceso competitivo, que financian estudios sobre la materia que empleen los resultados de las pruebas que aplica el instituto.

El proceso de reforma del Instituto ha estado acompañado de una política de acceso a las bases de datos para el desarrollo de investigaciones, manteniendo protegida la identidad de las personas que toman las pruebas y de las preguntas que se encuentran en éstas¹.

En particular, el *Seminario de investigación sobre la calidad de la educación* fue concebido como un espacio académico anual para la presentación de estudios innovadores adelantados por investigadores nacionales e internacionales, además de los

¹En su Artículo 4, la Ley 1324 de 2009 establece que "... los resultados individuales podrán comunicarse a terceros que los requieran con el fin exclusivo de adelantar investigaciones sobre educación, si garantizan que el dato individual no será divulgado sin consentimiento previo de la persona evaluada. Sin perjuicio de la comunicación de datos agregados, o para investigaciones, los datos relativos a cada persona pertenecerán a aquella y no podrán ser divulgados sino con su autorización...".

trabajos más destacados financiados por el *Programa de Investigación*. El seminario es un espacio académico anual para presentación de estudios sobre calidad educativa que busca ser un referente académico por presentar trabajos rigurosos, recientes e independientes; un espacio de diálogo de investigadores de varias disciplinas, educadores de todos los niveles de la educación, y diseñadores de política, interesados en el mejoramiento de la calidad educativa. El seminario tiene gradualmente auditorios en línea en distintas regiones del país, con participación doble vía, y acceso de sus conferencias offline, en cualquier momento del año, en la página web del Instituto, para los interesados.

2. El Seminario anual del ICFES edición 2012

A. Objetivos

El Seminario Anual de Investigación en Educación del ICFES se llevó a cabo los días 1 y 2 de Noviembre de 2012 en el Hotel Marriot de Bogotá, con participantes mediante videoconferencia, en auditorios de la Universidad de Cundinamarca (Fusagasugá), Universidad Simón Bolívar (Barranquilla) y Universidad Industrial de Santander (Bucaramanga). La política del Instituto es el aprovechamiento de las TIC para que en el seminario participen los interesados de distintas regiones de Colombia.

El propósito del evento fue dar a conocer ante la comunidad académica algunas de las investigaciones más recientes que emplean datos de pruebas estandarizadas sobre logro académico. Los trabajos fueron seleccionados por su excelencia académica, y procurando heterogeneidad en preguntas de investigación, disciplinas y métodos empíricos. En particular, para este año el Instituto enfatizó la presentación de estudios que contribuyeran a que los investigadores y otros usuarios de las pruebas de logro académico conozcan tanto los usos de las pruebas como también sus limitaciones. El Seminario también tuvo como propósito dar a conocer una selección de la producción académica del ICFES.

B. Participantes

Desde su primera versión, el Seminario ha estado dirigido a investigadores, profesores y estudiantes de educación superior, de diversas disciplinas incluyendo educación, psicología, sociología, economía, estadística y otras ciencias sociales desde las cuales exista interés en investigar sobre la calidad de la educación.

C. Programa Académico (Conferencias y Talleres)

El seminario se llevó a cabo en dos días, distribuido en conferencias principales y alternas, talleres, y exposición de estudios académicos mediante afiches.

Las *conferencias principales* estuvieron dedicadas a la presentación de estudios internacionales sobre calidad educativa. Como en años anteriores, los criterios para la selección de los estudios de las conferencias principales fueron los siguientes:

- Al seleccionar conferencias, se le da más importancia a las características de los estudios que a la trayectoria de los conferencistas.
- Se prefieren los estudios de investigación aplicada, que sugieran ideas para mejorar la calidad de la educación, bien sea en el aula o en el nivel de las políticas públicas.
- Los estudios deben emplear pruebas estandarizadas de logro académico, como las pruebas Saber.
- Se seleccionan estudios muy recientes, que ofrezcan enfoques y preguntas frescas, para enriquecer la discusión académica en el país.
- Se seleccionan estudios con diversidad de preguntas, disciplinas y métodos de investigación
- Se procura heterogeneidad de preguntas de estudios según el nivel educativo.
- Hay un especial esfuerzo por presencia de estudios de distintos países, con énfasis América Latina.
- Para el caso de las conferencias principales, éstos deben contar con alguna revisión previa de pares.

Las *conferencias alternas* tuvieron como propósito presentar estudios nacionales que reúnan las características arriba mencionadas, salvo las dos últimas.

En las *jornadas de la tarde* de ambos días se desarrollaron cuatro (4) talleres introductorios simultáneos, de capacitación sobre técnicas de investigación en calidad educativa, anunciados mediante la página web del Instituto, en la cual los interesados se inscribieron previamente.

El seminario concluyó con un conversatorio sobre el rol de las pruebas estandarizadas en la evaluación en Colombia, moderado por Adriana Molina, Jefe de la oficina asesora de gestión de proyectos de investigación del ICFES, donde intervinieron Juan Carlos Bayona, ex-rector del Gimnasio Moderno y ex-Sub Secretario de Educación de Bogotá, Diego Alexander Vargas, docente de matemáticas de la institución educativa El Chairá ubicada en el departamento del Caquetá, Francisco Piedrahita, rector de la Universidad ICESI de Cali, y Julián Mariño, Director de Evaluación del ICFES.

Jueves 1 de Noviembre (Mañana)

Conferencias principales

8:00	Registro de participantes
8:30	Apertura del III Seminario Internacional sobre Investigación en Calidad de Educación. Margarita Peña, Directora del ICFES
9:00	Usos correctos e incorrectos de las pruebas educativas para la evaluación. <i>Conferencista:</i> Daniel Koretz, PhD en Psicología del Desarrollo. Henry Lee Shattuck Professor of Education, Universidad de Harvard. <i>(Videoconferencia en vivo desde Boston, Estados Unidos. Traducción simultánea disponible).</i>
10:00	Preguntas (Bogotá y otras ciudades participantes)
10:50	Publicación de resultados de las pruebas y desempeño escolar. <i>Conferencista:</i> Vladimir Ponczek, PhD en Economía. Investigador de la Fundación Getulio Vargas, Brasil. <i>(Conferencia en inglés, traducción simultánea disponible)</i>
11:50	El sentimiento de auto-estima académica: situaciones relevantes para el sistema escolar chileno <i>Conferencista:</i> Rodrigo Roco Fossa, PhD. en Sociología y Economía de la Educación, Investigador Universidad de Chile.
12:50	Preguntas de los participantes a los conferencistas de la mañana. <i>Moderador:</i> Javier Sáenz Obregón, Centro de Estudios Sociales, Universidad Nacional. Miembro del comité asesor del Programa de Investigación del ICFES.
01:30	Cierre de la sesión

Jueves 1 de noviembre (Tarde)

Talleres y conferencias alternas

14:30 a 18:30	Taller de introducción a Modelos Lineales Jerárquicos (Sesión 1) <i>Instructor:</i> Andrés Martínez, PhD (Universidad de Michigan)	Taller sobre destrezas en la interpretación de pruebas estandarizadas de logro educativo <i>Instructores:</i> Richard Shavelson, PhD (Universidad de Stanford), Edward Wiley (Universidad de Colorado en Boulder) María Figueroa, PhD (Universidad de los Andes) y Javier Toro (ICFES).	Taller sobre uso de bases de datos de evaluaciones estandarizadas (sesión 1) <i>Instructores:</i> Carolina Lopera y Álvaro Uzaheta - ICFES	Conferencias alternas 5 conferencias de trabajos que hacen uso de las evaluaciones que realiza el ICFES (Ver programación abajo)
------------------	--	---	--	--

*** Sesiones de conferencias alternas I**

Moderado por *Lina María Saldarriaga*, PhD. Profesora Asistente, Departamento de Psicología, Universidad de los Andes. Miembro del comité asesor de Programa de Investigación del ICFES.

14:30	Calidad, Igualdad y Equidad en la Educación Colombiana. Análisis de la prueba SABER 2009. <i>Jesús Duarte, Soledad Bos y Martín Moreno. División de Educación del Banco Interamericano de Desarrollo. Washington D.C.</i>
15:00	Escolaridad, Habilidades Cognitivas y Crecimiento Económico Departamental en Colombia. <i>Camilo Leonardo Silva Díaz. Pontificia Universidad Javeriana.</i>
15:30	Relaciones entre programas y proyectos institucionales y el desempeño en pruebas SABER 5º y 9º. y SABER 11º en 12 instituciones públicas de Manizales. <i>Ligia Inés García Castro. Universidad Autónoma de Manizales, Observatorio de la Calidad de la Educación (FUNDECA) y la Fundación Luker.</i>
16:30	Evaluación de la escritura en el examen de estado de evaluación de la calidad de la Educación Superior en Colombia: la experiencia del ICFES. <i>Ernesto Cuchimaque; Carlos Ordóñez; Carlos Pardo. Instituto Colombiano para la Evaluación de la Educación – ICFES</i>
17:00	Métodos de equiparación en pruebas SABER 11º Colombia: Comparación de procedimientos en personas videntes e invidentes. <i>Aura Nidia Herrera Rojas, Lady Catheryne Lancheros Florián, Gillen Javier Jiménez López, María del Pilar Soler Parra y Angela María Espinosa Garzón. Universidad Nacional.</i>

Viernes 2 de noviembre (Mañana)

Conferencias principales

09:00	Clasificación de escuelas en la nueva institucionalidad educativa: contribución de modelos de valor agregado para una responsabilización justa <i>Conferencista: Ernesto San Martín, PhD en Ciencias. Profesor del Facultad de Educación y Facultad de Matemáticas, Pontificia Universidad Católica de Chile.</i>
10:00	Valor Agregado en la Educación Superior: Evidencia de Colombia <i>Conferencista: Ben Domingue, PhD. Universidad de Colorado en Boulder. (Conferencia en inglés, traducción simultánea disponible)</i>
11:20	La ampliación de la brecha socioeconómica en el logro escolar: nueva evidencia y posibles explicaciones <i>Conferencista: Sean Reardon, PhD en Educación. Profesor de la Universidad de Stanford, Facultad de Educación. (Conferencia en inglés, traducción simultánea disponible)</i>
12:20	Preguntas de los participantes a los conferencistas de la mañana. <i>Moderador: Francisco Pérez Calle, Máster en Administración Pública de la Universidad de Harvard, Director de G Exponencial Consultores.</i>
12:50	Clausura del seminario <i>Margarita Peña, Directora del ICFES.</i>
13:20	Cierre de la sesión

Viernes 2 de noviembre (Tarde)
Talleres, conversatorio y conferencias alternas.

14:30 a 18:30	Taller de introducción a Modelos Lineales Jerárquicos (Sesión 2)	Taller sobre uso de bases de datos de evaluaciones estandarizadas (sesión 2)	Taller de introducción a la inferencia causal en investigación cuantitativa de educación y ciencias sociales	14:30 Conferencias alternas II* <i>3 conferencias de trabajos que hacen uso de las evaluaciones que realiza el ICFES (Ver programación abajo)</i>
	<i>Instructor: Andrés Martínez (Candidato a PhD, Universidad de Michigan)</i>	<i>Instructores: Carolina Lopera y Álvaro Uzaheta - ICFES</i>	<i>Instructor: Sean Reardon, PhD (Universidad de Stanford)</i>	16:30 Conversatorio sobre el rol de las pruebas estandarizadas en la evaluación en Colombia. <i>Moderado por Adriana Molina, Jefe de la oficina asesora de gestión de proyectos de investigación del ICFES</i>

*** Sesiones de conferencias alternas II**

Moderado por *Adriana Molina*, Jefe de la oficina asesora de gestión de proyectos de investigación del ICFES

14:30	Desempeño relativo de los graduados en el área de educación en el examen de Estado del ICFES. <i>Juan D. Barón, Leonardo Bonilla, Mónica Ospina y Lina Cardona. Banco Mundial, Banco de la República, Universidad EAFIT.</i>
15:00	Diagnóstico del Valor Agregado en el nivel de lectura de los Egresados del Programa de Administración de Empresas de la Universidad Jorge Tadeo Lozano: Una propuesta desde el Benchmarking aplicando Data Envelopment Analysis – DEA <i>Federico Manuel Zevallos Peña. Universidad Jorge Tadeo Lozano.</i>
15:30	Benchmarking de universidades colombianas. <i>Daniel Bogoya, Universidad Jorge Tadeo Lozano.</i>

Las conferencias magistrales fueron transmitidas por video-conferencia en las Universidades mencionadas arriba a través de Renata, y pudieron seguirse por *video-streaming* a través de Internet.

3. Comparación de contenidos de los seminarios 2010 a 2012

La siguiente tabla muestra la distribución de las conferencias magistrales de los tres Seminarios anuales realizados hasta el momento de acuerdo a líneas de investigación. También se muestra la heterogeneidad de disciplinas de los estudios presentados. En cuanto a líneas de investigación (las filas de la tabla), éstas se organizan de investigación básica (las primeras filas) a investigación aplicada. Así, en la tabla hay tres grupos de líneas de investigación: metodología de investigación, evaluación de la

educación, sistemas de rendición de cuentas que emplean datos de logro académico, e intervenciones en campo para mejorar la calidad de la educación. En cuanto a heterogeneidad de disciplinas de los estudios seleccionados para las conferencias magistrales de los seminarios, éstas se agruparon en “educadores” (color naranja, con 4 conferencias en el periodo), “psicólogos y sociólogos” (color amarillo, con 5 conferencias en el periodo) y “economistas y estadísticos” (color rojo, con 10 conferencias en el periodo). La diferencia en cantidades de estudios presentados por disciplina está correlacionada con la probabilidad de que la respectiva disciplina utilice en sus estudios sobre calidad educativa datos de logro académico.

