



**CHRISTIAN CANCINO D.**  
*Doctor en Economía y Empresas, Universidad Autónoma de Madrid. Ingeniero Comercial, Universidad de Chile. Ingeniero en Información y Control de Gestión, Universidad de Chile.*



**ANTONIO FARIÁS L.**  
*Ph.D. in Management - IESE Business School. Ingeniero en Información y Control de Gestión, Universidad de Chile.*

## Instrumentos de Medición de la Calidad Educativa en Chile: COMENTARIOS DESDE EL CONTROL DE GESTIÓN

El conjunto de movilizaciones efectuadas a nivel nacional por estudiantes universitarios y secundarios han puesto al sistema educativo chileno en el centro de la discusión. Un informe de la Organización para la Cooperación y el Desarrollo Económico (OCDE) respecto de la educación superior en Chile señala múltiples problemas que debe enfrentar nuestro sistema, entre ellos la inequidad social en el acceso a la educación superior, prácticas de financiamiento obsoletas, instituciones que ya no se justifican y carreras largas e inflexibles (OCDE, 2009).

Se ha instalado en nuestro país la percepción de que el principal problema de la educación es su deficiente calidad. Este tema es motivo de debate y controversia en Chile y el mundo, entre otras razones porque el concepto de calidad es polisémico, tiene múltiples sentidos y se define de acuerdo con los fines de la sociedad y el rol que se le asigna a la educación dentro de ella (Colegio de profesores, 2006).

Dado que es necesario evaluar la calidad de la educación que reciben los alumnos, se utilizan pruebas estandarizadas que intentan medir lo aprendido. Sin embargo, se acusa que estas evaluaciones no logran captar la verdadera capacidad o aprendizaje de los estudiantes.

En la actualidad, son dos las evaluaciones más relevantes a las que son sometidos los estudiantes chilenos. El primero es el sistema de medición de la calidad de la educación, SIMCE, que evalúa la calidad educacional de los alumnos en determinados cursos de la enseñanza básica y media. La segunda evaluación es la denominada prueba de selección universitaria, PSU, que es una batería de pruebas cuyo fin es seleccionar a los alumnos que ingresan a las distintas universidades del país.

## DESCRIPCIÓN DE LOS INSTRUMENTOS SIMCE Y PSU

Explicaremos brevemente la naturaleza de ambos indicadores antes de explicar desde una visión de control el por qué sí podría resultar adecuada su utilización.

El SIMCE es el Sistema Nacional de Evaluación de resultados de aprendizaje del Ministerio de Educación de Chile. Su propósito principal es contribuir al mejoramiento de la calidad y equidad de la educación, informando sobre el desempeño de los estudiantes en diferentes subsectores del currículum nacional, y relacionándolos con el contexto escolar y social en el que ellos aprenden.

Las pruebas SIMCE evalúan el logro de los Objetivos Fundamentales y Contenidos Mínimos Obligatorios (OF-CMO) del Marco Curricular vigente en diferentes subsectores de aprendizaje, a través de una medición que se aplica a nivel nacional una vez al año a los estudiantes que cursan un determinado nivel educacional. Hasta el año 2005, la aplicación de las pruebas se alternó entre 4° Básico, 8° Básico y 2° Medio. A partir del año 2006, se evalúa todos los años a 4° Básico y se alternan 8° Básico y 2° Medio. Desde el año 2010 se incorpora la evaluación en 3° Medio del subsector Inglés.

Además de las pruebas asociadas al currículum, el SIMCE también recoge información sobre docentes, estudiantes y padres y apoderados a través de cuestionarios de contexto. Esta información se utiliza para contextualizar y analizar los resultados de los estudiantes en las pruebas SIMCE.

A continuación, explicaremos detalles del segundo indicador que se discute en este artículo, la PSU.

Desde el año 2003 las Universidades que componen el Consejo de Rectores, emplean una nueva batería de pruebas para seleccionar a los alumnos que ingresan a sus carreras. Ella está compuesta por cuatro pruebas llamadas Pruebas de Selección Universitaria, PSU, las que son desarrolladas, administradas, aplicadas y reportadas por el Departamento de Evaluación, Medición y Registro Educacional de la Universidad de Chile (DEMRE).

