



Serie Evidencias:

Evaluaciones de la PSU: propuestas para mejorar el sistema de admisión

31 de Enero de 2014

Año 3, Nº 25

La evaluación y el análisis continuo de la Prueba de Selección Universitaria (PSU) es, sin lugar a dudas, un tema importante a ser estudiado dado el alto impacto que esta prueba tiene en la vida de las personas.

Si bien han existido evaluaciones anteriores respecto a esta temática, no todas ellas se habían hecho públicas. En base a lo anterior, el Ministerio de Educación, junto con el Consejo de Rectores de Universidades Chilenas, encargó en 2011 un estudio a la reconocida empresa educativa Pearson para poder conocer la calidad de este instrumento. Sus resultados, presentados en este documento, apuntan a que la prueba requiere de cambios urgentes.

1. Introducción

En Chile, quienes finalizan estudios superiores obtienen salarios 160% superiores respecto a quienes sólo terminan la etapa escolar¹, mejorando así sus posibilidades de desarrollo personal y profesional. Por ello, estudiar en alguna de las Instituciones de Educación Superior (Centros de Formación Técnica, Institutos Profesionales y Universidades) es una aspiración de la gran mayoría de los chilenos.

La Prueba de Selección Universitaria (PSU) es el instrumento principal de selección utilizado por muchas instituciones para determinar quiénes ingresarán a estudiar en la Educación Superior. Esta evaluación es administrada y desarrollada por un grupo de universidades pertenecientes al Consejo de Rectores de Universidades Chilenas (CRUCH), a través del Departamento de Evaluación, Medición y Registro Educacional (DEMRE) de la Universidad de Chile. Asimismo, esta agrupación de universidades cuenta con un sistema de admisión común, al cual se han ido adscribiendo en el tiempo otras universidades no pertenecientes a dicho consejo². El objetivo de esta prueba es precisamente evaluar el conocimiento de los estudiantes en distintas materias, con el fin de ordenarlos según su rendimiento y determinar su ingreso a las universidades adscritas a dicho sistema de admisión común³. Así, la Prueba de Selección Universitaria es el instrumento que anualmente determina en gran parte el futuro de miles de estudiantes.

Dado lo anterior, es fundamental que este instrumento sea evaluado y analizado continuamente para asegurar su calidad y coherencia con los objetivos que se plantea. En base a lo anterior, podemos distinguir tres tipos de análisis de validación para evaluar pruebas estandarizadas⁴:

1 OECD. (2013). Education at a Glance.

2 Desde el proceso de admisión 2012 se integraron por primera vez 8 universidades privadas en calidad de adscritas al sistema de admisión.

3 No obstante lo anterior, otras Instituciones de Educación Superior utilizan también el puntaje PSU para sus procesos de selección.

4 Phelps, R. (2014). Large-scale testing. Uses and abuses. Presentación realizada en la Universidad Finis Terrae en enero de 2014.

1. Validez predictiva, la cual mide hasta qué punto los resultados en las pruebas estandarizadas predicen resultados futuros (por ejemplo, rendimiento en la universidad).

2. Cobertura de contenidos, que evalúa en qué medida el test cubre los contenidos que dice evaluar (por ejemplo, los del sistema escolar).

3. Diferencias entre subgrupos, las cuales evalúan las diferencias de rendimientos entre subgrupos, por ejemplo, sexo, tipo de establecimiento, etc.

Dichos análisis permiten hacer correcciones y así mejorar los déficits que puedan encontrarse en esta prueba para así no perjudicar a quienes tienen los méritos para seguir sus estudios pero el sistema de selección no los está detectando.

Si bien existió una evaluación anterior a este instrumento que comprendía varios de estos análisis, encargada por la Universidad de Chile a la empresa Educational Testing Service (ETS) en 2004, sus resultados no fueron dados a conocer por la institución mandante sino sólo hasta 2012, cuando el informe fue solicitado a través de Ley de Transparencia, lo que obligó a hacerlos públicos.