Ilustración 1 – Conferencias Magistrales de los Seminarios ICFES 2010 a 2012, por línea de investigación y disciplina

Seminario de Investigación en Calidad Educativa				
	Línea de investigación	2010	2011	2012
Investigación Básica	Metodología	Shavelson		
	Evaluación			Cuchimaque Herrera
	Medición			Koretz
Investigación Aplicada	Factores asociados al desempeño	Fernandes	Uzaheta	García
		Porta		Silva
	Valor agregado			San Martín / Domingue / Silva / Zevallos
	Sistemas de evaluación			Ponczek
	Intervenciones y logro			Roco
	Colegios	Bonilla	Bellei	
	Ciudadanía		Chaux	
			Cervantes, Cuevas	
	Currículo	Schmidt		
	Equidad			Duarte
			Willms	
	Género		Cabezas	
	Maestros		Lavy, Ome	Barón
	Pares		Martínez	
	Recursos		Valderrama	
Tiempo de estudio		Lavy, Menezes		
Universidades	Saavedra		Bogoya	

Convenciones:			
<i>ICFES in house</i>	Educador	Psicólogo/Sociólogo/ Ciencias Sociales	Economista / Estadístico / Ingeniero

4. Las presentaciones académicas

En este capítulo se presentan las principales ideas de las conferencias. Sus presentaciones al público y el texto del estudio respectivo se encuentran en la página web del ICFES. *Estos resúmenes no pretenden reemplazar la lectura de los libros y los estudios de los autores*, sino transmitir algunos de los planteamientos principales formulados en el evento. Antes de dar paso a las conferencias acerca de investigaciones nacionales e internacionales sobre educación, la directora del ICFES, Margarita Peña, dio unas palabras de introducción que serán resumidas a continuación.

Conferencista: Margarita Peña, PhD, Directora del ICFES

Intervención: Apertura del III Seminario Internacional sobre Investigación en Calidad de Educación

En sus palabras de instalación del Seminario, la directora del Instituto señaló que los sistemas de educación de distintos países están convergiendo hacia el uso recurrente de pruebas de logro educativo. En Brasil, por ejemplo, se adoptó desde 1998 el examen nacional de educación secundaria, ENEM. En México, desde 2008, se aplica la Evaluación Nacional del Logro Académico de Centros Escolares (ENLACE). En Chile se realizan desde los años ochenta las pruebas de lo que hoy se conoce como el Sistema Nacional de Medición de Resultados de Aprendizaje, SIMCE. En Estados Unidos una Ley del año 2001, conocida bajo el título “*No child left behind*” (que ningún niño quede rezagado), afianzó la medición del logro académico como mecanismo de rendición de cuentas en el sector educativo. Además de iniciativas al interior de los países, también se han consolidado las mediciones entre países como es el caso de las pruebas PISA, TIMSS e ICCS, que se realizan en forma periódica, y donde Colombia ha participado. En resumen, no hay señal de reversa sino de consolidación en el uso de las pruebas estandarizadas de logro en la educación.

Subrayó que en Colombia, el ICFES realiza pruebas estandarizadas para medir la calidad de la educación a los estudiantes del sistema escolar desde los 90, 20 años después de que iniciara en 1968 una importante tradición de medición con el examen de admisión a las universidades colombianas. Desde 2009, el instituto se volcó por mandato legal enteramente a la evaluación de la educación. La serie SABER, como se conocen ahora todas las evaluaciones nacionales de logro educativo que realiza el Instituto, es la concreción de un proyecto de muchos años, orientado a contar con herramientas válidas y confiables para evaluar la calidad de la educación y producir, a partir de ella, indicadores igualmente válidos con base en lo cual puedan tomarse decisiones.

Dijo también que en forma creciente, la sociedad y los decisores de política buscan en esta información bases objetivas para justificar decisiones de “consumo educativo” (por ejemplo ¿en qué colegio matriculo a mi hijo?), explicación sobre las causas del funcionamiento del sistema, explicaciones sobre qué funciona en educación para implementarlo de forma masiva. También se emplean para buscar “responsables”; el uso creciente de uso resultados de las evaluaciones “para pedir o rendir cuentas”, en ocasiones con efectos significativos sobre profesores, rectores, estudiantes e incluso mandatarios, llama la atención sobre la necesidad de saber con certeza, o por lo menos aproximarse a ello, lo que estas pruebas dicen sobre la calidad de la educación.

Concluyó este aparte diciendo que ningún médico se atrevería a diagnosticar el estado general de un paciente con base en una simple muestra de sangre... es una comparación exagerada, pero sirve para subrayar lo que está en juego. “De esto se trata este seminario”, aseguró.

Sobre la agenda de investigación del ICFES, mencionó que el Instituto desarrolla un programa de investigación para que el país disponga cada vez de mejores pruebas, y para fomentar la investigación aplicada mediante el uso de los datos que éstas suministran. Invitó a los investigadores de todas las disciplinas (educadores, economistas, sociólogos, antropólogos, estadísticos) a utilizar esta información para comprender mejor la educación, medirla, evaluar políticas, y analizar programas destinados a mejorarla.

Señaló que el Programa de Investigación del ICFES se va consolidando gradualmente, (i) con los estudios que hace el Instituto para mejorar las pruebas y producir conocimiento, (ii) con los estudios que financia mediante convocatorias competitivas para investigadores del país; (iii) y con un componente de capacitación de investigadores y actores de la educación, que se materializa en los seminarios anuales, como el que es objeto de estas memorias, y en cursos cortos de capacitación en metodologías de investigación cuantitativa en educación.

Concluyó diciendo que promover investigación empleando datos de las pruebas SABER implica que la comunidad académica y la sociedad en general conozcan su potencial, pero también las limitaciones que éstas tienen. Señaló la responsabilidad de mostrarle al país los usos apropiados de las pruebas SABER, pero también advertir sobre usos inapropiados de las mismas. Los resultados de la evaluación no pueden ser tomados a la ligera. Un buen uso de las pruebas nos ayuda a avanzar, a mejorar. Un mal uso de las pruebas, nos confunde.

Conferencista: Daniel Koretz, PhD en Psicología del Desarrollo. Henry Lee Shattuck. Profesor de Educación, Universidad de Harvard.

Conferencia: Usos correctos e incorrectos de las pruebas educativas para la evaluación.

Esta conferencia se basa en el libro titulado *Measuring Up*² (2008) del profesor Daniel Koretz, dedicado a mostrar en forma sencilla las complejidades del diseño y uso de las pruebas estandarizadas de logro educativo.

A manera de introducción, el profesor Koretz señaló que la investigación sobre los efectos en el logro escolar de aplicar pruebas estandarizadas, muestra que, los diseños de investigación y los datos obtenidos son débiles, y los efectos son modestos.

Los estudios disponibles muestran efectos mixtos sobre la práctica educativa: algunas mejoras, pero también algunos efectos indeseables, como la preparación inapropiada para las pruebas. Hace notar que los puntajes pueden ser inflados, es decir, aumentar mucho más que el aprendizaje real.

² El libro se conoce con el título de *El ABC de la evaluación educativa (2011)*, en una edición publicada como Ceneval, de México.

En su conferencia se refirió a los siguientes temas: (i) El “principio de muestreo” de las pruebas; (ii) Evidencia de inflación de puntajes; (iii) Respuestas a pruebas cuyos resultados generan consecuencias (high stakes testing): cómo se genera la inflación de puntajes; (iv) Implicaciones para el desarrollo de nuevos programas de pruebas y evaluación.

Ilustración 2 – Muestreo para construir una prueba de logro académico (Koretz)

1 Propósitos de la educación	
2 Logro de los estudiantes	Otros
3 Áreas seleccionadas para la prueba	Áreas no probadas
4 Partes probadas áreas seleccionadas	Partes de áreas no probadas
5 Muestra estudiada	Muestra no estudiada

A partir de la ilustración anterior, el profesor Koretz explicó que las pruebas cobijan una porción pequeña (la intersección de la última fila con la primera columna) de todos los temas que cobijan el trabajo educativo. Los efectos de lo que llamó “muestreo incompleto son: (i) la evaluación educativa es sistemáticamente incompleta; (ii) cuando la prueba se toma con baja presión, los efectos son modestos; (iii) cuando los resultados de las pruebas tienen importantes implicaciones para quienes las toman, los efectos son muy grandes: hay incentivos para concentrarse en la muestra utilizada, no en todo el área, se presentan riesgos de instrucción restringida y de mala preparación para la prueba, y hay inflación de puntajes.

Con relación a la inflación de puntajes, señaló que los puntajes sólo tienen significado si sirven para generalizar al área que se quiere medir. Si las ganancias son generalizables al área, éstas tienen que ser generalizables a otras pruebas dentro de la misma área. La inflación de puntajes socava la evaluación de dos maneras: bien sea porque la mejora total es exagerada, o porque la efectividad relativa (por ejemplo de las escuelas) es estimada incorrectamente.

El profesor Koretz dedicó parte de su presentación a mostrar las diferencias entre lo que llamó *buena* o *mala* preparación para las pruebas: la *buena preparación* entrega a los estudiantes conocimientos y destrezas que éstos pueden aplicar en cualquier parte, como estudios y empleos posteriores; en consecuencia, en otras pruebas. La *mala preparación* genera ganancias específicas en la prueba que no son generalizables más allá de la prueba.

Para aumentar los puntajes, señaló, existen caminos positivos y negativos. Entre los primeros se encuentran enseñar más, trabajar más duro y más efectivamente. Entre los caminos no deseables se encuentran la reasignación de estudiantes, el entrenamiento de estudiantes y el uso de trampas.

Sobre los sistemas de rendición de cuentas en educación basados en pruebas estandarizadas de logro, subrayó que el uso de una sola prueba para evaluar y rendir cuentas no es adecuado. Esto se debe, por un lado, a que las pruebas omiten muchos resultados importantes, y por otro, a que las pruebas que tienen consecuencias para el evaluado generan efectos mixtos en la práctica, y producen ganancias infladas en los puntajes.

Con respecto a los modelos de valor agregado (MVA) (aquellos que procuran estimar la contribución específica de la escuela o universidad al logro académico de sus estudiantes), el profesor Koretz señaló que son, en cierta medida, la mejor manera de medir resultados, pero implican temas adicionales. En particular, grandes cantidades de ruido aleatorio, resultados inestables entre pruebas, y dificultad para inferirlos verdaderos efectos de los maestros y las escuelas. Subrayó que los MVA no evitan el problema de la inflación de puntajes.

El profesor Koretz señaló que no hemos identificado los mejores tipos de evaluación basada en pruebas de logro ni los mejores sistemas de rendición de cuentas en educación. Por consiguiente debemos seguir investigando para poder entender cuáles programas maximizan mejoras reales, cuáles minimizan trucos y mala preparación de estudiantes para pruebas, y cómo evitar la inflación de puntajes.

Como sugerencias planteó la importancia de llevar a cabo un monitoreo y evaluación de la educación y de los sistemas de rendición de cuentas basados en pruebas de logro, en forma continua; sugirió evaluar el sistema de evaluación y no solo la educación; y probar nuevos diseños de pruebas de logro y nuevos diseños de sistemas de rendición de cuentas. Sugirió la necesidad de monitorear las respuestas de comportamiento de los educadores frente a las pruebas de logro que toman sus estudiantes, y la importancia de investigar sobre variaciones en los efectos, por ejemplo entre tipos de escuelas y tipos de estudiantes, para estimar mejor las ganancias reales y mejorar los incentivos.

Finalmente, sugirió minimizar en las pruebas la repetición innecesaria de contenidos y los estilos de presentación de los ítems. Y planteó la necesidad de construir pruebas “con auditoría”.

Daniel Koretz es Doctor en Psicología del Desarrollo de la Universidad de Cornell y profesor de la Facultad de Educación de Harvard. Su investigación ha estado dedicada a la valoración educativa como herramienta de política pública, incluyendo las pruebas estandarizadas y los sistemas de rendición de cuentas en educación.

Información en línea sobre el trabajo de Daniel Koretz:

<http://www.gse.harvard.edu/directory/faculty/faculty-detail/?fc=47648&flt=k&sub=all>

Conferencista: Vladimir Ponczek, PhD en Economía. Investigador de la Fundación Getulio Vargas, Brasil.

Conferencia: Publicación de resultados de las pruebas y desempeño escolar.

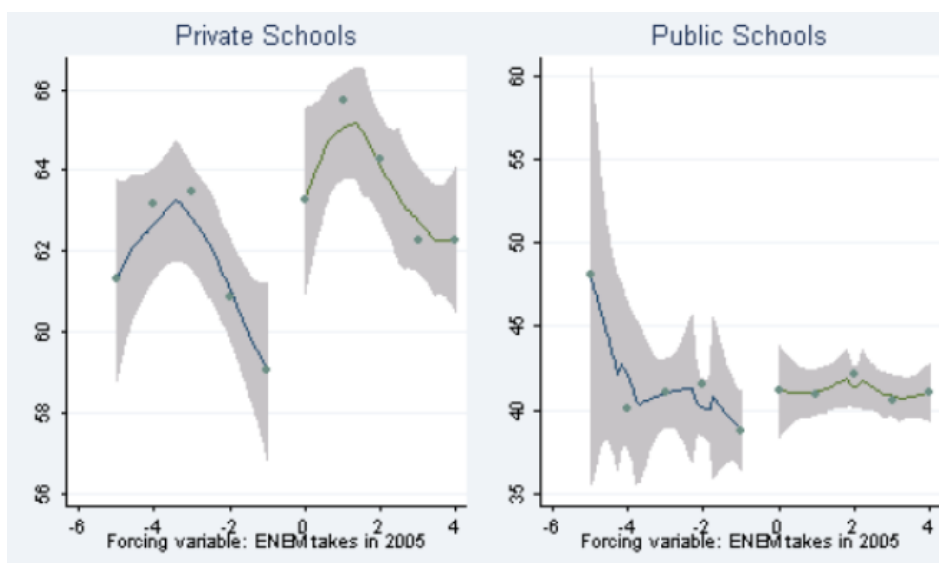
En el estudio, “Publicación de resultados de las pruebas y desempeño escolar” Vladimir Ponczek, Braz Camargo, Rafael Camelo y Sergio Firpo exploran los efectos de las señales de la calidad escolar (específicamente la publicación de los resultados de las pruebas), sobre el desempeño de los estudiantes en la prueba ENEM³ del Brasil. Puntualmente, se preguntan ¿cómo reaccionan los estudiantes y las escuelas a señales de calidad escolar?

La pregunta principal del estudio es si la publicación de los resultados de las pruebas puede servir como incentivo para mejorar la educación. El argumento es que la publicación de resultados puede afectar: (i) el esfuerzo de los estudiantes o padres, pues la información sobre la calidad escolar modifica la decisión del estudiante sobre cuánto esfuerzo hacer en su actividad escolar; una señal negativa induce más esfuerzo; (ii) el esfuerzo de la escuela (maestros, rectores, insumos). Sobre este último punto, el estudio reconoce el hecho de que la escuela ignora la función de producción de la educación, es decir, la combinación necesaria de esfuerzos para mejorar los puntajes. No obstante, la señal que proviene de los puntajes revela las debilidades de la escuela. En ese sentido, se podrían esperar impactos sobre los insumos observables en la escuela.