Las PSU son instrumentos de evaluación educacional que miden la capacidad de razonamiento de los postulantes egresados de la Enseñanza Media, teniendo como medio, los contenidos del Plan de Formación General de Lenguaje y Comunicación, de Matemática, de Historia y Ciencias Sociales y de Ciencias. Esta última incluye a Biología, Física y Química.

Los candidatos deben rendir en forma obligatoria Lenguaje y Comunicación y Matemática, y elegir entre Historia y Ciencias

Sociales y Ciencias. Sin embargo, si lo estiman, pueden optar por las cuatro pruebas.

## CRÍTICAS A LOS INSTRUMENTOS DE MEDICIÓN

Es frecuente escuchar en distintas discusiones argumentos que apuntan a señalar al SIMCE y a la PSU como malos instrumentos de medición. Las críticas más frecuentes dicen relación con que el puntaje obtenido en estas evaluaciones se correlaciona en forma directa con el nivel de escolaridad de los padres del alumno. A mayor nivel de escolaridad de los padres, mayor es el puntaje obtenido por el postulante. Del mismo modo, se explica que existe una fuerte relación de los puntajes con el ingreso bruto familiar. A mayor ingreso, mejor es el resultado en el SIMCE y en la PSU.

Adicionalmente a la correlación directa existente entre escolaridad, ingreso de los padres y puntaje SIMCE, la literatura encuentra fuertes diferencias entre los puntajes obtenidos por los estudiantes pertenecientes a la educación particular pagada versus los pertenecientes a la educación municipal (Sapelli & Vial, 2002; Chumacero & y Paredes, 2008; Paredes y Pinto, 2009). En Lenguaje se llegaría a una diferencia de 132 puntos en promedio y de 145 puntos en Matemáticas, lo que es bastante significativo considerando que una escuela en promedio obtiene 250 puntos. La brecha se reduce al comparar educación particular subvencionada con municipal llegando a los 30 puntos promedio.

La literatura también señala que la utilización de pruebas estandarizadas, tipo SIMCE, trae consecuencias negativas (Eyzaguirre & Fontaine, 1999):

- La enseñanza se puede reducir, enfocándose sólo a lo que se puede medir con facilidad y dejando de lado aspectos relevantes de más difícil evaluación (Williams, Blythe, White, Li, Stenberg & Gardner, 1996).
- También se puede limitar la enseñanza sólo a lo que se sabe va a ser evaluado. Para entender este efecto es necesario recordar que las pruebas estandarizadas por lo general son sólo una muestra, una selección de todo el contenido que puede abarcar una asignatura o un campo de conocimientos. Se ha detectado que los profesores tienden frecuentemente a limitar la enseñanza a aquellos temas dentro de una asignatura que se sabe serán evaluados, disminuyendo así, de hecho, las oportunidades de aprendizaje de sus alumnos.
- Puede también darse una ejercitación excesiva del test mismo en desmedro de la enseñanza de la materia. Por ejemplo,

cuando se asigna tiempo de clases para que los alumnos hagan ensayos con facsímiles. Smith (1991) constató que el tiempo destinado por los alumnos de EE.UU. a la preparación y ejecución de pruebas estandarizadas tomaba un tiempo promedio de 100 horas de las 1.000 horas lectivas anuales. Esta proporción es más significativa si se considera que el tiempo efectivo de enseñanza, para ese mismo período, es sólo de 300 horas en promedio (Weinstein & Mignano, 1993). El entrenamiento en la lógica de las pruebas puede conducir a aumentos de puntajes que no se condicen con lo realmente aprendido.

- Otra vertiente de crítica se refiere a que las pruebas estandarizadas tienen un sesgo cultural y utilizan un vocabulario propio del grupo social dominante y, por ello, discriminan negativamente a los niños de niveles socioeconómicos bajos y con carencias culturales. Esta crítica sería válida solamente cuando se quiere condicionar los incentivos económicos a los diferentes establecimientos según su desempeño en una prueba, o cuando los resultados de las pruebas se traducen en un cierre de oportunidades educacionales para los alumnos con resultados insatisfactorios.
- Finalmente, hay quienes sostienen que un sistema de evaluación externa crea en los colegios un clima de competencia y ansiedad contrapuesto al placer de aprender por aprender e incluso a un desarrollo emocional sano.