Por ello, y considerando la importancia de realizar y transparentar los resultados de los análisis asociados a instrumentos que tienen consecuencias tan importantes, el Ministerio de Educación, junto con el Consejo de Rectores de Universidades Chilenas, luego de un proceso de licitación internacional, encargó un nuevo trabajo el año 2011 a la empresa de reconocida trayectoria en temas educativos Pearson, que permitiera conocer la calidad de la PSU.

El presente documento busca aportar a la mejora de este instrumento, mediante la revisión de diversos estudios realizados. Para ello, una primera sección resume los principales resultados del estudio de evaluación llevado a cabo por el Educational Testing Service (ETS) el año 2004. Luego, una segunda sección analiza los resultados del trabajo encargado a Pearson en 2011, en el cual se enfoca este documento. Para finalizar se entregan algunas conclusiones y recomendaciones derivadas del estudio.

2. Evaluación realizada por ETS

La evaluación realizada por ETS en 2004 analizó la batería de instrumentos que compone la PSU⁵ y generó ciertas conclusiones y recomendaciones orientadas a la mejora de este instrumento, las que se resumen a continuación:

- **Clarificar propósitos:** Es esencial que todos los propósitos y usos de los resultados de la PSU sean clarificados y que se desarrolle un plan para reunir evidencia con el fin de validar los puntajes de la evaluación para todos los propósitos que se les intente dar a las pruebas (por ejemplo para la admisión a la universidad, la asignación de ayudas estudiantiles, etc.).

5 Educational Testing Services. (2005). Evaluación Externa de la Prueba de Selección Universitaria (PSU).

- **Plan de equiparación de puntajes en el tiempo:** El estudio comprobó que los puntajes PSU de un año y otro no eran comparables, lo que debía ser corregido. El plan de equiparación (equating)⁶ debía considerar todas las pruebas de las evaluaciones pasadas y presentes, con el fin de permitir la comparabilidad para cada instrumento.

- **Uso de métodos de Teoría de Respuesta al Ítem (TRI):** Con el fin de estudiar la confiabilidad de la prueba, los grados de dificultad de las preguntas y la prueba en su conjunto, y hacerse cargo de las diferencias poblacionales que pudieran haber entre el grupo sobre el que se pilotean las preguntas y la población que rinde la prueba.

- **Plan comprensivo de equidad:** Que incluya la evaluación de confiabilidad de la prueba para distintos subgrupos, así como también un análisis de Funcionamiento Diferencial en Pruebas (DIF, por sus siglas en inglés)⁷.

- **Realizar investigaciones sobre las pruebas con resultados públicos para toda la comunidad.**

3. Estudio desarrollado por Pearson

Posteriormente al estudio de ETS, el MINEDUC junto con el CRUCH encargaron la realización de un estudio de evaluación de la PSU, de carácter público cuyo objetivo fue evaluar de forma extensa la calidad de la batería completa de las pruebas de la PSU, con el fin de entregar información que permitiera generar mejoras en el instrumento.

El trabajo fue encomendado a la empresa Pearson⁸, la cual fue elegida mediante un proceso de licitación. El análisis se centró en: (i) evaluación de procesos e ítems (procesos de construcción de preguntas -ítems- de la PSU) y (ii) evaluación de la validez del instrumento (puntajes de la PSU).

Para cumplir con los requerimientos anteriores, se consideraron varias fuentes principales de información:

- Bases de datos con información relevante (Puntajes PSU, rendimiento académico escolar y universitario, entre otras).

- Información de la PSU, pilotaje y versión operativa, desde la primera PSU hasta 2010.

- Documentación formal acerca de los procedimientos usados en el desarrollo de la PSU, así como en la generación de estos puntajes.

6 La equiparación es el proceso estadístico que se emplea para ajustar los puntajes en las pruebas a fin de que estos puedan ser utilizados de forma comparable e intercambiable.

7 Proceso que analiza la equivalencia del instrumento de medición entre subgrupos que componen la población de evaluados.

8 Pearson. (2013). Evaluación de la PSU Chile http://www.mineduc.cl/usuarios/mineduc/doc/201301311058200.ChilePSU-Resumen_Ejecutivo.pdf

- Entrevistas individuales a actores claves del DEMRE, MINEDUC, Consejo Técnico Asesor del CRUCH y otras instituciones.