El estudio asume que existe heterogeneidad en las reglas de publicación de los resultados de las pruebas de logro escolar de Brasil (ENEM), es decir que no todas las escuelas censadas en el área metropolitana de Sao Paulo están expuestas a la publicación de los resultados. Por lo anterior, utiliza una regresión discontinua que permite comparar el grupo *tratamiento* (en este caso, expuestos a la publicación de resultados, que es el objeto de la evaluación) y el grupo de *control* (en este caso, el grupo de los *no* expuestos a la publicación de resultados). La diferencia entre ambos grupos es el efecto o impacto evaluado (**Ilustración 3**).

³Los puntajes ENEM son empleados como criterio de admisión en varias universidades públicas y privadas. También se utilizan en la selección de beneficiarios del programa federal de becas universitarias. No es un examen obligatorio. Desde 2006, el gobierno divulga anualmente el promedio de los puntajes de los colegios el año anterior. Solamente se divulgan puntajes de colegios cuando 10 o más estudiantes son evaluados. El puntaje del colegio es el promedio de todos los estudiantes que terminaron secundaria ese año. Los autores del estudio emplean en particular las bases de datos de la ENEM de 2005 y 2007, y el censo de escuelas.

Ilustración 3 – Impacto de la publicación de resultados en colegios públicos y privados



En la ilustración se aprecia que solamente existe diferencia entre los resultados de los grupos tratamiento y control en los colegios privados. La publicación de puntajes no revela una diferencia en el desempeño de los colegios oficiales de Brasil.

Los autores señalan que la publicación de puntajes de pruebas en el 2005 no tiene impacto en las características observables de las escuelas en 2007. No obstante, identifican que la publicación tiene un impacto sobre los puntajes del 2007 en los colegios privados. Este impacto es heterogéneo, pues mientras que en los mejores colegios la publicación no tiene ningún efecto, en los peores colegios hay efectos positivos. Con respecto al esfuerzo estudiantil, el estudio encuentra evidencia de que la divulgación de puntajes de pruebas de logro sí impacta el esfuerzo, observando esta variable en el trabajo y clases extra.

Los autores señalan cuatro eventuales limitaciones de sus estimaciones empíricas, a saber: (i) los tomadores de las pruebas podrían “estafar” al sistema de tal suerte que escuelas tratadas (con publicación de sus resultados) podrían inducir a tomar el examen sólo a los mejores estudiantes. No obstante, los estudiantes son los responsables de la inscripción, y la participación en la prueba en los colegios privados es cercano al 90%; (ii) un riesgo de composición (de la forma como están compuestos los colegios en términos de las características de sus estudiantes): los mejores estudiantes podrían asistir a escuelas cuyos resultados se publican; (iii) selección de escuelas: solo las buenas escuelas, entre las tratadas, sobreviven. No obstante, sólo 45 escuelas (6%) desaparecen de la muestra, y no hay diferencia significativa entre tratadas y no tratadas; (iv) puede haber preocupación por la trayectoria: escuelas tratadas y no tratadas podrían asignar diferentes probabilidades a la publicación futura de sus puntajes promedio en las pruebas, caso en el cual el sesgo de la estimación del efecto de la publicación de las pruebas en los puntajes ulteriores sería hacia abajo.

Como conclusiones, los autores señalan que la publicación de los resultados de las pruebas de logro escolar en Brasil mejora el desempeño promedio en estas pruebas en colegios privados. En ese sentido, consideran que los incentivos de mercado repercuten

significativamente sobre la educación. Los autores sostienen que las mejoras observadas provienen básicamente de aumentos en el esfuerzo, pues no encontraron ningún cambio en la composición de los estudiantes, ni de insumos en las escuelas.

Vladimir Ponczek es Doctor en Economía de la Universidad de Princeton y Profesor de la Facultad de Economía de la Fundación Getulio Vargas de Sao Paulo, Brasil. Dedicó su investigación a la economía del desarrollo y la economía laboral.

Información en línea sobre el trabajo de Vladimir Ponczek:

<https://sites.google.com/site/vponczek/working-papers>

Conferencista: Rodrigo Roco Fossa, PhD en Ciencias de la Educación por el IREDU/CNRS de la Universidad de la Bourgogne, Francia. Investigador Universidad de Chile.

Conferencia: El sentimiento de auto-estima académica: situaciones relevantes para el sistema escolar chileno.

El trabajo presentado por el autor tiene el objetivo de analizar fenómenos que son parte de la vida y el contexto escolar y que poseen valor como dimensiones en sí mismas, pero que a la vez influyen y determinan las posibilidades de mejorar la equidad y calidad de los aprendizajes escolares. Así mismo, busca confirmar o descartar la presencia del efecto “Pez-grande en Laguna Pequeña” (BFLPE, por su sigla en inglés) y sus características en el contexto nacional e identificar variables relevantes asociadas al sentimiento de autoestima académica que permitan enriquecer la discusión sobre políticas educacionales en Chile. Para ello, se analizan datos de 8vo Básico de 2004 y de 2009, construyendo tres índices de autoestima académica, dos de orden general y uno referido específicamente a matemáticas. Se discuten además las posibilidades de construcción y validación de tales índices, para luego concentrarse en algunas variables significativas de la realidad escolar chilena (género y repetencia).

En su conferencia, el profesor Roco justifica la importancia de las percepciones subjetivas de los estudiantes sobre su experiencia escolar, en su vida escolar y futura. Dentro de estas percepciones menciona la autoestima académica, el bienestar en la escuela, la percepción del apoyo y exigencia de los padres, las expectativas de escolarización posterior, la relación con los profesores, la afinidad declarada para con las disciplinas, la importancia declarada respecto a las disciplinas y las razones generales para ir a la escuela. Su estudio se concentra en las dos primeras percepciones: autoestima académica y bienestar en la escuela, y su relación con el logro.

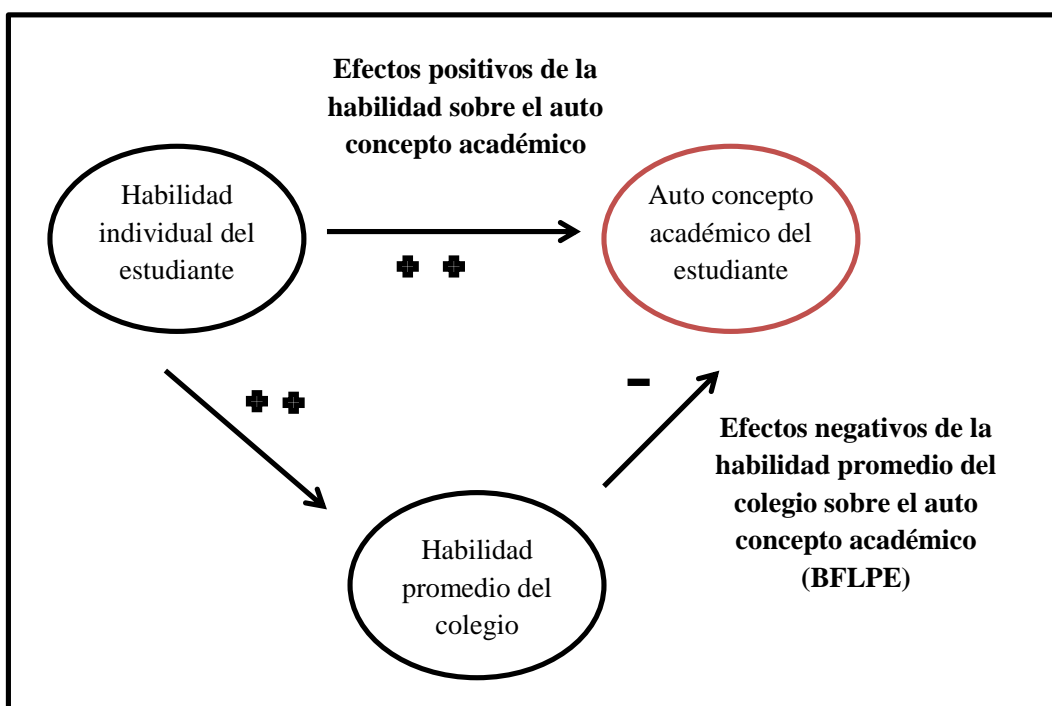
Las llamadas dimensiones subjetivas de la experiencia escolar (mal llamadas “no-cognitivas”, según indica el autor) son parte substancial de ambientes que facilitan o dificultan el trabajo escolar; a nivel individual operan como factores estrechamente asociados a los aprendizajes académicos; poseen valor intrínseco en tanto tales; constituyen aprendizajes y forman parte del desarrollo de la persona.

Citando a Hascher (2008) define el bienestar en la escuela como una experiencia emocional caracterizada por la dominancia de sentimientos y cogniciones positivos hacia la escuela, las personas en la escuela y el contexto escolar. El bienestar en la escuela representa una evaluación subjetiva, emocional y cognitiva de la realidad

escolar y puede verse como un desbalance entre aspectos positivos y negativos. El profesor Roco subraya que la relación entre el bienestar escolar y los resultados escolares es más fuerte durante la adolescencia (Knuver y Brandsma, 1993).

Para su ejercicio empírico, el autor formula como hipótesis para Chile la existencia de la presencia del fenómeno “pez grande en laguna pequeña” (BFLPE por sus siglas en inglés). El efecto BFLPE señala que estudiantes con habilidades semejantes tenderán a tener una menor autoestima académica en colegios de alto desempeño que en colegios de menor desempeño. Este efecto es resumido en la **Ilustración 4**.

Ilustración 4 – Predicciones teóricas del efecto Pez-grande en Laguna Pequeña



Como método empírico de investigación, el autor realiza un análisis factorial de constructos suyos sobre auto-estima de los estudiantes, con base en lo que éstos reportaron en un cuestionario de factores asociados a la calidad, aplicado en conjunto con las pruebas de logro SIMCE de Chile. Como controles, el autor incluyó un amplio número de variables tales como el índice de nivel socioeconómico (NSE) del grupo familiar (escolaridad ambos padres, ingreso per cápita, bienes), el capital cultural familiar (libros en el hogar), el género del alumno (mujer v/s hombre), el tipo de familia (biparental v/s otra), la existencia de hermanos, la repitencia escolar previa, el paso por pre-escolar temprano, las expectativas de escolarización posterior (alumno y apoderados), el tamaño de la clase, la heterogeneidad académica de la clase y otros.

Con base en este ejercicio, el autor concluye que la autoestima académica y el bienestar escolar son componentes presentes e importantes de la realidad escolar chilena en sus diferentes contextos. Concluye también que el género y la repitencia tienen una importante asociación con la autoestima académica (destaca la autopercepción en matemáticas en el caso de las mujeres). Señala que la repitencia está muy relacionada

con el bienestar en la escuela. Concluye también que el fenómeno de comparación social BFLPE está presente en la escuela chilena: ambientes más selectivos tenderían a erosionar, a condiciones comparables e igual desempeño académico, la autoestima de los jóvenes del nivel octavo básico en el caso chileno.

Rodrigo Roco es PhD en Ciencias de la Educación por el IREDU/CNRS de la Universidad de la Bourgogne, Francia. Fue presidente de la Federación de Estudiantes de la Universidad de Chile y es miembro del Senado de la Universidad.

Conferencista: Ernesto San Martín, PhD en Ciencias. Profesor del Facultad de Educación y Facultad de Matemáticas, Pontificia Universidad Católica de Chile.
Conferencia: Clasificación de escuelas en la nueva institucionalidad educativa: contribución de modelos de valor agregado para una responsabilización justa

El Congreso de Chile adoptó recientemente nuevos mecanismos de rendición de cuentas y presión por resultados en su sistema escolar. En particular, creó un Sistema Nacional de Aseguramiento de la Calidad de la Educación, que incluye a la Superintendencia de Educación y la Agencia de Calidad de la Educación. Esta última busca transformar la relación existente entre Estado, familias y establecimientos educativos, mediante el uso intensivo de pruebas estandarizadas de logro de los estudiantes, y la difusión pública de los resultados y consecuente categorización de escuelas. La Agencia deberá clasificar a los establecimientos educativos en términos de su efectividad escolar. Dadas las múltiples metodologías disponibles, *la clasificación que realice la Agencia podría potencialmente juzgar y responsabilizar injustamente a algunos establecimientos escolares, ya sea subestimando su desempeño, o bien sobrestimando el de aquellas escuelas que sólo en apariencia obtienen buenos resultados*. Estimar el efecto-escuela constituye uno de los desafíos metodológicos de mayor complejidad de la investigación educativa contemporánea. El estudio presentado por Ernesto San Martín contribuye al debate, mediante una ilustración empírica de las marcadas diferencias entre la posición que ocupan las escuelas en un esquema de clasificación a partir de un modelo de valor agregado, y la posición que ocupan a partir de promedios SIMCE a nivel de escuelas.

El profesor San Martín dedicó la primera parte de su exposición a presentar algunas características del sistema educacional chileno y su nueva institucionalidad. Al respecto dijo que el cuasi-mercado educacional chileno se caracteriza por tener un sistema de *vouchers* (subsidios a la demanda) instaurado a finales de los años 80, que permite a los padres elegir la escuela a la cual envían a sus hijos, y disponibilidad de información sobre rendimientos educativos agregados a nivel de escuela. Los rendimientos educativos son medidos por medio de la prueba nacional SIMCE (Sistema de Medición de la Calidad de la Educación). Inicialmente el SIMCE era aplicado alternadamente a los niveles 4, 8 y 10. Desde el 2005, el SIMCE se aplica cada año al nivel 4, y de forma alternada a los niveles 8 y 10.

La ley chilena introduce un Sistema Nacional de Aseguramiento de la Calidad de la Educación: la Superintendencia de la Educación y la Agencia para la Calidad de la Educación. La ley aprobada define que las escuelas serán clasificadas dentro de cuatro grupos, según los niveles de desempeño alcanzados, cuya ponderación incluirá, en no menos de un 67% del total, los estándares de aprendizaje medidos en pruebas censales

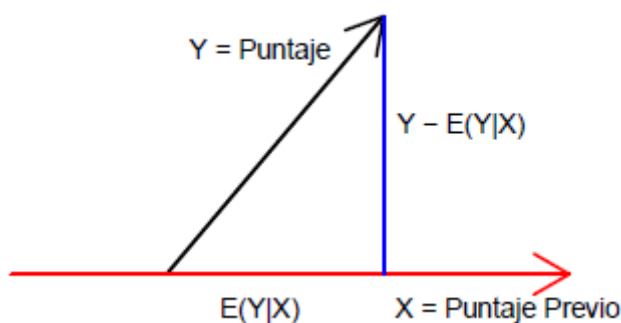
(e.g. SIMCE) alcanzados por los establecimientos. *La clasificación tiene consecuencias, desde la intervención de escuelas hasta su cierre.*

La clasificación que realizará la Agencia de Calidad se relaciona con lo que la literatura internacional ha llamado *efectividad escolar*. La Agencia de Calidad debe, por lo tanto, *conceptualizar el efecto escuela*. Debe además emplear metodologías de estimación del efecto escuela que reflejen válidamente la contribución neta de los establecimientos al progreso escolar de los estudiantes. La Ley chilena plantea que el procedimiento para realizar dicha clasificación podría eventualmente incluir modelos de valor agregado.