Por su parte, y con foco en la selección que realizan las universidades para sus alumnos, vemos que existe una serie de críticas a la Prueba de Selección Universitaria como instrumento de medición. Primero, se evidencia que, salvo contadas excepciones, a los estudiantes de mayores recursos económicos les va mejor que a los de menores recursos, brecha que continúa agrandándose año tras año. Es decir, la correlación entre ingresos familiares y resultados PSU es muy notoria en Chile. Segundo, muchos señalan que el instrumento PSU es poco confiable, pues el proceso de selección de los estudiantes a distintos tipos de carreras, con necesidades formativas distintivas, no aplica adecuadamente. La PSU podría estar siendo un instrumento adecuado de selección sólo para algunas carreras.

Es importante recordar que cuando se aplica un instrumento de medición educacional como las PSU a un grupo de estudiantes, se tiene el propósito de obtener información de cada uno de ellos, la que será empleada para tomar decisiones que lleven a ordenar a los postulantes desde el que tiene el más alto rendimiento hasta el de menor desempeño, y a partir de esta ordenación, seleccionar a los que tienen la mayor probabilidad de éxito académico.

En consecuencia, es de suma importancia que los resultados obtenidos representen razonablemente el rasgo que se está midiendo, lo que permitirá afirmar que las PSU empleadas son altamente confiables, y, por ende, el ordenamiento realizado asegura que la selección que se hará a partir de él, está cumpliendo con el supuesto: “A mayor puntaje de selección, mayor probabilidad de éxito en el primer nivel universitario”. La confiabilidad, apunta a descubrir qué parte de la variabilidad de los puntajes es producto de las diferencias individuales entre los sujetos en la variable medida y qué parte es atribuible a errores aleatorios o al azar. En efecto, si en dos oportunidades a un grupo de individuos se les somete a la aplicación de un instrumento que mide cierto rasgo, nunca se obtienen resultados que se dupliquen exactamente, siempre se apreciarán diferencias. Sin embargo, las mediciones reiteradas, que se realizan en las mismas condiciones, presentarán cierta consistencia y esta propensión de las mediciones sucesivas a ser consistentes es lo que se denomina confiabilidad.

Las cuatro pruebas de selección universitaria, tanto a nivel de grupo completo como por cada uno de los subconjuntos determinados por las variables Región, Género, Dependencia y Año de Egreso, presentan coeficientes de confiabilidad <sup>™</sup> iguales o superiores a 0,91, valor que, de acuerdo con estándares internacionales y nacionales, es considerado muy satisfactorio, por cuanto indica que el 91% o más de la varianza de los puntajes es producto de las diferencias individuales en rendimiento que presentan los postulantes.

Este resultado asegura que la ordenación que se hace a partir de los puntajes es muy precisa, lo que, a su vez, avala la calidad de la selección que se lleva a cabo, que es el fin último del proceso.

## VISIÓN DE CONTROL PARA LOS INSTRUMENTOS SIMCE Y PSU

A pesar que las críticas a los instrumentos pueden ser válidas, en este artículo esbozamos un análisis de las pruebas SIMCE y PSU desde la perspectiva exclusiva del control de gestión. El objetivo es agregar esta mirada al debate de la educación en Chile, de tal modo que la perspectiva que ofrecen los enfoques de control también pueda ser ponderada a la hora de evaluar estos instrumentos de medición. Bajo este punto de vista observamos que el SIMCE en sí, no constituye un mal indicador. ¿Por qué lo anterior?