- Datos de la PSU relativa a respuestas de los examinados, además de resultados psicométricos a nivel de ítem.

- Paneles de opinión, los cuales incluyeron a estudiantes, profesores y expertos en currículum.

A lo largo del estudio hubo una extensa interacción entre Pearson y una "Contraparte Técnica" a cargo de MINEDUC, integrada por tres personas nombradas por el Ministerio y tres por el CRUCH. Adicionalmente el Ministerio se asesoró por dos destacados expertos internacionales, Ronald Hambleton y Christina Stage.

3.1 Objetivos de la evaluación

Los dos grandes ámbitos de la evaluación (de procesos e ítems y de la validez del instrumento), se subdividieron en 18 objetivos, los que se reagruparon en 3, tal como se muestra en la Tabla 1.

Tabla 1: Objetivos de la evaluación desarrollada por Pearson

Agrupación de objetivos	Objetivo	Objetivos de evaluación
Desarrollo y uso de la PSU y sus ítems	Evaluar el proceso de desarrollo de la PSU y sus ítems	Desarrollo de ítems
		Pilotaje de ítems
		Construcción de pruebas
		Banco de ítems
		Muestreo piloto y selección de ítems
		Desempeño de ítems operacionales vs pilotos
		Fuentes de DIF exploratorias
		Puntajes de pruebas estandarizados
		Confiabilidad y error de medición condicionados (EEMCOND)
		Recomendación de un modelo para derivar puntajes de corte
Análisis de puntaje	Examinar los puntajes de las pruebas que surgen de la PSU con el fin de evaluar la calidad y confiabilidad de éstos.	Análisis del proceso usado para derivar un puntaje único para Ciencias
		Evaluar métodos TRI para calibrar ítems y equiparar puntajes
		Evaluar software para el análisis de ítems y banco de ítems
		Evaluar información de puntajes
Estudios de validación	Evaluar el grado en que la implementación de la PSU se orienta a cumplir los objetivos de dicho instrumento.	Estructura interna de constructo
		Validez de contenidos
		Cambio en el puntaje de la prueba
		Predicción de resultados universitarios

Fuente: Pearson (2013)

Para la evaluación de la PSU se usaron como base estándares profesionales de evaluación utilizados comúnmente a nivel internacional para analizar pruebas de este tipo, entre los que destaca Standards for Educational and Psychological Testing⁹, International Guidelines for Test Use¹⁰ y Program Evaluation Standards¹¹.

Si bien las siguientes secciones presentan un resumen de los hallazgos y recomendaciones en los tres grupos de objetivos presentados, el foco de este documento está en el análisis de los resultados de los estudios de validación, dada la importancia de este ámbito tanto en la construcción de los ítems como en la evaluación de los resultados de esta prueba en base a los objetivos que se plantea.

3.2 Principales recomendaciones

A partir de los hallazgos, el estudio presenta 124 recomendaciones para los tres grupos de objetivos antes mencionados. No obstante, es posible identificar tres recomendaciones generales, las cuales se presentan a continuación:

- **La base para el desarrollo¹² de pruebas de la PSU debería desplazarse desde la Teoría Clásica de Pruebas (TCT) hacia la Teoría de Respuesta al Ítem (TRI):** Ello porque el método TRI (i) modela específicamente la relación entre la habilidad del examinado y la dificultad de un ítem dado; (ii) permite analizar el grado de precisión de la medición como función de la habilidad del examinado, lo que implica que los desarrolladores de pruebas podrían determinar qué parte de la escala de puntaje de la PSU está siendo más exacta al medir; (iii) entrega un marco que permite calibrar ítems pilotos y ubicarlos en la misma escala del instrumento, y cuando los ítems están bien calibrados, los valores estadísticos de cada ítem no dependen de las características de la muestra específica y; (iv) en alguna de sus formulaciones, permite corregir por adivinación de una manera técnicamente más adecuada que la que se emplea actualmente (penalización por respuestas incorrectas), un mecanismo cuestionado por la literatura.