El autor entonces clasifica a las escuelas usando promedios SIMCE y responde a una primera pregunta: ¿cuál fue el desempeño de las escuelas municipalizadas (MC y MD) y particulares subvencionadas (PS) según los resultados del SIMCE2009? Luego se pregunta *¿qué tan justa resulta esta clasificación de desempeño escolar?* Responder a esta pregunta significa caracterizar todo lo que confluente en un puntaje SIMCE, e identificar *cuánto de este aprendizaje es posible atribuir a las escuelas*. Para ello, es necesario desarrollar una discusión conceptual en torno a dos conceptos claves: *efecto escuela y valor agregado*.

En la **Ilustración 5** a continuación se encuentra una representación de una intuición geométrica sobre el Valor Agregado. En la gráfica, Y representa el puntaje SIMCE-2009. Se busca explicar esta variable utilizando la medición SIMCE-2005, denotada por X. Estos requerimientos pueden ser ejemplificados por medio de una intuición geométrica básica, que a su vez tiene la ventaja de permitir una formalización general.

Ilustración 5 –Intuición Geométrica Básica sobre el Valor Agregado



Las relaciones geométricas anteriores se pueden resumir de la siguiente manera:

$$Y = E(Y|X) + \{Y - E(Y|X)\}$$

Donde:

- (i) $E(Y|X)$ es la proyección ortogonal de Y sobre el espacio de funciones de X. En otras palabras, esta expresión representa la proyección de los puntajes SIMCE-2009 sobre los puntajes SIMCE-2005.
- (ii) $\{Y - E(Y|X)\}$ es la proyección ortogonal sobre el espacio ortogonal al espacio de funciones de X. Típicamente, esta diferencia se llama residuo de Y sobre X. En

términos concretos, el residuo se refiere a los puntajes SIMCE-2009 que no tienen relación con los puntajes SIMCE-2005.

(iii) $E(Y|X)$ y $\{Y - E(Y|X)\}$ son no-correlacionados u ortogonales. En otras palabras, el hecho de que X explique Y significa que existe una relación estructural de dependencia, pero que ésta no es necesariamente explicada por X en su totalidad.

Según los modelos de Valor Agregado, si una escuela tiene un indicador de *valor agregado positivo*, podemos concluir que dicha escuela fue capaz de llevar a los estudiantes más allá de donde hubiesen llegado de haber sido tratados por una escuela promedio. Si una escuela tiene un indicador de *valor agregado negativo*, podemos decir que dicha escuela no fue capaz de llevar a los estudiantes más allá de donde hubiesen llegado de haber sido tratados por una escuela promedio.

La definición de valor agregado muestra que los indicadores de valor agregado *son relativos a una población de escuelas*: si ésta cambia, aquéllos pueden cambiar. Así mismo, el valor agregado depende del conjunto de covariables observadas, las cuales caracterizan la *escuela promedio*. En un sistema de rendición de cuentas en educación que utilice datos de logro, la elección de estas covariables es una señal de política pública, más que una decisión metodológica que busca estabilidad de los estimadores de valor agregado. Las consideraciones anteriores permiten formular un criterio de selección de covariables: incluir factores que describen procesos de selección de estudiantes por parte de las escuelas.

El autor diferencia entre efecto escuela y valor agregado. El *efecto escuela* es lo que permite explicar la heterogeneidad de las variables dependientes anidadas en dicha escuela y que no es explicada por las covariables. Por otro lado, con puntajes esperados en la escuela promedio, el *valor agregado* es una comparación de puntajes esperados condicionales a un conjunto de covariables y al efecto escuela. De acuerdo con lo anterior, el hecho de que una escuela tenga un indicador de *valor agregado positivo* indica que se espera que los estudiantes que asisten a esa escuela tengan un puntaje superior al que hubieran obtenido en caso de haber asistido a una *escuela promedio*, dentro de la *población de escuelas* consideradas en el estudio.

En cuanto a la metodología, el autor señala la existencia de endogeneidad estructural: en efecto, los estudios de efectividad escolar utilizan (al menos) dos mediciones a nivel individual, que el autor llama puntaje previo y puntaje actual/contemporáneo. Excepto por algunos diseños específicos, *el puntaje previo ya ha sido afectado por el efecto escuela* debido a la poca movilidad de los estudiantes entre una y otra medición. Esta situación no es excepcional y se presenta, por ejemplo, cuando se estudia la persistencia del valor agregado en el tiempo. Decimos entonces que *el puntaje previo es endógeno con respecto al efecto escuela*. Para resolver el problema de endogeneidad, el autor utiliza variables instrumentales, es decir, una variable que no está correlacionada con el efecto escuela.

En las estimaciones empíricas presentadas por el profesor San Martín, la variable dependiente es el puntaje matemáticas SIMCE 2009. Las covariables endógenas son el puntaje en matemáticas SIMCE 2009, y el puntaje promedio de estudiantes que pertenecen a un mismo establecimiento educativo en 2009. La covariable exógena es la proporción de padres que reporta que sus hijos/as fueron seleccionados por la escuela en

la que están en el 2009. Sobre las variables instrumentales, San Martín recordó que el número de instrumentos debe ser mayor o igual al número de factores endógenos considerados en el modelo. Luego escogió el nivel educativo de los padres (la movilidad social chilena de los últimos 20 años justifica su elección como variable instrumental); el número de personas que habita en el hogar del estudiante (este es un factor que las escuelas no controlan, lo que justifica su elección como instrumento); y la proporción de estudiantes que, al momento de responder el SIMCE, escribió correctamente su número de identidad nacional en ambas aplicaciones (esta información es ciega para cada establecimiento educativo, lo que justifica su elección como variable instrumental).

El autor subraya que si los estimadores obtenidos de valor-agregado no tienen en cuenta el problema de endogeneidad no se verían diferencias de efectividad; cuando, por el contrario, si se controla la endogeneidad, se encuentran diferencias entre las escuelas.

El profesor San Martín concluyó señalando que ignorar los problemas de endogeneidad estructural tiene un impacto dramático sobre el ordenamiento de las escuelas. Por lo tanto, controlar los problemas de endogeneidad asegura una responsabilización justa. La elección de instrumentos es, por tanto, un aspecto clave. Un sistema de responsabilización debe conceptualizar los instrumentos antes de la recolección de datos y de la aplicación de los instrumentos de medición. Los modelos de valor agregado nos permiten identificar escuelas efectivas, pero no nos permiten explicar a qué se debe esa efectividad. Sólo observando las escuelas efectivas y no-efectivas podremos adquirir una noción de la efectividad escolar.

Ernesto San Martín es PhD en Ciencias. Profesor del Facultad de Educación y Facultad de Matemáticas, Pontificia Universidad Católica de Chile. Sus áreas de investigación son la psicometría y la medición educacional; la estadística matemática, la historia de la estadística y la probabilidad, y la enseñanza de las probabilidades. Estas investigaciones las ha desarrollado por medio de diversos proyectos FONDECYT, proyectos bilaterales con la K.U. Leuven (Bélgica), proyectos de innovación y desarrollo (FONDEF, CORFO-Innova), y proyectos con el Ministerio de Educación (Estudio Piloto de Valor Agregado; Proyectos FONIDE de análisis de videos del Docente Más; entre otros)⁴.

Conferencista: Ben Domingue, PhD en Educación. Universidad de Colorado en Boulder.

Conferencia: Valor Agregado en la Educación Superior: Evidencia de Colombia.

Los modelos de valor agregado, o las estimaciones del cambio en el desempeño de los estudiantes en un periodo de tiempo atribuible a los maestros o las escuelas, ha sido un área importante de investigación, así como un componente de algunas iniciativas de política, en Estados Unidos y Europa. Estos estudios se han centrado en el nivel de primaria y secundaria y no se han desarrollado en educación superior debido, en parte, a que en otras partes del mundo no existen resultados de pruebas de todos los estudiantes universitarios al finalizar su formación. La prueba SABER PRO en Colombia es tal resultado. En esta conferencia, Ben Domingue esbozó muchos de los problemas de la

⁴Información extraída de la página web de la Universidad Católica de Chile (<http://mideuc.cl>).

estimación de modelos específicos de valor agregado con ejemplos empíricos tomados del contexto colombiano.

El estudio de Domingue está dirigido a medir si las instituciones educativas de educación superior influyen el aprendizaje estudiantil. Para lograrlo, plantea, idealmente se requiere una medida de logro estudiantil antes y después de que los estudiantes reciban educación superior. En Colombia estas medidas son el Saber 11 y el examen Saber PRO.

El estudio de Domingue examina la efectividad de las universidades usando datos longitudinales, obtenidos de vincular puntajes SABER PRO y SABER11. El puntaje de salida de un estudiante de universidad (resultados) será el puntaje del estudiante de la administración de SABER PRO 2011. El puntaje de entrada de un estudiante a la universidad (inicial) será el puntaje del estudiante en SABER 11. También utilizó datos de la encuesta administrada a estudiantes cuando tomaron SABER PRO.

Para alcanzar el objetivo de medir la efectividad universitaria utilizando datos longitudinales, se requiere una medida de variación en la efectividad entre instituciones educativas que genere igualdad de condiciones con respecto a diferencias en el desempeño de quienes entran y las características de los estudiantes. Los modelos estadísticos que intentan medir la efectividad de las instituciones educativas se conocen como *Modelos de Valor Agregado (VA)*.

Un Modelo de Valor Agregado representa una aproximación estadística que procura aislar el efecto específico de las instituciones educativas de otros actores, por ejemplo la riqueza de los padres y la motivación estudiantil. Aunque hay muchas maneras para especificar el Modelo de Valor Agregado todas tienen dos cosas en común: (i) todos los modelos de Valor Agregado controlan por el nivel de logro de los estudiantes que llegan. Esto se hace *para asegurar que a la institución educativa no se le pide injustamente responder por diferencias preexistentes entre los estudiantes*. (ii) Ningún Modelo de Valor Agregado sirve para explicar *por qué* algunas instituciones aparecen más efectivas que otras. Comprender las diferencias en prácticas educativas que causan que los estudiantes aprendan más en determinados ambientes requiere de investigación adicional focalizada.

Una opción para hacer el estudio es tomar el puntaje de cada estudiante en SABER PRO y restarle el puntaje SABER 11, y después promediar la diferencia para cada universidad. Sin embargo, esto sería insuficiente para explicar las diferencias entre instituciones en cuanto a los tipos de estudiantes que reciben. Por ejemplo, algunas reciben estudiantes con altos puntajes en SABER 11 y otras reciben estudiantes con bajos puntajes en SABER 11. Adicionalmente, las instituciones tienen mezclas muy diferentes en las características socio demográficas de sus estudiantes. El autor señala, por ejemplo, que la mediana de estrato socioeconómico de los estudiantes en algunas instituciones se ubica en el primer estrato, cuando en otras instituciones la mediana está en el quinto. ¿Cómo podemos hacer comparaciones justas y sensatas entre grupos tan diferentes?

Aquí es donde entran los Modelos de Valor Agregado. Se construye un modelo de Valor Agregado en forma tal que procure ajustar por diferencias en los grupos de

estudiantes que asisten a las distintas universidades: diferencias en puntajes SABER 11, y/o diferencias en características demográficas de los estudiantes, como el estatus económico o la educación de la madre.

Señala el autor que, al construir un modelo de VA para este estudio, se formuló varias preguntas: ¿Qué hacer si tenemos puntajes de SABER 9 y SABER 11? ¿Debemos usarlos ambos? ¿Debemos ajustar por diferencias en la población estudiantil usando cada una de las medidas demográficas de los estudiantes que están disponibles? ¿Solamente un subconjunto? ¿Algunas variables son más importantes que otras? ¿Cómo manejar los datos faltantes? ¿Debemos excluir estudiantes que no tienen los datos completos? La mayoría estudiantes completó SABER PRO en los semestres 9 y 10 después de matricularse. Sin embargo, algunos estudiantes tomaron la prueba en el primero, segundo, o tercer semestre. ¿Debemos tener en cuenta estas diferencias en el modelo de Valor Agregado?

El principal desafío al hacer análisis de Valor Agregado, en consecuencia, es que cada decisión tomada en el camino *puede generar cambios en la estimación de la efectividad de algunas universidades*. En la mayoría de los casos no hay una única combinación obvia decisiones generalmente aceptadas por la mayoría de estadísticos y diseñadores de políticas. En consecuencia, desarrollar un modelo de VA requiere de una gran cantidad de previsión, investigación, y consideración de alternativas.

Puesto que controlar por diferencias demográficas será importante para hacer comparaciones razonables, es importante definir las variables demográficas que vamos a utilizar. La inclusión de variables demográficas permite que los puntajes esperados de SABER PRO varíen de acuerdo a estas variables. Por ejemplo, podríamos esperar un menor crecimiento para estudiantes que trabajaban mientras asistían a la escuela. Es importante recordar que no solamente los estudiantes varían en términos de sus características demográficas; las instituciones también tienen distintos tipos de estudiantes según sus características demográficas.

En materia de crecimiento versus estrato económico, como se espera, el autor encuentra una correlación positiva entre el estrato socioeconómico del estudiante promedio la institución y el estimador de efectividad. Dicho esto, a lo largo del espectro de composición socioeconómica seguimos observando variación sustancial en efectividad entre instituciones que comparten niveles similares de composición socioeconómica.

De la misma manera como los modelos VA puede ser usados para estimar la efectividad de instituciones de educación superior, también pueden ser usados para estimar la efectividad de otros aspectos de la educación terciaria. Por ejemplo, el autor encuentra que las instituciones acreditadas fueron estimadas en 0.3 unidades más efectivas que las instituciones no acreditadas. Relativo al estudiante general (Desviación estándar de 1 en esta prueba), es de un tamaño importante.