La gestión basada en indicadores plantea varias dificultades, entre ellas, que los indicadores no siempre se utilizan para responder a su problemática u objetivo, es decir se ocupa un indicador como respuesta a un problema distinto para el que fue diseñado. En el caso del indicador SIMCE, vemos que éste

podría ser un instrumento adecuado para informar a los padres y apoderados hacia dónde, qué colegio, enviar a sus hijos, de acuerdo con la calidad promedio que se observa en aquellos alumnos que rinden el examen de cada colegio. El problema nace, cuando los apoderados no ocupan este indicador para tomar decisiones, y sólo lo ocupan para presionar a los colegios a que aumente el puntaje promedio a obtener en el SIMCE. La iniciativa esperada con el SIMCE es que los padres envíen a sus hijos a otros colegios con mayor puntaje, no que presionen por aumentos de puntajes al mismo colegio donde ya estudian sus hijos. Si padres y apoderados, bajo las características del sistema educacional chileno orientado a la demanda, no toman decisiones adecuadas con el instrumento de medición que ofrece el Estado, éste último podría estar midiendo una problemática distinta de la que intenta resolver.

En segundo lugar, los indicadores son proxis, es decir, variables aproximadas, las cuales no explican la totalidad de una realidad. Nadie pretende explicar la rentabilidad real de una empresa o proyecto cuando se ocupan indicadores como ROI, ROA o ROE. Estos últimos son valores aproximados que intentan dar información parcial sobre la tendencia o proyección de una situación futura. Entonces ¿por qué se recomienda utilizar valores aproximados y no una visión completa de la realizada bajo estudio?

Existe un problema de factibilidad que es necesario entender cada vez que se quiere estudiar de forma periódica un problema. Claramente, seleccionar a 300 mil estudiantes por un único set de pruebas de selección universitaria podría ser un indicador incompleto de la real selección que debería hacerse año a año en las universidades. Seguramente, también debería ser posible realizar otras pruebas complementarias: como entrevistas personales, focus group, entrevistas y evaluaciones en grupos o equipos, pruebas individuales por tipo de carreras, de acuerdo con sus requerimientos y características, etc. Es decir, para explicar la problemática completa de selección universitaria no sólo deberíamos utilizar la PSU, sino también un conjunto mayor de pruebas y test complementarios como los descritos. Todo lo anterior, de seguro nos llevaría a mejorar el proceso de selección. Entonces ¿Cuál es el problema? La factibilidad.

Bajo la visión de control de gestión la elección de indicadores debe realizarse bajo una mirada que responda a una factibilidad técnica, económica y organizacional. En la primera, factibilidad técnica, debemos preguntarnos si existen o no las capacidades, tecnología e infraestructura, para poder dar anualmente un conjunto de pruebas de selección que permitan que sobre 300 mil alumnos realicen la serie de evaluaciones - PSU, focus group, entrevistas personales, en equipos, prácticas, etc. - y logremos seleccionar a aquellos que ingresarán a la universidad y a sus respectivas carreras. Evidentemente, resulta al menos

cuestionable pensar que en el país exista la infraestructura suficiente para realizar lo anterior de forma periódica. Se tendría que solicitar semanalmente a los colegios públicos horarios protegidos para dar todos los tipos de pruebas de selección, no sólo en diciembre como es hasta ahora, sino en todo momento del año. Asimismo, se deberían crear grupos de trabajos de profesionales que estén todo el año monitoreando y realizando los exámenes.

Respecto de la factibilidad económica, es posible ver que para realizar el conjunto de pruebas de selección se requiere que el Estado se haga cargo, al menos en parte, de financiar la infraestructura y tecnología que dará soporte a todo el sistema. Este financiamiento, el cual podría ser incluso compartido con aquellos alumnos que dan las pruebas, pudiese ser demasiado caro y difícil de mantener en el tiempo. Ya no es sólo financiar la PSU, sino el conjunto complementario de pruebas de selección propuestos.

Finalmente, una evaluación de la factibilidad organizacional debe hacerse, entendiendo que es probable que no resulte fácil establecer los criterios particulares que diferencian tipos de carreras ni perfiles para los alumnos. La sociedad y las instituciones se han acostumbrado a las reglas del juego planteadas en el proceso de selección universitaria bajo PSU y podría resultar difícil implementar cambios. Los temas culturales pasan a tomar una importancia tremenda al respecto.