- **Las pruebas de la PSU deberían ser equiparadas:** La equiparación es el proceso estadístico que se emplea para ajustar los puntajes en las pruebas a fin de que estos puedan ser comparables e intercambiables en el tiempo y entre pruebas¹³. En la actualidad ningún puntaje PSU es comparable en el tiempo, por lo que Pearson propone que se realice una equiparación, utilizando el método TRI. Lo anterior es particularmente relevante debido a que los puntajes de la PSU pueden ser utilizados durante dos años.

9 AERA, APA & NCME. (1999). Standards for educational and psychological testing. Washington, D.C.

10 International Test Commission. (2012). ITC guidelines for quality control in scoring, test analysis, and reporting test scores.

11 Yarbrough, D., Shulha, L., Hopson, R. & Caruthers, F. (2011). The program evaluation standards: A guide for evaluators and evaluation users (3rd ed.).

12 Pearson argumenta las bondades de TRI para el análisis y construcción de pruebas (i.e. ensamblaje, equiparación y análisis varios de ítems).

13 Kolen, M., & Brennan, R. (2004). Test equating, scaling and linking: Methods and practices (2nd ed.). New York: Springer-Verlag.

Por otro lado, el estudio destaca que el DEMRE afirma que entrega un puntaje que, supuestamente, sería comparable entre los alumnos que rinden los distintos bloques de la prueba de Ciencias (recordar que ésta combina los resultados de una porción de ítems comunes para todos, y una sección electiva entre Biología, Química y Física¹⁴). Sin embargo, Pearson considera que bajo este diseño no es posible realizar procesos de equiparación válidos y que, en la práctica, estos puntajes no pueden ser considerados como comparables. En vistas de lo anterior, se recomienda la separación de las pruebas de Biología, Física y Química (o, al menos, la separación de puntajes), de manera de reconocer las diferencias entre ellas, permitiendo una mejor medición de los constructos subyacentes a cada disciplina. La separación de las pruebas permitiría evaluar el conocimiento de Biología, Física y Química, haciendo la distinción entre las tres áreas de medición.

• **El programa de la PSU debería desarrollar una estructura de investigación continua** para validar los usos e interpretaciones de las pruebas, entendiendo que las condiciones pueden cambiar en el tiempo y que una constante medición permite una permanente mejora de los instrumentos.

3.3 Estudio de validación de la PSU

La validez se refiere al grado en que la evidencia y la teoría respaldan la interpretación de puntajes de las pruebas en cuanto a los objetivos planteados por ésta¹⁵, lo cual constituye una medida esencial para la evaluación continua de la PSU.

En base a lo anterior es que si bien el informe presentado por Pearson incluye diagnóstico y recomendaciones para todos los objetivos de medición presentados, en la presente sección se profundizará en los objetivos relacionados a los estudios de validación de la PSU.

3.3.1 Hallazgos y recomendaciones

A continuación se presentan los principales hallazgos y recomendaciones relativas al proceso de validación de la PSU, desagregando según el objetivo de evaluación, de acuerdo a lo expuesto en la Tabla 1: i) Estructura interna del constructo, ii) Validez de contenidos, iii) Cambio en el puntaje de la prueba y, iv) Predicción de resultados universitarios.

i) Estructura interna de constructo

El propósito de este objetivo de evaluación es examinar la estructura interna de los ítems de la batería de pruebas de la PSU, lo cual representa la medida en que los puntajes de las pruebas son suficientes para recoger los componentes principales de las definiciones conceptuales de la prueba. Para lo anterior se utiliza un análisis factorial de los ítems, usando datos del proceso de

14 Este año se agregó una cuarta sección electiva solo elegible por estudiantes de la modalidad Técnico-Profesional y que sería similar al bloque común.

15 AERA, APA & NCME. (1999). Standards for educational and psychological testing. Washington, D.C.

admisión 2012. El análisis factorial a nivel de pregunta se realizó separadamente para todas las pruebas de la PSU, con el fin de determinar si cada una de las pruebas representa un único factor subyacente (unidimensionalidad).