El profesor Domingue concluyó recordando que Colombia está en una posición global única para hacer este tipo de investigación dados los datos que tiene el ICFES para hacerlo. Y dijo que un uso potencial de los estimadores de efectividad obtenidos con modelos de VA sería ofrecer algo así como una medida de “efecto por dólar” para todas

las instituciones educativas del país. Incluso a pesar de las limitaciones, considera, éste sería un importante indicador para los actores involucrados.

Ben Domingue es PhD en Metodologías de Investigación y Evaluación de la Universidad de Colorado en Boulder. Es investigador asociado del Instituto de Ciencias del Comportamiento en esa universidad. Recientemente, ha sido coautor de estudios acerca de los efectos de los factores genéticos y sociales sobre las redes de amistades, el papel de las políticas escolares y los contextos sociales como moderadores ambientales, la inequidad y las escuelas entre otros.

Conferencista: Sean Reardon, Doctor en Educación. Profesor de la Universidad de Stanford, Facultad de Educación.

Conferencia: La ampliación de la brecha socioeconómica en el logro escolar: nueva evidencia y posibles explicaciones.

Esta conferencia se basa en el capítulo del autor, contenido en el libro de R. Murnane & G. Duncan (Editores) (2011) *¿Dónde hay oportunidades? La desigualdad creciente y las oportunidades de vida inciertos de niños de bajos ingresos*⁵.

A manera de contexto, el profesor Reardon señala que, en promedio, los estudiantes de familias de mayores estratos socio-económicos tienen mejor desempeño en pruebas académicas, consiguen mayor escolaridad y como consecuencia consiguen mejor nivel socioeconómico cuando son adultos. Sin embargo, el alcance de la reproducción social (la fortaleza de la correlación entre el estatus socioeconómico de los padres y los resultados de los hijos) es variable: puede cambiar en el tiempo y el espacio, como resultado de políticas, normas, valores y condiciones económicas.

Las preguntas principales que se formula el autor en su estudio son las siguientes: ¿Qué tan grande es la brecha de logro académico en Estados Unidos? Específicamente, ¿qué tan grande es la brecha ingreso-logro? ¿Cómo se compara con otros países? ¿Cómo ha cambiado esta brecha en los últimos 50 años? ¿Qué explica los cambios observados?

Para realizar su estudio, el profesor Reardon combina 13 estudios⁶ con información de logro educativo, los cuales incluyen cohortes de estudiantes nacidos entre 1943 y 2001 y que presentaron pruebas entre 1960 y 2009. Todos los estudios analizados por el autor cumplen tres criterios: tienen muestras representativas a nivel nacional, incluyen resultados en pruebas estandarizadas de logro, e incluyen información sobre ingresos familiares.

El autor señaló las dificultades de medir la brecha entre estudiantes de altos y bajos ingresos teniendo 13 estudios distintos para hacerlo: el ingreso es continuo, no una variable binaria; la distribución del ingreso cambia en el tiempo; el ingreso se reporta en categorías diferentes en los estudios; el ingreso se reporta con error. Además, hay complicaciones al comparar *brechas* entre los distintos estudios porque el contenido de las pruebas es diferente; las escalas de las pruebas son diferentes; la confiabilidad de las pruebas es diferente; y las muestras son diferentes (en edad/grado, representatividad).

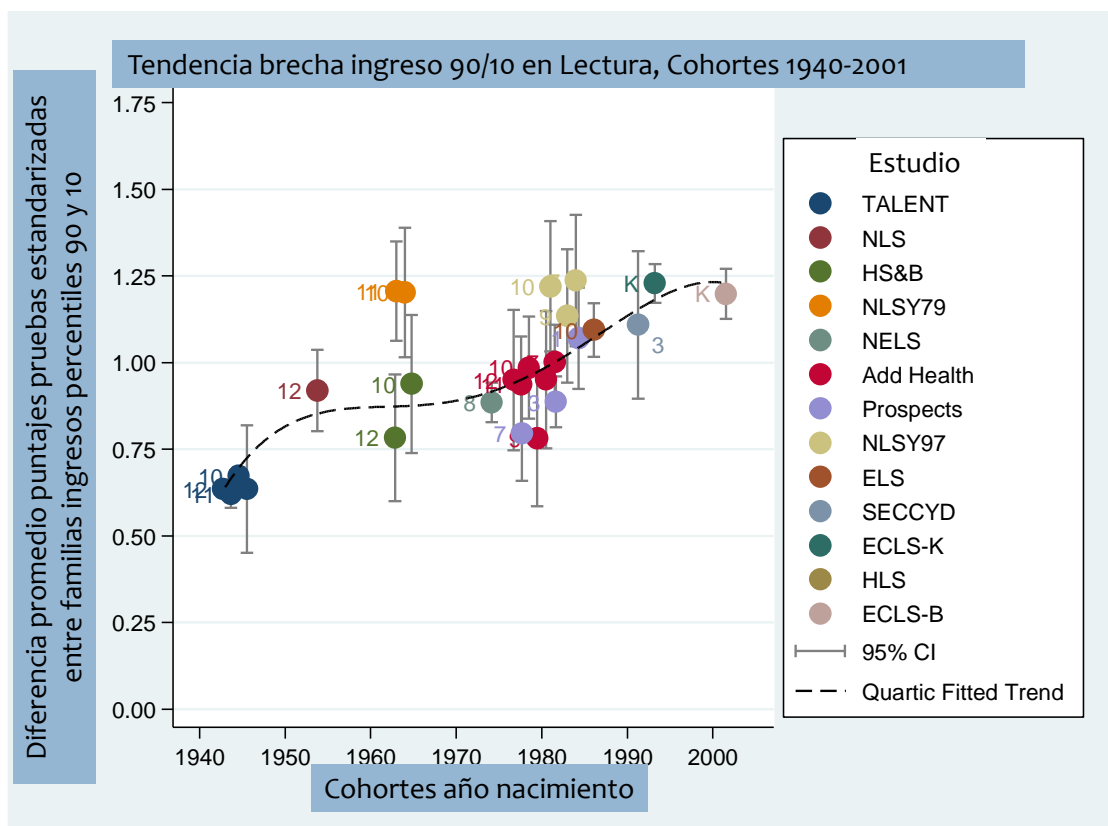
Para calcular las brechas de ingreso/logro requeridas para hacer el estudio, las soluciones del profesor Reardon a estas dificultades metodológicas fueron las siguientes: estandarizar resultados de pruebas dentro de cada estudio; usar datos categóricos de ingreso para estimar logro promedio de hijos en familias en los percentiles 90 y 10 de distribución del ingreso; ajustar por confiabilidad estimada del ingreso; ajustar por confiabilidad estimada de logro de la prueba; usar estudios longitudinales para valorar si/cómo las brechas varían por edad/grado; análisis de

⁵Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children, New York: Russell Sage Foundation Press, 2011.

⁶Los estudios son: Project TALENT, NLS72, HS&B, NLSY79, NELS, Add Health, Prospects, NLSY97, ELS, SECCYD, ECLS-K, HSLs, ECLS-B.

sensibilidad para valorar distintos diseños muestrales; y ponderar estimaciones con inverso de varianza de muestra.

Ilustración 6 – Brechas de ingreso/logro en Estados Unidos por cohorte de nacimiento de los estudiantes. Calculadas con 13 estudios de logro



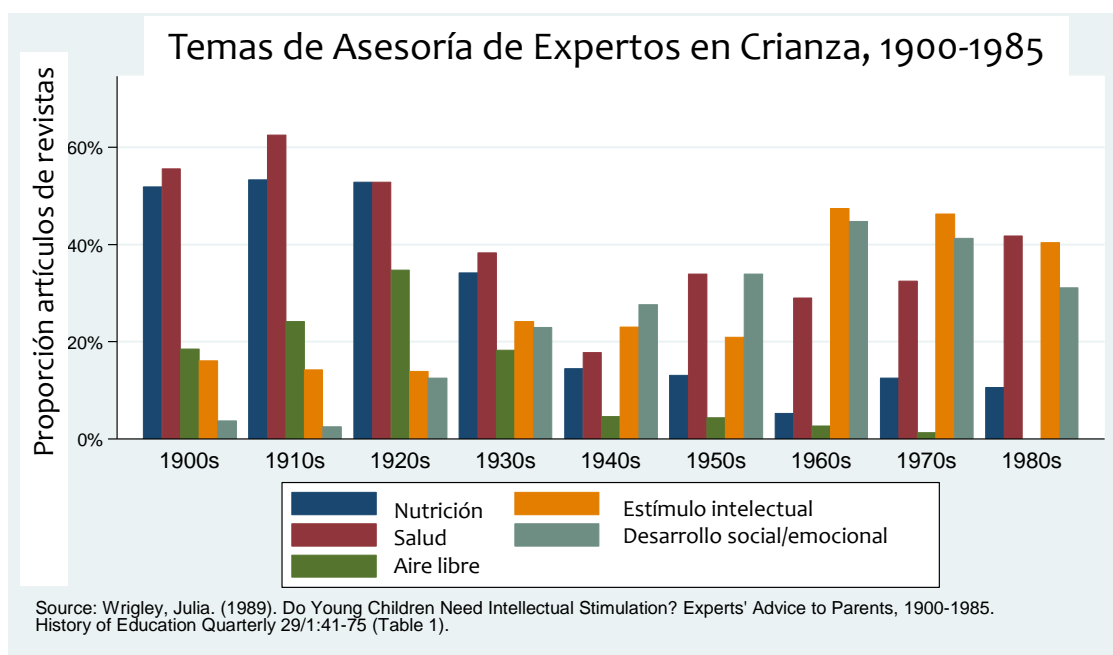
Para dimensionar qué tan grandes son las brechas encontradas, el profesor Reardon señaló que una desviación estándar es equivalente a la diferencia entre percentiles 31 y 69. Si la brecha es una desviación estándar, esto implica que el estudiante promedio en una familia del percentil 10 en la distribución del ingreso tiene un puntaje por debajo del 84% de estudiantes de familias en el percentil 90 de la distribución del ingreso. Una desviación estándar, además, es la cantidad típica que un estudiante aprende en un año en kínder y primero, 3 años en primaria y 6 años en secundaria.

Al procurar explicar por qué ha crecido la brecha ingreso/logro, el profesor Reardon exploró las siguientes opciones: ¿creciente inequidad del ingreso? ¿Cambios en los patrones de inversión de las familias? ¿Retornos al ingreso crecientes? ¿Inversiones en el desarrollo cognitivo de los hijos (entre familias de altos ingresos) crecientes? ¿Prácticas parentales cambiantes? ¿Correlación entre ingresos y otros recursos familiares creciente? ¿Segregación socioeconómica creciente? ¿Políticas sociales (red de protección social para los pobres) cambiantes?

En particular señaló que la visión de los padres sobre la crianza ha cambiado a lo largo del siglo XX. Hay un énfasis cada vez mayor en la importancia que le dan los padres al apoyo al desarrollo cognitivo de sus hijos. Las clases medias/altas educan a sus hijos bajo el llamado “crecimiento concertado”, mientras que la clase trabajadora considera el

“logro por crecimiento natural”. La **Ilustración 8** a continuación muestra la evolución en los temas de asesoría de expertos en crianza en las primeras ocho décadas del siglo XX. Como es posible apreciar, a lo largo del tiempo el tema de la nutrición tiende a perder protagonismo, mientras que la salud sigue teniendo un papel importante. Análogamente, el estímulo intelectual y el desarrollo social y emocional han cobrado mayor importancia, especialmente a partir de 1950.

Ilustración 7 – Visiones Cambiantes sobre la Crianza 1900-1985 (Wrigley, 1989)



El autor señaló que hay retornos progresivos a la educación universitaria y destrezas cognitivas. Igualmente, que el ingreso está más fuertemente asociado a la educación de los padres y a las destrezas cognitivas ahora que antes. Subrayó un emparejamiento selectivo entre parejas. Dijo que las familias de altos ingresos no solo tienen más ingresos sino también otros recursos importantes (padres duales, alto nivel educativo & destrezas cognitivas, hogares más pequeños, pocas madres muy jóvenes).

En resumen, las tendencias encontradas por el profesor Reardon en el periodo son las siguientes: (i) las brechas de ingreso/logro han crecido marcadamente en décadas recientes (desde cohortes de nacimiento 1970s o antes); (ii) las brechas de logro por ingreso han crecido más rápidamente en la mitad superior de la distribución del ingreso; (iii) las brechas de logro por ingreso ahora son más grandes que la brecha de logro entre blancos y negros; (iv) las brechas están presentes desde cuando los estudiantes comienzan la escuela; (v) las diferencias en inequidad, sumadas a una asociación estable entre ingreso y logro educativo, parecen insuficientes para explicar los patrones de asociación entre inequidad y brechas de ingreso y logro. Es último punto, explica Reardon, se puede explicar porque la asociación entre ingreso y logro también ha cambiado.

Sobre esto último, la hipótesis provisional del profesor es la siguiente: para trabajadores jóvenes, los retornos del grado universitario se han duplicado entre 1980-2000; la

creciente importancia de la educación en el mercado laboral y la movilidad económica han hecho que el éxito educativo sea aún más importante; esto ha modificado el comportamiento de inversión en la crianza, es decir, ha cambiado la forma como los padres piensan sobre sus hijos. También ha cambiado cómo pensamos sobre rol de las escuelas (mayor énfasis en el logro escolar, medido con pruebas de logro), lo que conduce a una competencia creciente por ventajas educativas. Tener dinero y otras formas de capital es una ventaja en esta competencia; el ingreso importa más que antes.

Las implicaciones planteadas por el autor sobre sus hallazgos son las siguientes: (i) el vínculo entre ingreso familiar y logro de los hijos, sumado a la importancia cada día mayor de las destrezas cognitivas en determinar los ingresos, producen un ciclo de retroalimentación que *conduce a baja movilidad socioeconómica y creciente inequidad*. (ii) Este ciclo de retroalimentación operaría parcialmente a través de la enseñanza, aunque las escuelas (en un sentido estrecho) no parecen ser la causa primera de esta tendencia. No es claro si las escuelas (por si solas) puedan reversar esta tendencia, aunque serían un mecanismo que puede ayudar.

En materia de políticas públicas, las implicaciones de sus hallazgos serían, a juicio del autor, las siguientes: (i) reducir la desigualdad económica; (ii) mayor inversión en primera infancia; (iii) programas focalizados son probablemente los más rentables (apoyo a familias de bajos ingresos, reparar/fortalecer la red de protección, programas para desarrollar destrezas de crianza, incrementar el foco de políticas educativas para estudiantes de comunidades y familias con bajos ingresos, desarrollar y probar estrategias para mejorar la enseñanza/aprendizaje de estudiantes de bajos ingresos).