Al analizar el instrumento de medición PSU bajo una perspectiva del control de gestión, advertimos que la selección de estudiantes tiene una confiabilidad satisfactoria, incluso para estándares internacionales, adicionalmente el costo de tomar esta prueba no resulta elevado. Consideramos que en este caso el problema no es el instrumento con el que se mide a los postulantes. La PSU podría ser un buen indicador para ayudar al proceso de selección de algunas carreras, como por ejemplo aquellas del área negocios o de las ingenierías. Sin embargo, no necesariamente el indicador resulta igualmente efectivo en otras áreas, como en el caso de las Artes u otras carreras donde será necesario complementar con distintos tipos de exámenes. En este caso el indicador no es el malo, sino que aplicaría sólo para algunos tipos de carreras. Otro factor positivo de la PSU es que si bien, no capta la totalidad de la real aptitud de un alumno para ingresar a la universidad, es una buena proxi. La prueba de selección universitaria presenta una alta factibilidad económica, su costo es relativamente bajo en comparación a otros métodos de evaluación de alumnos, adicionalmente no impacta negativamente, ni demanda alto tiempo para su rendición.

Respecto del SIMCE y las críticas como indicador, vemos que pocos se atreverían a descartarlo como un medio para evaluar la educación en Chile. Si bien el SIMCE no explica la totalidad

del aprendizaje escolar, constituye una buena proxy, tiene una alta factibilidad económica, organizacional y técnica, es una medición que permite abarcar un alto número de estudiantes a un costo razonable.

Los efectos perversos de evaluar masivamente con pruebas objetivas y estandarizadas como el SIMCE, pueden disminuirse con un cuidadoso diseño del test, con una adecuada supervisión de clases y con políticas educacionales sensatas. Los inconvenientes de la aplicación masiva de evaluaciones se potencian cuando se



*La sociedad y las instituciones se han acostumbrado a las reglas del juego planteadas en el proceso de selección universitaria bajo PSU y podría resultar difícil implementar cambios. Los temas culturales pasan a tomar una importancia tremenda al respecto.*

asigna a estas evaluaciones un papel que no les corresponde. Por ejemplo, cuando ellas se transforman en el único instrumento para tomar decisiones críticas acerca de financiamiento, promoción de profesores o para determinar el futuro académico de los alumnos. En ese sentido, proponemos el uso de una malla de indicadores que permitan evaluar diversos resultados del proceso educacional desde diferentes perspectivas, lo que en el caso de la educación, es especialmente pertinente.

Cuando no están claramente establecidas las metas educacionales de cada etapa, las pruebas pueden llegar a ser la única herramienta para determinar el currículo escolar e incluso el contenido de los libros de texto. Esto da a las pruebas un peso

desmesurado en el quehacer educacional. Por último, cabe decir que las pruebas de evaluación alcanzan su máximo potencial de utilidad sólo cuando forman parte de un sistema coherente con metas educacionales claras y bien especificadas, donde hay una adecuada y transparente comunicación acerca de los objetivos de la medición, las características del instrumento y los resultados obtenidos, y complementación con otros métodos para asegurar la calidad del sistema (Eyzaguirre & Fontaine, 1999).

## COMENTARIOS FINALES

La economía chilena en los últimos años ha sido caracterizada como de rápido crecimiento y desigual distribución del ingreso. En este contexto, varias investigaciones muestran que la distribución de los ingresos en Chile es una de las más desiguales en el mundo, pero que dicha desigualdad se ha mantenido relativamente estable desde una perspectiva de largo plazo (Contreras, 1999). En efecto, en nuestro país el 20% más rico de la población recibe

17 veces más ingresos que el 20% más pobre. Por contraste, en los Estados Unidos esta misma relación alcanza a 8.9 veces; y en Perú y Corea del Sur estas relaciones alcanzan a 10.5 y 5.7 veces, respectivamente (PNUD, 1995). Estas desigualdades en los ingresos, producen una notable concentración de los estudiantes de más altos ingresos en la educación superior, principalmente en las universidades, aún cuando ya es sabido que los talentos están repartidos democráticamente en el sistema escolar. Efectivamente, cuestionarse la efectividad del sistema actual resulta ser una necesidad imperiosa al momento de hablar de la educación como un derecho y no el privilegio de quienes puedan pagarla.