Un segundo análisis fue el Funcionamiento Diferencial de Pruebas (FDP), el cual permite identificar si algún grupo en particular de la población que rinde la prueba obtiene mejores puntajes, generando evidencia preliminar¹⁶ sobre la equidad de la prueba.

Los análisis determinaron:

(i) La presencia de una única dimensión subyacente para cada una de las pruebas que componen la PSU, condición necesaria para utilizar el modelo de TRI.

(ii) Se encuentra evidencia de que hay diferencias en los resultados de acuerdo al grupo evaluado. Por ejemplo, estos son inferiores para los egresados de la modalidad Técnico-Profesional (en relación a los de Científico-Humanista) en las pruebas de Ciencias-Biología, Matemática, y Lenguaje y Comunicación. Asimismo, son mejores para los egresados de establecimientos particulares pagados (en relación a los del sector municipal) en las pruebas de Matemática, Lenguaje y Comunicación, Historia y Ciencias Sociales y Ciencias-Biología y para los de establecimientos particulares subvencionados (frente a los municipales) en Matemática. Debe tenerse en cuenta que este tipo de análisis no entrega información respecto de la causa de este “funcionamiento diferencial”, que puede ser un problema de la prueba, de un ítem en particular o el reflejo de inequidades existentes en el sistema educacional y no en el instrumento de medición.

Recomendaciones

• La **utilización del marco TRI** para el proceso de construcción de pruebas, los análisis de nivel de ítem y la equiparación.

• La realización de **análisis adicionales para entender mejor las diferencias** entre alumnos egresados de instituciones privadas y subvencionadas versus los del sector municipal, en particular en las pruebas de Lenguaje y Comunicación y Matemática. De la misma forma para el caso de los alumnos egresados de la modalidad Científico-Humanista versus Técnico-Profesional.

ii) Validez de contenidos

El objetivo de este análisis es evaluar la alineación de la PSU con el contenido que dice evaluar (currículo nacional). También se busca lograr una comprensión más profunda del alineamiento de la PSU desde la perspectiva del profesor de Enseñanza Media y del profesor universitario, específicamente en lo que se refiere a la alineación con el currículo implementado en la sala de clases (que no siempre es semejante al currículo teórico descrito en las bases curriculares) y los conocimientos y destrezas requeridos

16 Un análisis FDP por sí solo no es concluyente respecto de temas de equidad o sesgo, pero da luces sobre cuáles son los grupos donde dirigir las sospechas, que deben ser confirmadas o refutadas mediante análisis más profundos y exhaustivos.

por la Educación Superior para el éxito académico de los estudiantes (que tampoco corresponden necesariamente a todos los conocimientos y destrezas descritos en el currículum).

Lo anterior se realizó mediante la revisión de la alineación de las pruebas con las bases curriculares vigentes en base a las dimensiones del análisis de Webb¹⁷.

Los resultados del estudio de alineamiento indican que:

(i) Para casi todas las pruebas el nivel de alineamiento de la PSU con los Objetivos Fundamentales (OF) y los Contenidos Mínimos Obligatorios (CMO) del currículum chileno es bajo.

(ii) Existe una desconexión fundamental entre el propósito y uso de la PSU (seleccionar alumnos para su ingreso a las universidades) y el marco sobre el que teóricamente se construiría la PSU (alineación estrecha con el currículum de la Enseñanza Media). El marco actual podría ser apropiado para pruebas cuyo propósito fuese medir el grado de cobertura y adquisición del currículum por parte de los estudiantes de Enseñanza Media, lo que no es propósito principal de una prueba de admisión.

(iii) En cuanto al uso de la PSU como un predictor de éxito en la Educación Superior, un tema importante en las entrevistas fue que la PSU no captura todas las aptitudes o habilidades necesarias para tener éxito en la Educación Superior.

Recomendaciones

• **Revisión de la política de alinear estrechamente el marco curricular con las pruebas de la PSU.** Se propone el desarrollo de un marco que describa las aptitudes (por ejemplo, habilidades) y variables no cognitivas relevantes (como por ejemplo, motivación) que los estudiantes necesitan para tener mayor probabilidad de éxito en la Educación Superior.