Sean Reardon es Doctor en Educación de la Universidad de Harvard y Profesor de la Facultad de Educación de la Universidad de Stanford, donde dirige el programa de entrenamiento doctoral en análisis cuantitativo de políticas educativas. Su investigación se concentra actualmente en las causas y los patrones de las disparidades socioeconómicas en el logro, los efectos de las políticas de integración en los colegios sobre los patrones de segregación y los resultados educativos, y los efectos sociales y educativos de la inequidad en el ingreso.

Información en línea sobre el trabajo de Sean Reardon:
<http://ed.stanford.edu/faculty/seanreardon>

5. Conferencias alternas

En las sesiones de la tarde, los días 1 y 2 de Noviembre, se presentaron conferencias en las que se expusieron estudios que hacen uso de las pruebas SABER. La selección de estos estudios se realizó mediante una invitación pública a presentar trabajos en el seminario. A continuación se presenta un resumen de cada una de ellas:

Calidad, Igualdad y Equidad en la Educación Colombiana. Análisis de la prueba SABER 2009. Jesús Duarte, Soledad Bos y Martin Moreno. División de Educación del Banco Interamericano de Desarrollo. Washington D.C.

El objetivo de la investigación es describir la desigualdad en los aprendizajes de los alumnos colombianos. Para este fin, analiza la equidad en la distribución de los recursos

y procesos que ocurren en la escuela y su relación con los aprendizajes. Para efectos de este estudio, los autores diferencian entre igualdad y equidad. Definen *igualdad* como la distribución de los aprendizajes según el nivel socioeconómico (NSE) de los estudiantes, la zona geográfica donde residen y el sector de la escuela que atienden, sea público o privado. Así mismo, definen *equidad* como la distribución de los aprendizajes teniendo en cuenta como se distribuyen los recursos y procesos escolares según la condición socioeconómica de los estudiantes

Para estimar el NSE del estudiante construyen un índice basado en el cuestionario socio-demográfico de SABER 5 y 9 2009 muestral. Los ítems incluidos fueron: condición de la vivienda, hacinamiento y nivel educativo de los padres.

En su análisis, los autores emplean los conceptos de *riesgo relativo* y *riesgo atribuible* a la población, conceptos usados en epidemiología, y que los autores adaptan a la investigación en educación. En estos términos, su pregunta de investigación es por el riesgo de que los estudiantes no alcancen al menos el nivel satisfactorio en las pruebas.

Como medida de *riesgo relativo* emplean la vulnerabilidad (no lograr nivel satisfactorio en SABER) asociada con estar expuesto a un factor de riesgo (ser pobre, no tener suficiente acceso a un recurso, etc.) comparado con la vulnerabilidad de los que no están expuestos al riesgo. Y en cuanto al *riesgo atribuible a la población* emplean la proporción de la vulnerabilidad total que disminuye si la exposición al factor de riesgo desaparece. Estas medidas, dicen los autores, tienen en cuenta el tamaño de la proporción expuesta al riesgo.

El análisis muestra que en Colombia existen altas desigualdades en los resultados académicos de los estudiantes, asociadas al nivel socioeconómico de sus familias y al tipo de gestión y zona geográfica de las escuelas a las que asisten. Esta relación es más importante entre las escuelas que al interior de las mismas, denotando una alta segregación de las escuelas colombianas según el nivel socioeconómico de los alumnos.

En cuanto a la distribución de recursos y procesos escolares, encuentran que hay altas inequidades con clara desventaja para las escuelas que atienden a los alumnos más pobres y las escuelas públicas urbanas y rurales, que se asocia a su vez con mayor riesgo de sus alumnos de obtener resultados insatisfactorios en la prueba SABER. Los resultados de las estimaciones de los modelos multinivel, donde se analizan de manera conjunta las interacciones entre factores escolares y los resultados de las pruebas, indican que mejores condiciones físicas de las aulas, la conexión adecuada a servicios públicos, la jornada completa, la presencia de reglas en el aula, los menores niveles de violencia en las escuelas y mayor satisfacción docente están asociados de manera significativa con más altas probabilidades de que los estudiantes logren resultados adecuados en las pruebas. Por lo anterior, afirman que la distribución de los recursos y procesos en las escuelas que atienden a diferentes poblaciones tienen un efecto significativo sobre el riesgo de vulnerabilidad para el caso de las poblaciones pobres, las escuelas rurales y las escuelas públicas.

Con base en estos hallazgos, los autores señalan que mejorar la distribución de los recursos y procesos en las escuelas que atienden a la población pobre tendría un alto potencial de incrementar al mismo tiempo la calidad y la equidad de los aprendizajes escolares en Colombia. Finalmente, señalan que las variables de infraestructura y

jornada completa tienen el mayor potencial para disminuir la inequidad, dentro de las que evaluaron.

Escolaridad, Habilidades Cognitivas y Crecimiento Económico Departamental en Colombia. Camilo Leonardo Silva Díaz. Pontificia Universidad Javeriana.

El estudio busca validar la posibilidad de que en el caso colombiano el crecimiento económico no dependa del nivel inicial de crecimiento ni del grado de escolaridad, sino de las habilidades cognitivas de las personas. El propósito de este ejercicio es analizar los resultados obtenidos por Hanushek y Woessmann (2009) para el caso de Colombia.

El autor utiliza un modelo que busca explicar la tasa de crecimiento real del PIB per cápita departamental en función del capital humano, otros factores que afectan al crecimiento, y un término estocástico del error. La variable de capital humano depende a su vez de los años de escolaridad, la calidad de esa escolaridad, factores familiares, la salud, los compañeros, experiencia, etc.

El autor realiza distintas regresiones econométricas y concluye que no es posible encontrar evidencia de que exista una relación global entre las habilidades cognitivas reportadas en la prueba Saber 11 y el crecimiento económico departamental. Si bien existe una relación causal, ésta sólo se observa en las siguientes materias: matemáticas, física, biología e inglés. Por lo anterior, no se puede afirmar que las demás materias no se relacionan con el crecimiento económico departamental.

A través de sus resultados, el autor afirma que la escolaridad sólo está relacionada con el crecimiento económico al no incluirse la variable en la regresión. Por tanto, predice que los resultados entregados por Hanushek y Woessmann (2009) no son significativos al involucrar otras pruebas internacionales.

Relaciones entre programas y proyectos institucionales y el desempeño en pruebas SABER 5° y 9°. y SABER 11° en 12 instituciones públicas de Manizales. Ligia Inés García Castro. Universidad Autónoma de Manizales, Observatorio de la Calidad de la Educación (FUNDECA) y la Fundación Luker.

Este trabajo tiene por objeto estudiar la siguiente pregunta: ¿cuáles son las posibles relaciones entre programas y proyectos implementados en las instituciones educativas y los resultados obtenidos en pruebas Saber entre los años 2002 - 2010 en 14 instituciones educativas públicas de Manizales.

Para el efecto, la investigadora emplea una metodología mixta cuantitativa (análisis descriptivo de los resultados) y cualitativa (estudio de caso). En su análisis emplea las bases de datos del ICFES entre los años 2002 y 2010.

Concluye la investigadora que existen relaciones entre los propósitos de formación de los proyectos institucionales y las competencias que se evalúan en las pruebas Saber. Concluye también que los resultados en las Pruebas Saber dependen de múltiples factores, entre ellos las intervenciones que se hacen desde los proyectos institucionales.

Evaluación de la escritura en el examen de estado de evaluación de la calidad de la Educación Superior en Colombia: la experiencia del ICFES. Ernesto Cuchimaque; Carlos Ordóñez; Carlos Pardo. Instituto Colombiano para la Evaluación de la Educación – ICFES

Este trabajo se concentra en estudiar la prueba de escritura en el examen SABER PRO, que tiene como propósitos evaluar la calidad de la educación superior del país con estándares internacionales de calidad y ofrecer información que permita introducir ajustes en los programas de formación. Una de las competencias que se ha considerado pertinente evaluar es la de comunicarse por escrito.

La competencia se definió desde enfoques que permiten ver el proceso de producción escrita como una dinámica que involucra la definición de la intención comunicativa, la estructuración de las ideas y la selección de las formas idóneas de expresarlas. La prueba no explora el dominio de conocimientos específicos sino la capacidad para organizar distintos elementos relacionados con el tema que se propone desarrollar y a escoger la mejor forma de presentarlos ante determinado auditorio.

El modelo de análisis reportado por el autor es el modelo de Rasch de facetas. En este modelo se considera que *evaluado*, *tarea* (pregunta) y *calificador* son facetas independientes. El software empleado por el autor convierte valoraciones ordinales dadas por distintos calificadores a una misma tarea, en una escala continua, de tipo logístico, que permite calibrar la habilidad del evaluado, reportando además una serie de indicadores, como los de severidad/indulgencia, para valorar la calidad de las preguntas y de los calificadores (Linacre, 2012). A partir de este análisis se produce un resultado denominado justo, que tiene en cuenta el grado de severidad o indulgencia de los calificadores que le han correspondido a un evaluado en particular.

Como conclusión de su ejercicio, los autores reportan que evaluar la escritura es un proceso complejo, y que la prueba muestra que los procesos de escritura de los estudiantes evaluados son deficientes.

Métodos de equiparación en pruebas SABER 11° Colombia: Comparación de procedimientos en personas videntes e invidentes. Aura Nidia Herrera Rojas, Lady Catheryne Lancheros Florián, Gillen Javier Jiménez López, María del Pilar Soler Parra y Angela María Espinosa Garzón. Universidad Nacional.

La pregunta de investigación de este trabajo es la siguiente: ¿cuál método de equiparación (TCT o IRT) permitirá estimar la equivalencia de puntuaciones con menor error, en pruebas de aplicación masiva para población vidente e invidente?

Como antecedentes, los autores señalan que, en pruebas de logro de aplicación masiva, el desempeño de los invidentes es significativamente inferior. Señalan que se aplica la misma prueba a videntes e invidentes, aunque éstos operan por canales sensoriales diferentes (Rojas y cols., Inédito).

Los autores emplean en su ejercicio los datos de la aplicación del examen de estado SABER 11 en el segundo semestre del año 2008, diferenciada por población vidente e invidente (237.102 videntes y 87 invidentes). Emplean un diseño cuantitativo experimental. El instrumento empleado es la subprueba de Lenguaje SABER 11 aplicado en 2008 II, que tiene 36 ítems (de selección múltiple con única respuesta) los cuales conformaban dos formas de prueba cada una de 24 ítems.

Las variables dependientes del estudio son el error de equiparación para cada método y el coeficiente de variación. Las variables independientes son los métodos de equiparación: (Modelo TCT: Método lineal de Tucker y Método lineal de Levine), el Modelo TRI y los métodos Media/Sigma.

De su ejercicio, los autores concluyen que los procedimientos de equiparación son una herramienta viable para obtener puntuaciones intercambiables entre la población invidente y la vidente cuando se mide el mismo constructo. Estos procedimientos favorecen procesos de evaluación equitativos, al contribuir en interpretaciones validas sobre el nivel de habilidad de una población determinada. En particular, el método media/sigma aporta un menor error de equiparación en cada matriz simulada. Por su parte, el método lineal de Tucker presenta una adecuada estabilidad a lo largo de las réplicas. Señalan los autores que es necesario elaborar pruebas objetivas que permitan evaluar el constructo medido a lo largo de la dimensión de la habilidad con el fin de tener información en sus diferentes niveles. Terminan diciendo que estos estudios ponen de relevancia la importancia de generar mayor conocimiento sobre la calidad de la evaluación dirigida a población minoritaria.

Desempeño relativo de los graduados en el área de educación en el examen de Estado del ICFES. Juan D. Barón, Leonardo Bonilla, Mónica Ospina y Lina Cardona. Banco Mundial, Banco de la República, Universidad EAFIT.

Como motivación de este trabajo, los autores señalan que los estudiantes colombianos en educación básica y media muestran un desempeño bajo en pruebas internacionales (últimos lugares en los resultados de la prueba PISA-2009 y TIMSS-2007) y que la *calidad* del docente es un determinante del rendimiento académico de los estudiantes (Eide et. al., 2004; De Paola, 2000; Hanusheck, 1992; Rivkin, 2005). Mediante un análisis descriptivo el estudio investiga cómo se comparan los estudiantes y graduados en el área de *educación* con los estudiantes y graduados *en otras áreas*, en términos de sus resultados en la prueba Saber 11°. La hipótesis de los autores es quea mejor desempeño en el examen del SABER 11° (ICFES), menor es la probabilidad de estudiar y graduarse de educación.

Para probar su hipótesis los autores corren un modelo probabilístico con los siguientes datos: pruebas estandarizadas de Estado, SABER 11, presentadas entre el 2000 y 2007, y los registros del SPADIES (Individuos registrados en instituciones de educación superior acreditadas).

Los resultados reportados por los autores son los siguientes: los estudiantes de educación tienen un desempeño *más bajo* en el componente de matemáticas y de lenguaje de la prueba SABER 11° comparados con estudiantes en otras áreas. En el caso de las matemáticas, la diferencia en el desempeño entre graduados en educación y otras áreas en la prueba Saber 11 es *mayor* que la diferencia entre los que inician sus estudios en educación y en otras áreas. Igualmente, en lenguaje la diferencia en el desempeño en la prueba Saber 11 entre graduados en educación y otras áreas es *mayor* que la diferencia entre los que inician sus estudios en educación y en otras áreas.

Por otro lado, el estudio revela que la probabilidad de desertar en programas de educación es mayor en los estudiantes de más bajo desempeño. Además, a mejor desempeño en SABER 11, menor es la probabilidad de estudiar educación. Es decir que una persona con resultados bajos en SABER 11 (dentro de todos los estudiantes de educación superior) tiene una mayor probabilidad de estudiar y de graduarse en educación que en otras áreas de conocimiento. En el caso particular de las mujeres, existe una mayor probabilidad de estudiar educación que los hombres.

Para atraer a los mejores estudiantes a la docencia, los autores recomiendan mejorar el estatus de la profesión en términos de beneficios materiales y no materiales. Una medida podría ser el aumento de salarios, pues Mizala y Ñopo (2011) revelan que en América Latina los docentes ganan menos que el promedio en otras profesiones. Por otro lado, mencionan la creación de becas para estudios superiores en educación, cuyo monto depende de los resultados de la prueba estandarizada.

Diagnóstico del Valor Agregado en el nivel de lectura de los Egresados del Programa de Administración de Empresas de la Universidad Jorge Tadeo Lozano: Una propuesta desde el Benchmarking aplicando Data Envelopment Analysis – DEA. Federico Manuel Zevallos Peña. Universidad Jorge Tadeo Lozano.