Contrario a la creencia de que el contraste en los puntajes PSU se debe al tipo de establecimiento educacional del que provienen los estudiantes, recientes estudios señalan que la gran diferencia en puntajes no se debe tanto al tipo de colegio del que proviene un alumno (privado, municipal o particular), sino al ingreso del grupo familiar. Mientras que en los grupos con ingresos familiares de \$1,5 millones o más el puntaje promedio asciende a sobre 640 puntos, cuando los ingresos no superan los \$144 mil mensuales el promedio en la prueba de Matemática es ligeramente sobre 450 puntos, en una escala donde el mínimo puntaje es 150 puntos y el máximo es de 850 puntos.

Una vez que entendemos el problema de la desigualdad en los ingresos familiares de los alumnos, advertimos que la PSU no es la culpable de las brechas de inequidad social y educacional. Simplemente las refleja año tras año. El problema no está en criticar el termómetro, el indicador, o los criterios que las universidades utilizan para seleccionar sus postulantes. Lo que es necesario, de una vez por todas, es implementar medidas drásticas que mejoren la calidad de la educación en todos los niveles, y que se compense decididamente los mayores costos y dificultades de educar a los que tienen menos recursos, para evitar seguir con las serias deficiencias observadas en el ámbito educacional.

En conclusión, sostenemos que los instrumentos de medición utilizados para evaluar la calidad educacional de nuestros niños y jóvenes, no son culpables de los malos resultados obtenidos. El SIMCE y la PSU son sólo el termómetro con el que evaluamos la educación en Chile, si bien, como todo indicador, son perfectibles, no nos parece correcto criticarlos sólo porque nos comunican malas noticias. Reconocemos que la educación nacional presenta problemas de calidad, de inequidad en el acceso a la educación superior, y de altos costos, sin embargo estas deficiencias no son producto de los indicadores utilizados, sino son consecuencia del proceso educacional que se lleva a cabo en nuestras escuelas, colegios, liceos, centros de formación técnica, institutos profesionales y universidades, son producto de una muy desigual distribución de los ingresos, de un sistema educacional que segrega a sus alumnos y de las prácticas

educacionales que hemos utilizado hasta ahora. La solución no pasa por cambiar los instrumentos de medición, sino por implementar políticas que apunten a una mejor calidad en la educación y a reducir la brecha entre los individuos de mayores y menores ingresos. [E&A](#)

## REFERENCIAS

Chumacero, R. & Paredes, R. (2008). Should for profit schools be banned? Documento de Trabajo, Departamento de Ingeniería Industrial, Universidad Católica de Chile.

Contreras, D. (1999). Distribución del ingreso en Chile. Nueve hechos y algunos mitos, *Perspectivas*".

Colegio de Profesores de Chile (2006). La crisis del sistema educativo chileno: Diagnóstico y propuestas.

Eyzaguirre, B. & Fontaine, L. (1999). ¿Qué mide realmente el SIMCE? *Estudios Públicos*.

Organización para la Cooperación y el Desarrollo Económico OCDE. (2009). La educación superior en Chile.

Paredes, R. & Pinto, J.I. (2009). ¿El fin de la educación pública en Chile? *Estudios de Economía*, 36/1, pp. 47-66.

PNUD (1995). Human Development Report 1995. New York: New York University Press.

Sapelli, C. & Vial, B. (2002). The Performance of Private and Public Schools in the Chilean Voucher System. *Cuadernos de Economía*, 39/118, pp. 423-454.

Smith, M. L. (1991). Put to the Test: The Effects of External Testing on Teachers. *Educational Researcher*, 20/5.

Weissbluth, M. (2008). Crónica de una inequidad anunciada. *La Tercera*, 28 diciembre de 2008.

Weinstein, C. & Mignano, A. (1993). *Elementary Classroom Management: Lessons from research and practice*. Nueva York: McGraw-Hill.

Williams, W., Blythe, T., White, N.; Li, J., Stemberg, R. & Gardner, H. (1996). *Practical Intelligence for School*. Nueva York: Harper Collins.

Sitio Web DEMRE. <http://www.demre.cl/psu.htm>

Sitio Web Sistema de Medición de Calidad de la Educación. [www.simce.cl](http://www.simce.cl)