• **En caso de mantener la alineación estrecha entre la PSU y el currículum de Enseñanza Media, se recomienda mejorar la calidad de ésta** (evaluada por Pearson como baja), y migrar plenamente al nuevo marco curricular del año 2009.

• **Revisar los tipos de ítems** usados en las pruebas para abordar el nivel de complejidad percibido en las pruebas.

iii) Cambio en el puntaje de la prueba

El seguimiento en el tiempo de los puntajes de admisión de determinados grupos permite localizar tendencias descendentes y ascendentes de determinadas subpoblaciones, y así diseñar políticas que apunten a disminuir las brechas entre grupos. Sin embargo, dado que los puntajes en las pruebas no son comparables

¹⁷ Las perspectivas consideradas en el método de Webb son (i) alineación del contenido que se evalúa; (ii) distribución de las preguntas de las evaluaciones; (iii) requerimiento cognitivo de las preguntas en las evaluaciones y; (iv) número de contenidos de los estándares que se están evaluando. Case, B., Jorgensen, M., Zucker, S. (2004). Alignment in educational Assessment

en el tiempo (no se ha implementado la recomendación de ETS en 2004 de equiparación o equating), la utilidad de este estudio respecto a este tema es limitada y sus conclusiones deben de ser consideradas con cautela.

Los análisis de tendencia arrojaron que:

(i) En promedio, los puntajes PSU se han mantenido con una pequeña tendencia al alza desde el año 2007 (considerando a quienes participaron del sistema único de admisión).

(ii) En línea con esta tendencia, se observa que los establecimientos particulares y los provenientes de la rama Científico-Humanista han subido sus puntajes, mientras que los establecimientos municipales y la rama Técnico-Profesional mantuvieron sus puntajes estables¹⁸.

Recomendaciones

• **Equiparación del puntaje de la prueba** de manera anual, con el fin de generar comparabilidad entre años, especialmente considerando que en Chile los puntajes de la PSU pueden usarse durante dos años.

• **Análisis continuo de evolución de puntajes entre subpoblaciones** relevantes de postulantes, con el fin de detectar tendencias en subpoblaciones.

iv) Predicción de resultados universitarios

Mediante el análisis de un conjunto de datos longitudinales para postulaciones universitarias que comprendieron el periodo entre 2004 y 2012, se analizó:

• La capacidad de los puntajes PSU y del desempeño académico del estudiante durante la Enseñanza Media para predecir sus resultados académicos universitarios.

• El valor de predicción del ranking de notas "puro", es decir, la posición relativa del estudiante respecto con sus compañeros de generación (en el mismo establecimiento)¹⁹.

• Si los puntajes en la PSU y el desempeño académico en la Enseñanza Media predicen el rendimiento y las tasas de graduación universitaria

¹⁸ El aumento de los beneficios estudiantiles en este periodo (2004-2011) determinó que hubiera un aumento en el número de estudiantes de establecimientos municipales y técnico-profesional que rendían la PSU.

¹⁹ Esta definición de ranking es distinta a la utilizada por el DEMRE para sus puntajes, que establece una ordenación en base a la posición del alumno con respecto a las últimas tres generaciones de su establecimiento educacional, incluyendo también en el algoritmo de "puntaje ranking" una valoración de su promedio de notas absoluto.

Los hallazgos del estudio indican que:

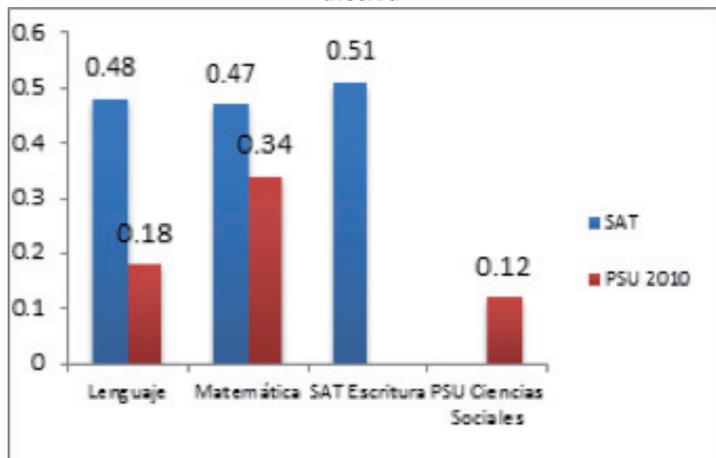
(i) La PSU tiene la capacidad de predecir los resultados universitarios hasta un cierto punto (respecto de los promedios del primer y segundo año universitario).