El objetivo de esta investigación es construir indicadores para la estimación del valor agregado en lectura, adquirido por los Egresados de Administración de Empresas de la Universidad de Bogotá Jorge Tadeo Lozano (UJTL), mediante un análisis comparativo con otras universidades de Bogotá.

El estudio busca averiguar si los logros potenciales de los estudiantes de administración de la UJTL, en el componente ECAES-comprensión lectora son similares a los del grupo de estudiantes de otras universidades considerados para el estudio. Para este propósito, el autor emplea un modelo de Análisis Envolvente de Datos (conocido en inglés como *Data Envelopment Analysis*, DEA), el cual se basa en la metodología de estimación de la eficiencia técnica, en la que dado un nivel de insumos se estima la capacidad de obtener el mayor producto posible. En el modelo de DEA, las unidades de toma de decisiones (DMU) son los estudiantes de Administración de Empresas que presentaron las pruebas ECAES entre los años 2007 a 2009. Las variables de entrada del modelo, es decir, los insumos, son los puntajes Saber 11 en lenguaje. Las variables de salida son.

Para un estudiante dado, el modelo compara su logro real en el puntaje ECAES en comprensión lectora respecto a este mismo logro obtenido por otros estudiantes más eficientes (pares eficientes) que registran entradas similares, es decir, puntajes similares en la prueba Saber 11 en el componente de lenguaje. El autor calcula dos clases de eficiencia técnica para los estudiantes de Administración de la UJTL: (i) *productividad en su grupo* (ESG), respecto a los estudiantes de administración de la UJTL; y *productividad en todos los grupos* (ETG), respecto a todos los estudiantes de administración evaluados.

Los principales resultados reportados por el autor son los siguientes: al interior del grupo de estudiantes de Administración de la UJTL, la eficiencia técnica mínima en el ECAES es de 82.2% y la máxima, de 100%. El autor subraya que cinco (5) estudiantes se constituyeron en “pares eficientes” dentro de este grupo. El promedio de productividad de este grupo fue de 94.6% y la mediana de 96.0%. En cuanto a la productividad en todos los grupos, el autor estimó una eficiencia técnica mínima de 73.2%, y máxima de 89.8 %. El autor subraya que ningún estudiante de la UJTL alcanzó a formar parte de la frontera eficiente en la evaluación que incluyó a otras universidades, pues el 50% de los estudiantes evaluados de administración de la UJTL lograron resultados de salidas (puntajes de ECAES en el componente de lectura) entre el 73.2% y el 82.6% respecto a sus “pares eficientes” para un nivel similar de entradas (puntaje de lenguaje en el ICFES). Ello indica que existen logros potenciales significativos por cubrir para el 50% de los estudiantes. Según el autor, estos logros podrían alcanzarse replicando las mejores experiencias didácticas en el ejercicio de Capacidades de Lecto-escritura y prácticas de Alfabetización Académica al interior de las universidades evaluadas.

Benchmarking de universidades colombianas. Daniel Bogoya, Universidad Jorge Tadeo Lozano.

El objetivo del estudio es ofrecer indicadores de *benchmarking* entre las instituciones de educación superior colombianas, con base en los resultados alcanzados por los estudiantes que presentaron las pruebas SABER PRO en el año 2011. Se analizan los indicadores tanto para cada institución como para el país, con el fin de enriquecer la comparación.

La investigación se llevó a cabo bajo la hipótesis de que si bien el aprendizaje alcanzado en una institución de educación superior podría estar signado por las condiciones socioeconómicas y académicas que los estudiantes demuestran al inicio de sus estudios,

siempre será posible pensar en construir y desplegar estrategias pedagógicas para modificar sensiblemente esta tendencia y lograr que los estudiantes con mayores dificultades venzan las barreras propias del entorno y también puedan desarrollar altos niveles de comprensión.

La metodología empleada para alcanzar el *benchmarking* partió de la información disponible en las bases de datos de las pruebas SABER PRO en la versión 2011 y siguió con el diseño y estimación del valor de los indicadores. Se agruparon los estudiantes de acuerdo con la institución a la que pertenecen y el estrato socioeconómico suministrado por ellos mismos en el formulario de inscripción. Luego, se calculó el promedio de puntaje asignado en cada uno de los componentes de las pruebas genéricas. La disposición de los resultados de cada institución frente a los respectivos resultados del país en su conjunto permite hacer una comparación directa del impacto de cada proyecto educativo, para valorar fortalezas y debilidades relativas que conduzcan al diseño e implementación de programas focalizados y lograr mayor equidad en el aprendizaje de los estudiantes.

El estudio presenta indicadores que se pueden interpretar como una señal objetiva y cuantitativa que muestra el resultado de aprendizaje alcanzado por los estudiantes, luego de experimentar toda la complejidad de las diversas actividades inherentes a la vida académica y cultural de las instituciones donde cursaron sus estudios. En el segmento de resultados globales, se presenta también el promedio y desviación estándar (de los promedios hallados por componente), y el número de estudiantes evaluados, para el país y para cada institución con datos de al menos diez estudiantes. Por otro lado, se calcula un índice de rendimiento, el cual corresponde al promedio global de la institución sobre el respectivo promedio de la institución con resultados más destacados, multiplicado por cien, el estrato socioeconómico promedio y la condición de acreditación institucional.

En cuanto a los resultados por componente, la investigación calcula un promedio de puntaje obtenido por los estudiantes evaluados, para la institución y para el país, respectivamente. Los componentes considerados son: pensamiento crítico, entendimiento interpersonal, solución de problemas, inglés, comunicación escrita y comprensión lectora, en el primer semestre; y escritura, inglés, lectura crítica y razonamiento cuantitativo, en el segundo.

En el estudio se han resumido en un indicador global los resultados de todos los estudiantes que pertenecen a una institución, independiente del programa que han cursado. El indicador global por tanto es una señal integral de tendencia que funde toda la diversidad visible de una institución: resultados de estudiantes con desempeños destacados compensan los de los estudiantes con mayores dificultades y resultados de programas académicos con debilidades también son equilibrados con los de los programas más fuertes académicamente.

6. Estudios presentados en formato poster

Los participantes en el seminario en el auditorio de Bogotá pudieron observar en los pasillos resúmenes de estudios nacionales sobre calidad de educación, que emplean datos de pruebas estandarizadas de logro educativo. La siguiente es una reseña de estos estudios con sus respectivas preguntas de investigación.

- Julio César Alonso y Beatriz Gallo (2012) “*Determinantes de la excelencia en inglés de los bachilleres colombianos*”. Universidad ICESI. Preguntas de investigación: ¿Cuáles son las características del estudiante y de su entorno que están relacionadas con un desempeño “anormalmente” bueno? ¿Qué tan importante es reconocer la existencia de una estructura jerárquica al evaluar esta relación?
- Alberto Aragón Muriel, José Peñaranda y Germania Micolta (2012) “*Nuevas metodologías de enseñanza en química a nivel de educación media*”. Universidad del Valle, y Coordinación PNAT. Preguntas de investigación: ¿Influye la implementación de nuevas metodologías de enseñanza en química en los resultados de las pruebas SABER 11° realizadas por los participantes de la segunda y tercera cohorte del PNAT organizado en la ciudad de Santiago de Cali? ¿Existe un mayor interés en los jóvenes bachilleres por estudiar carreras que se relacionen con la química como ciencia básica al promover nuevas estrategias de aprendizaje en esta área?
- Juan Carlos Arana, (2012) “*Análisis espacial de los resultados obtenidos en las pruebas SABER 11° aplicadas por el ICFES en el departamento del Quindío (2010-II)*”, Especialización en Análisis Espacial. Departamento de Geografía, Universidad Nacional de Colombia. Pregunta de investigación: ¿Cuál es la correlación, que existe de acuerdo con la localización geográfica de los puntajes y qué factores asociados, pueden influir en el resultado?
- Cristian Bohórquez, (2012) “*Aplicabilidad de la lingüística textual y la psicolingüística en la composición de preguntas del Examen de Estado SABER 11*”, Grupo de investigación Estudios del Lenguaje y la Educación. Universidad Tecnológica de Pereira. Pregunta de investigación: ¿Pueden identificarse errores de composición en ítems de pregunta del ICFES usando los estudios del lenguaje?
- Marly Tatiana Celis, Óscar Andrés Jiménez y Juan Felipe Jaramillo (2012) “*¿Cuál es la brecha de la calidad educativa en Colombia en la educación media y en la superior?*” Maestría en Economía, Universidad de Manizales. Preguntas de investigación: ¿Cuáles son las diferencias en la calidad educativa, o brechas educativas, en el nivel de educación media y en educación superior en Colombia? ¿Cuál es la diferencia en la calidad de la educación que reciben los distintos grupos de la población? ¿Qué factores determinan la brecha educativa y qué características tienen los individuos en brecha educativa?
- Luisa Josefa Corsi, María Constanza García, Marcela Jiménez y Jeny Paola Niño (2012) “*Factores asociados a desempeños destacados y no destacados en las pruebas SABER*”, Universidad Javeriana, Gimnasio Santa María del Alcázar, Institución Educativa Simón Rodríguez, Instituto Tecnológico Los Andes. Pregunta de investigación: ¿Es posible identificar relaciones existentes entre factores asociados a desempeños destacados y no destacados de los estudiantes en la Pruebas Saber 11 2009?
- Martha Ligia Cuevas y Víctor Cervantes (2012) “*Diferencia de la Dificultad como procedimiento para detectar funcionamiento diferencial del ítem (DIF) en SABER 5 y 9*”, Subdirección de Estadística – Instituto Colombiano para la Evaluación de la Educación (ICFES). Pregunta de investigación: ¿Cómo funciona la diferencia

de la dificultad bajo diferentes condiciones teniendo en cuenta su potencia y tasas de error tipo I?

Enrique Chau y Manuela León (2012) “*Actitudes homofóbicas entre adolescentes: Prevalencia y factores asociados en 6 países de América Latina*”, Departamento de Psicología, Universidad de Los Andes. Preguntas de investigación: ¿Qué diferencias en actitudes homofóbicas existen entre adolescentes de 6 países de América Latina (Chile, Colombia, Guatemala, México, República Dominicana y Paraguay)? ¿Qué variables explican las actitudes homofóbicas?

Lorena Fontecha, Johana Gantiva y Yinzú Nairouz (2012) “*Percepción de los estudiantes frente a las prácticas de evaluación en el aula y su relación con el desempeño académico medido en las Pruebas SABER 9 2009*”. Fundación Cardioinfantil y Liceo de Cervantes. Pregunta de investigación: ¿Cuáles son las prácticas de evaluación en el aula desde la perspectiva del estudiante y qué relación tienen éstas con el desempeño académico de los estudiantes medido en la Prueba SABER 9?

Fabián Armando Gil, Viviana Alejandra Rodríguez, Martín Alonso Rondón, Carlos Gómez-Restrepo (2012) “*Variabilidad en los resultados por estudiante y universidad de las pruebas Saber Pro de Medicina*”, Departamento de Epidemiología Clínica y Bioestadística. Pontificia Universidad Javeriana. El objetivo del estudio es evaluar si existe algún tipo de asociación entre los resultados de la prueba SABER PRO, las características de las instituciones (Universidades/Facultades) y las características de los estudiantes de medicina.

Silvia Gómez Soler, (2012) “*Ingenuidad al Comparar Promedios: Cuantificando el desempeño de los estudiantes de Administración de Empresas de la Universidad Javeriana en el ECAES*”. Economista y MSc. Economía de la Universidad Javeriana, MA Desarrollo Internacional, de American University. Pregunta de investigación: ¿Los rankings de universidades realizados a partir de los promedios de los resultados del ECAES capturan realmente el desempeño de los estudiantes de la universidad bajo estudio?

Elizabeth León, Jonatán Gómez, Juan Pablo Neira y Eduar Moisés Castrillo (2012) “*Modelo de Datos para el Análisis Multidimensional y el Procesamiento Analítico en Línea de las Pruebas SABER: Caso particular Pruebas SABER PRO 2011*”, Departamento de Ingeniería de Sistemas – Facultad de Ingeniería, Universidad Nacional de Colombia. Pregunta de investigación: ¿Es posible aplicar técnicas computacionales de análisis a las pruebas SABER PRO?

Natalia Solano, (2012) “*Aproximación a la construcción de un Índice de Calidad de la Educación básica y media en Colombia*”, Universidad ICESI. Objetivo del estudio: construir un indicador que considere un conjunto de variables que proporcionen información funcional de análisis y evaluación de la calidad de la educación básica y media en Colombia.

7. Los talleres de capacitación sobre metodologías de investigación

A continuación se presenta la descripción de los talleres que se llevaron a cabo durante las jornadas de la tarde del seminario:

Título: Taller de Introducción a los Modelos Lineales Jerárquicos

Instructor: Andrés Martínez (Candidato a PhD), Universidad de Michigan.

Descripción: Los modelos lineales jerárquicos (MLJ), también conocidos como multinivel, mixtos o de efectos aleatorios, son modelos de regresión estadística que permiten estudiar asociaciones contexto-dependientes. Estos modelos proporcionan un marco conceptual y un conjunto flexible de herramientas de análisis para estudiar una variedad de procesos sociales y de desarrollo.

Por ejemplo, en educación, los MLJ pueden ser utilizados para estudiar la relación entre el desempeño académico y el acervo socio-económico de un estudiante cuando se presume que dicha asociación varía de un colegio a otro. Bajo el esquema clásico de regresión múltiple, se hace necesario especificar un modelo para cada colegio para permitir que las estimaciones de los parámetros varíen de un colegio a otro. Cuando el número de estudiantes por colegio es pequeño o el número de colegios es grande el esquema clásico es altamente inconveniente y técnicamente inadecuado. En cambio, los MLJ permiten un modelaje más apropiado y relativamente sencillo de los datos provenientes de éstas y otras estructuras sociales en las que existe una agrupación de las unidades bajo estudio (en este caso, de estudiantes en colegios).

El objetivo de este taller fue introducir los MLJ partiendo desde el modelo clásico de regresión múltiple. Se comenzó por aplicar el modelo clásico de regresión para estudiar la relación entre el desempeño en matemáticas y el bagaje social y económico de los estudiantes de un colegio. Luego se estudió cómo dicha asociación varía de un colegio a otro y cómo esta dependencia conduce naturalmente a la especificación de un MLJ. Luego se estudió la especificación e interpretación de resultados de un MLJ utilizando los datos de una de las aplicaciones recientes de las Pruebas SABER 11, que incluye resultados en la prueba de matemáticas así como información social y económica a nivel del estudiante. Finalmente, se presentaron ejemplos de otras aplicaciones y herramientas que permiten la aplicación de este tipo de modelos.

Prerrequisitos: Comprensión básica de inferencia estadística y habilidad para interpretar los resultados de una regresión múltiple.

Título: Taller sobre destrezas en la interpretación de pruebas estandarizadas de logro educativo

Instructores: Richard Shavelson, PhD (Universidad de Stanford), Edward Wiley (Universidad de Colorado en Boulder) María Figueroa, PhD (Universidad de los Andes) y Javier Toro (ICFES).

Descripción: El propósito del taller fue sensibilizar a los usuarios de grandes bases de datos, como las del ICFES, sobre las fortalezas y limitaciones de los puntajes de pruebas de logro para medir resultados educativos. Dicho de otra manera, se buscó que los participantes comprendieran las fortalezas y debilidades de los puntajes de las pruebas como proxy de aspectos del logro estudiantil que queremos conocer. El taller ofreció técnicas conceptuales, cuantitativas y estadísticas para juzgar la interpretación

de los puntajes. El taller procuró mostrar que aquellos que interpretan los puntajes frecuentemente prestan poca atención a qué tanto dichos puntajes significan las cosas que éstos intentan significar. El puente entre los dos puede ser fuerte o débil; cualquier debilidad plantea la necesidad de cautela en la interpretación de los puntajes de logro. El taller planteó varios pasos en el proceso de incrementar la fortaleza de la inferencia, desde los puntajes (productos) hacia lo que queremos conocer (resultados). Los pasos plantean supuestos básicos asociados con cada uno de los mismos, con especial atención a los compromisos que frecuentemente deben hacerse por razones de eficiencia del tiempo, costos y seguridad. Este proceso fue ejemplificado vinculando los estándares educativos nacionales de Colombia con las pruebas estandarizadas del ICFES y con los distintos ítems que se encuentran en las pruebas ICFES. Los investigadores y maestros deben usar no solamente puntajes en pruebas de logro sino también otros indicadores de resultados educativos (motivación, identidad, compromiso, selección de la carrera, cursos tomados, deserción), para valorar los resultados educativos.

Programa del taller:

- I. Introducción y visión general
- II. Estándares Nacionales de Educación de Colombia
 - a. Enumeración de resultados buscados
 - b. Discusión sobre los resultados que estarían incluidos en pruebas ICFES
 - c. Consideración de resultados que no están en las pruebas ICFES para la interpretación de los puntajes en pruebas
- III. Revisión del marco de una prueba ICFES
 - a. Contenido
 - b. Tipo de contenidos/competencias
 - c. Pensamiento/razonamiento crítico
 - d. Otros
 - e. Tipos de ítems de la prueba (resumen de áreas/contenidos medidos)
 - f. Discusión de implicaciones para interpretar los puntajes
- IV. Interpretación de puntajes en pruebas de logro
 - a. Confiabilidad de los puntajes en pruebas de logro
 - b. Validez de la información
 - c. Utilidad de la información para investigación y mejoramiento
- V. Discusión acerca de la inferencia, desde los puntajes en la prueba ICFES hacia los resultados en los estándares nacionales de educación.

Título: Taller sobre uso de bases de datos de evaluaciones estandarizadas

Instructor: Carolina Lopera y Álvaro Uzaheta - ICFES

Descripción: El taller, de carácter práctico, estaba dirigido a estudiantes e investigadores con conocimientos básicos en inferencia estadística. El objetivo principal es discutir las complejidades metodológicas, implicaciones en el análisis, procedimientos y métodos cuantitativos y estadísticos para el uso de bases de datos de las pruebas estandarizadas censales y muestrales que realiza el ICFES.

En particular, el taller mostró las diferencias en los objetivos, diseño metodológico, características de aplicación y generación de escalas de calificación en las evaluaciones

de SABER 11° y SABER 5° y 9°. Así mismo, se refirió a las implicaciones de las especificaciones propias de cada prueba en la implementación de técnicas estadísticas para realizar cálculos y análisis de puntajes promedios, diferencias significativas entre grupos de referencia (género, sector, nivel socioeconómico y tipo de establecimientos) y análisis de datos longitudinales.

Programa del taller:

Sesión 1

- I. Aspectos metodológicos y conceptuales de SABER 11o.
 - a. Objetivo de la evaluación
 - b. Características de la aplicación
 - c. Características del cálculo de resultados
 - Individuales
 - Instituciones
 - d. Uso de resultados longitudinales
- II. Usos y advertencias para el uso e interpretación de los resultados de SABER 11o.
- III. Ejercicios prácticos

Sesión 2

- I. Aspectos metodológicos y conceptuales de SABER 5o. y 9o.
 - a. Objetivo de la evaluación
 - b. Características de la aplicación
 - c. Características del cálculo de resultados
 - Calculo de Escalas
 - d. Diseño de bloques
 - e. Teoría de respuesta al ítem
 - f. Metodología de valores plausibles y errores de estimación
 - Determinación de niveles de desempeño

Título: Taller de introducción a la inferencia causal en investigación cuantitativa de educación y ciencias sociales

Instructor: Sean Reardon, PhD (Universidad de Stanford)

Descripción: Históricamente buena parte de la investigación educativa y de ciencias sociales no ha sido diseñada de una forma que les permita a los investigadores hacer inferencias causales creíbles sobre los efectos de políticas y prácticas. En parte, esto se debe a que numerosos estudios cuantitativos en educación y en ciencias sociales son esencialmente de correlaciones: pueden mostrar la presencia de asociaciones estadísticas entre políticas y prácticas con variables de resultados, pero no pueden proveer evidencia contundente de causalidad.

Establecer los efectos causales de políticas y prácticas sociales y educativas es uno de los objetivos de la investigación en educación y ciencias sociales. Los investigadores en educación, por ejemplo, pueden interesarse en el efecto que una determinada práctica de enseñanza tenga en el aprendizaje estudiantil, cuáles son los efectos de las políticas de rendición de cuentas en las prácticas de los maestros, o cuál es el efecto de los programas de educación de la primera infancia en la preparación escolar, y así sucesivamente. Los sociólogos pueden interesarse en conocer cuál es el efecto de determinadas condiciones de los barrios en el desarrollo infantil, o cuál es el efecto de las redes sociales en el comportamiento de los individuos. En estos casos, las decisiones sobre políticas y prácticas, incluyendo la asignación de recursos, deben depender de la

más creíble evidencia disponible sobre efectos causales de las políticas o prácticas en cuestión.

En décadas recientes, sin embargo, el llamado “contrafactual” o modelo de resultados potenciales (también llamado “Modelo Causal Rubin”) y desarrollos relacionados, ha cambiado dramáticamente la forma como los científicos sociales piensan sobre la causalidad. La más dramática influencia de este modelo ha ocurrido en economía; el modelo ha empezado a ganar recientemente influencia en el trabajo cuantitativo en los campos de la sociología, educación y la ciencia política. El nuevo esquema causal no es tanto un conjunto de modelos técnicos sino un marco lógico preciso para pensar acerca de la causalidad – y qué constituye evidencia de causalidad – en las ciencias sociales.

El taller, de carácter introductorio, estaba dirigido a estudiantes de postgrado con algún tipo de entrenamiento previo en investigación cuantitativa (por ejemplo, los participantes se sintieran cómodos con técnicas aplicadas de regresión). Se presentó en forma general una batería de métodos cuantitativos que permiten hacer inferencia causal en ausencia de un verdadero experimento aleatorio. Estos métodos incluyen 1) experimentos naturales, 2) variables instrumentales, 3) cuasi-experimentos, 4) métodos longitudinales, incluyendo métodos de series de tiempo interrumpidas y métodos de diferencia en diferencia, 5) regresión discontinua, 6) estimadores de aparejamiento, incluyendo *propensity score matching*, 7) estimadores de efectos fijos. Históricamente, estos métodos han sido muy pocas veces utilizados en la investigación educativa y en ciencias sociales distintas a la economía, pero ofrecen considerable poder a los investigadores interesados en generar evidencia convincente y creíble sobre efectos causales.

8. Los participantes y su evaluación sobre el seminario

En el seminario 2012 participaron un total de 483 personas, distribuidos en el auditorio de Bogotá (76%) y en 45 municipios adicionales, de todo el país, vía videoconferencia o *webstreaming*. En la mitad de los casos, los participantes fueron docentes de distintos niveles educativos.

En las evaluaciones sobre el seminario, los participantes fueron consultados si el evento les aportaba “nuevos elementos conceptuales y/o metodológicos útiles” para su desempeño académico y profesional, con lo cual coincidió el 90% de los participantes que diligenciaron la evaluación (**Ilustración 8**).

Ilustración 8 –Me aportónuevos elementos útiles para mi desempeño profesional

	2010	2011	2012
Totalmente de acuerdo	60%	76%	67%
Parcialmente de acuerdo	34%	19%	23%
Parcialmente en desacuerdo	2%	2%	3%
Totalmente en desacuerdo	4%	3%	7%

Al ser consultados sobre su interés por participar en siguientes versiones del seminario o en eventos similares organizados por el ICFES, el 92% de los participantes manifestó su interés en hacerlo (**Ilustración 9**).

Ilustración 9 – Me gustaría participar en siguientes seminarios

	2010	2011	2012
Totalmente de acuerdo	95%	93%	89%
Parcialmente de acuerdo	2%	5%	3%
Parcialmente en desacuerdo	0%	1%	0%
Totalmente en desacuerdo	3%	1%	8%

De la misma manera, el 90% de los participantes que diligenciaron el formato de evaluación del seminario están de acuerdo con que las temáticas tratadas en este seminario los motivan a hacer investigaciones o seguir investigando sobre temas relacionados con la calidad de la educación (**Ilustración 10**).

Ilustración 10 –Me motiva a investigar sobre calidad de educación

	2010	2011	2012
Totalmente de acuerdo	79%	79%	78%
Parcialmente de acuerdo	17%	18%	12%
Parcialmente en desacuerdo	1%	1%	1%
Totalmente en desacuerdo	3%	2%	8%

A su vez, el 94% de los participantes que diligenciaron la evaluación estuvieron de acuerdo con que “las temáticas de las conferencias presentadas son pertinentes con respecto a la discusión sobre calidad de la educación”.

Ilustración 11 –Las temáticas son pertinentes sobre calidad de educación

	2012
Totalmente de acuerdo	72%
Parcialmente de acuerdo	22%
Parcialmente en desacuerdo	2%
Totalmente en desacuerdo	4%

Al responder la pregunta abierta sobre los aspectos positivos del seminario, las respuestas predominantes subrayan la calidad de los expositores y de los temas de las conferencias. En cuanto a los temas por mejorar, se subrayan la importancia de la puntualidad en la agenda del evento, y la mejora del sonido y demás servicios tecnológicos.

Los posters con resúmenes de estudios que emplean datos de pruebas estandarizadas de logro se utilizaron en el seminario por primera vez en 2011, pero solo hasta 2012 fueron evaluados por los participantes. Para el 91% de los participantes los posters son una modalidad de presentación interesante, novedosa y útil (**Ilustración 12**). Para el 90% de los participantes los posters contenía información clara y suficiente (ibíd). Para el 92%

de los participantes los posters dan una idea del tipo de investigaciones que se pueden hacer con las bases de datos del ICFES (ibíd).

Ilustración 12 – Percepción sobre los posters

	Novedoso y útil	Info clara y suficiente	Muestra ejemplos
Totalmente de acuerdo	64%	49%	72%
Parcialmente de acuerdo	27%	41%	20%
Parcialmente en desacuerdo	4%	6%	3%
Totalmente en desacuerdo	4%	4%	6%

9. Conclusiones

El seminario en su versión 2012 tuvo como característica central la presentación de trabajos dirigidos a compartir con la comunidad participante las posibilidades y limitaciones de los datos de pruebas estandarizadas de logro educativo, de tal forma que sus usuarios tengan un mejor conocimiento sobre las formas apropiadas de utilizar este tipo de datos.

En efecto, temas como el muestreo de estudiantes y contenidos, la inflación de puntajes y los usos inadecuados de las pruebas cuando forman parte de sistemas de rendición de cuentas de alto impacto fueron ampliamente discutidos durante el Seminario. De esta manera, el ICFES buscó que, en su política de masificación del uso de esta información, ese uso sea creciente pero responsable, en todos los dominios, se trate de estudiantes, maestros, padres de familia, autoridades del sector educativo o investigadores, en los distintos propósitos que todos estos actores utilizan esta información. Los participantes tienen ahora mejores elementos de juicio para interpretar apropiadamente lo que dicen y lo que no dicen las pruebas de logro.

Durante el seminario se profundizó, además, en la metodología de valor agregado, priorizada por el ICFES para tratar de dar respuesta a la medición de la calidad de la educación de las instituciones educativas, en una forma justa, por oposición a la simple publicación de resultados de los puntajes, que equivocadamente le atribuye a las instituciones aciertos o errores que no son tales, sino simples diferencias en condiciones socioeconómicas de sus estudiantes.

En esta línea de investigación, los estudios presentados mostraron su potencial para proveer medidas acertadas de la contribución específica que hacen las instituciones educativas a sus estudiantes, observable en las pruebas. Pero igualmente mostraron las altas exigencias metodológicas que requieren los estudios de valor agregado, y la sensibilidad de sus resultados a distintos diseños metodológicos. El caso chileno, en particular, presentado en el seminario, ofrece además la posibilidad de ver los desafíos de emplear este tipo de metodologías, como parte de un sistema de rendición de cuentas del sector educativo basado en los datos de las pruebas de logro.

La suma de posibilidades y limitaciones de las pruebas de logro como de los métodos de valor agregado hicieron a la vez un caso en favor de sistemas de calidad y rendición de cuentas que hagan empleo de distintas fuentes de información. El diagnóstico del

paciente, dijo en una metáfora la directora del Instituto, requiere distintos exámenes sobre su salud, pues mal se haría en diagnosticar con base en uno solo.

En materia de líneas de investigación, el conjunto de conferencias principales presentadas en el seminario en los tres años de operación del mismo cobijan un amplio conjunto de temas de investigación básica a investigación aplicada, con heterogeneidad disciplinar y metodológica.

Desde un punto de vista institucional, la realización por tercer año consecutivo del seminario, con un creciente aumento en calidad de los trabajos presentados y en la jerarquía académica de los conferencistas, va en la dirección correcta para consolidar al Seminario como un referente natural de la comunidad académica del país.