(ii) La variable "ranking en la enseñanza media"²⁰ contribuye a mejorar la predicción de los resultados universitarios luego de controlar los puntajes de la PSU y las Notas de Enseñanza Media (NEM). En esta misma línea, Gallegos & Meneses (2006)²¹ encuentran que la utilización de un ranking puro de las notas de Enseñanza Media, aumenta la validez predictiva del instrumento.

(iii) Los resultados por tipo de carrera arrojaron conclusiones similares a las encontradas a nivel general. En particular, se encontró que los puntajes de la PSU de Matemática y de Ciencias y el desempeño académico de la Enseñanza Media (NEM y ranking) mostraron una mayor capacidad predictiva que los puntajes en las pruebas de Lenguaje y Comunicación e Historia y Ciencias Sociales.

A modo de complemento del trabajo de Pearson, se presentan los resultados del estudio realizado por Richard Phelps²², que indican que la validez predictiva de la PSU es inferior al nivel presentado por el examen estandarizado que se usa para el proceso de admisión de las universidades de Estados Unidos (SAT). La siguiente figura muestra la comparación entre ambas pruebas²³.

Gráfico 1: Comparación PSU-SAT en cuanto a su validez predictiva



Fuente: Phelps (2014)

²⁰ Ver nota anterior.

²¹ Gallegos, S., Meneses, F. (2006). ¿Es eficiente el sistema de ingreso a la universidad? El uso del ranking en la Universidad Católica de Chile.

²² Presentado en la Universidad Finis Terrae en enero de 2014.

²³ La validez predictiva fluctúa entre -1 y 1. Mientras más cerca del 0, el valor indica menor capacidad predictiva.

Recomendaciones

• **Desarrollar evidencia de respaldo para el uso y sentido de la PSU**, como estudios de validez para establecer evidencia que respalde el proceso de toma de decisiones.

• **Incorporar en el análisis de validez predictiva** el desempeño de los alumnos en etapas posteriores, como continuación de estudios de postgrado, inserción en el mundo laboral y salario al egresar de la carrera.

• **Investigar la uniformidad en las prácticas de asignación de notas entre diferentes instituciones de Educación Superior** para una misma carrera, para avalar los hallazgos de las medidas de validez predictiva que emplean el promedio de notas.

• **Sofisticar los análisis para evaluar la capacidad predictiva de la prueba en la población general.** Los análisis hasta ahora evalúan el desempeño de la prueba para predecir los desempeños de alumnos que fueron seleccionados en las universidades estudiadas. No sabemos si esos resultados son aplicables para predecir los desempeños potenciales de los estudiantes que quedaron fuera.

4. Principales conclusiones

Dada la importancia de evaluar un instrumento con consecuencias tan relevantes en la vida de las personas, como las PSU, en el presente documento se dan a conocer tanto las conclusiones de la evaluación realizada a dicho instrumento en 2004 por ETS cuyos resultados no se hicieron públicos sino hasta 2012, como del trabajo de Pearson realizado el mismo año. Además de dar una mirada general a los resultados, se indagó en el análisis de validez de la PSU, debido a su importancia en la construcción de los ítems y en la evaluación de los resultados de las pruebas respecto a si cumple con los objetivos que se plantea, presentándose los hallazgos y recomendaciones más relevantes.

En primer lugar, preocupa que los resultados del estudio del ETS no fueran hecho públicos sino hasta siete años después de completado, perjudicando la transparencia y credibilidad del sistema. Asimismo, al observar las falencias encontradas a la PSU en dicho trabajo y verificar que una proporción importante de ellas se mantiene en el tiempo, se concluye que varias generaciones de estudiantes fueron perjudicadas al no corregirse dichos problemas, tales como la comparabilidad en el tiempo de los instrumentos y las bajas tasas de predictibilidad de sus resultados.

En segundo lugar, una conclusión general del trabajo de Pearson y que sigue la línea de lo presentado por el ETS en el año 2004 es la necesidad de introducir la metodología de Teoría de Respuesta al Ítem (TRI), tanto para el análisis de las preguntas como para el ensamblaje de las pruebas y la equiparación de los puntajes en el tiempo.

En tercer lugar, es urgente investigar la existencia de los potenciales sesgos de las pruebas hacia egresados de establecimientos particulares y del área Científico-Humanista, para, en caso de validarlos, determinar sus causas y tomar medidas lo antes posible.

En cuarto lugar, se requiere realizar mejoras a la prueba de Ciencias, puesto que en la actualidad se utilizan indistintamente los resultados de las diferentes pruebas, sin ser ellas equiparables. Se sugiere reconocer la diferencia entre las tres áreas evaluadas (Biología, Física y Química), a través de la separación de los puntajes o de las evaluaciones.

Lo anterior indica que desde que se reemplazó a la Prueba de Aptitud Académica (PAA) por la PSU hasta la fecha, se afectó a varias generaciones y en particular a determinados grupos identificados en el análisis, como los alumnos provenientes de establecimientos que imparten la modalidad Técnico-Profesional, y de establecimientos municipales, perjudicando sus posibilidades de acceder a mejores condiciones de vida.

Por ello la señal entregada por el Consejo de Rectores de las Universidades Chilenas en 2013²⁴ respecto de algunos cambios y anuncios, que se presentan a continuación, es un indicio muy positivo para los jóvenes.

Cambios:

• A partir del proceso de admisión 2014 se generó un nuevo módulo diferenciado electivo de la prueba de Ciencias para estudiantes provenientes de establecimientos que imparten la modalidad Técnico-Profesional con el fin de cubrir las diferencias en el currículum entre ambas modalidades.

Anuncios:

• Incorporación de manera progresiva el método de TRI en la PSU. No se ha especificado, sin embargo, la voluntad de utilizar este método para hacer comparables las pruebas en el tiempo.

• Eliminación en el proceso de admisión 2015 el uso de puntuación que penaliza por respuestas erróneas.

• Se incluirán los análisis de diferencias por subgrupos (según modalidad de enseñanza y nivel socioeconómico).

Si bien lo anterior es un avance, aún existen aspectos pendientes, tales como:

• **Incorporación de la medición de habilidades:** En este sentido, Koljatic et al (2012)²⁵ argumentan que hay que tener precauciones al usar pruebas de conocimientos en países donde existe una desigual distribución de los contenidos entregados, como ocurre en el caso de Chile, y que perjudica a los alumnos provenientes de sectores más vulnerables.

• **Alineación de contenidos PSU en base a currículum,** en caso de continuar con la evaluación de contenidos.

24 Presentación de Consejo de Rectores de Universidades Chilenas (2013). Disponible en http://www.consejodirectores.cl/web/pdf/acta-2013/543/Anexo_20.pdf

25 Koljatic, M., Silva, M., Cofré. (2012). Achievements versus aptitudes in college admissions: A cautionary note based on evidence from Chile.

• **Separar las pruebas de Ciencias,** con el fin de reconocer de que las distintas disciplinas científicas son diferentes y los desempeños en una y otra no son equiparables.

• **Urgencia de generar pruebas comparables,** sobre todo ahora que los puntajes de admisión son válidos por más de un período. Si bien esto requiere de resguardos extraordinarios ante la filtración de preguntas (dado que para comparar las pruebas se debe repetir un grupo de preguntas año a año), esto ha sido posible de manejar tanto en Chile en otras pruebas, como en otros países.

Por lo tanto, si bien hay avances en la calidad de la PSU, se requiere urgentemente abordar todos los aspectos pendientes, para no perjudicar el futuro y bienestar de nuevas generaciones.