

Medición y evaluación para los procesos de admisión de la educación superior: Hallazgos desde California

Measurement and Evaluation for Higher Education Admissions: Findings from California

Saul Geiser

Center for Studies in Higher Education, University of California, Berkeley, EE.UU.

Resumen

El presente artículo entrega una visión general de los esfuerzos de medición y evaluación que han conformado el núcleo del sistema de admisión de la Universidad de California a lo largo del siglo pasado, enfocándose particularmente en los avances de los últimos 30 años. Describe la experiencia, las tensiones y los desafíos del proceso de admisión que se asocian con el uso de pruebas estandarizadas, así como los caminos de admisión alternativos implementados más recientemente. El artículo concluye con una visión general de los desafíos futuros para la institución de educación superior más selectiva en California, los que también pueden ser relevantes para otras instituciones de educación superior de todo el mundo.

Palabras clave: admisión a las universidades, pruebas de admisión, SAT, estatus socioeconómico

Correspondencia a:

Saul Geiser
Center for Studies in Higher Education
University of California, Berkeley, USA.
771 Evans Hall #4650, Berkeley, CA, USA
Correo electrónico: sgeiser@berkeley.edu

© 2016 PEL, <http://www.pensamientoeducativo.org> - <http://www.pel.cl>

ISSN: 0719-0409 DDI: 203.262, Santiago, Chile
doi: 10.7764/PEL.53.1.2016.7

Abstract

This article provides an overview of the measurement and evaluation efforts that have been at the heart of the University of California's admissions system throughout the last century, with a particular focus on developments over the past 30 years. It describes the admissions experience, tensions, and challenges presented by the use of standardized testing, as well as alternative admissions paths implemented more recently. The article ends with an overview of the challenges that lie ahead for the most selective higher education institution in California; these may also be relevant to other higher education institutions around the world.

Keywords: university admissions, admissions testing, SAT, socioeconomic status

Durante las últimas décadas, la Universidad de California (University of California, UC) ha estado a la vanguardia de la discusión nacional sobre el papel de las pruebas estandarizadas en la admisión a las universidades, impulsada por el marginal aporte de las puntuaciones de las pruebas estandarizadas a la predicción de los resultados de quienes asisten a la universidad y su significativo impacto negativo en los estudiantes pertenecientes a minorías. Motivada por la fuerte baja en el ingreso de estudiantes pertenecientes a minorías como consecuencia de la prohibición de la acción afirmativa aprobada por los votantes de California en 1996, la UC ha encabezado la implementación de nuevas políticas y procesos de admisión, explorando alternativas que posiblemente resulten útiles para otras instituciones.

El presente artículo entrega una visión general de aquella discusión sobre políticas de admisión y de la experiencia de la UC en cuanto a instaurar políticas alternativas. El artículo comienza con una descripción de la historia de las pruebas estandarizadas en Estados Unidos y luego se concentra en la experiencia de la Universidad de California y el SAT, el crecimiento de las pruebas con referencia a criterios y las lecciones derivadas de la implementación de políticas después de 2014. Me baso en los resultados de la agenda de investigación que he desarrollado durante los últimos quince años, primero como jefe de la Oficina de Investigación sobre Admisión de la Oficina del Presidente de la Universidad de California y luego como investigador en el Centro de Estudios sobre Educación Superior de la Universidad de California en Berkeley (Geiser, 2009, 2014, 2015; Geiser & Atkinson, 2013; Geiser & Caspary, 2005; Geiser & Santelices, 2006, 2007; Geiser & Studley, 2002).

Pruebas estandarizadas

La evaluación estandarizada para la admisión universitaria comenzó en 1926 con el SAT, en ese entonces conocido como la Prueba de Aptitud Académica [Scholastic Aptitude Test]. Hasta ese momento, las pruebas de admisión en Estados Unidos habían involucrado exámenes escritos intensivos sobre diferentes materias. El SAT ofrecía algo completamente nuevo: una prueba de opción múltiple fácil de calificar para medir la capacidad general o aptitud de los estudiantes para aprender.

El SAT se originó a partir de la experiencia de las pruebas de Coeficiente Intelectual (CI) aplicadas a los soldados durante la Primera Guerra Mundial. Los creadores de ambas pruebas compartían varios supuestos que la mayoría de los especialistas hoy consideran problemáticos: que la inteligencia es un rasgo unitario y heredado; que no está sujeta a cambios a lo largo de la vida; y que puede medirse con un número único.

Sin embargo, el SAT resultó ser atractivo para los institutos y universidades estadounidenses por muchas razones. Era estandarizada, a diferencia de las calificaciones de la escuela secundaria y, por lo tanto, podía usarse para comparar a los postulantes de diferentes escuelas. De acuerdo a sus promotores, podía identificar a estudiantes prometedores de escuelas más pobres que de otro modo podrían no ser aceptados. Por sobre todo, el SAT ofrecía una herramienta de predicción que les entregaba a los encargados de admisión un medio para distinguir entre postulantes que se desempeñarían bien o mal en la universidad. Es fácil entender por qué la prueba se hizo popular en los años que siguieron a la Segunda Guerra Mundial.

Fiel a sus orígenes en la evaluación del CI, el SAT durante largo tiempo ha sido conocido por sus ítems complicados, similares a los de un puzzle. La Figura 1 muestra un tipo de ítem muy discutido, denominado *analogía verbal*, que se usó hasta hace unos pocos años.

DOCTRINA: TEÓLOGO
 (A) Predecesor: heredero
 (B) Hipótesis: biólogo
 (C) Solitario: rivalidad
 (D) Arrogancia: persecución
 (E) Guitarrista: banda de rock

Respuesta: (B)

Figura 1. Ejemplo de ítem del SAT verbal (pre-2005). Fuente: College Board.

Este tipo de ítem fue muy criticado porque los estudiantes normalmente no aprenden analogías verbales como parte de sus estudios regulares en la educación secundaria. Aunque tenía la intención de medir el razonamiento verbal, sus críticos afirmaron que solo medía la capacidad de los estudiantes de pagar servicios de preparación para la prueba, en los que se les enseñaban analogías verbales. Este tipo de ítem fue retirado del SAT en 2005.

El ACT entra en escena

El ACT (un acrónimo de American College Testing [Exámenes de Ingreso a las Universidades Estadounidenses]) fue presentado en 1959 como un competidor del SAT. El ACT encarnaba un enfoque alternativo para las pruebas de admisión. En lugar de evaluar la capacidad o aptitud general de los estudiantes, el ACT fue diseñado para medir el logro de los estudiantes, es decir, cuánto habían aprendido en la escuela y el grado en el cual habían logrado dominar las diferentes asignaturas. La percepción común es que, mientras el SAT mide la capacidad innata de los estudiantes, el ACT mide cuánto aprendieron en sus estudios escolares.

Consistente con esta filosofía, el ACT se alinea más estrechamente con los currículos de las escuelas secundarias que el SAT. Los ítems de las pruebas se desarrollan a partir de los cursos ofrecidos a nivel secundario en todos los Estados Unidos. El ACT parece ser menos entrenable, y el consenso es que el ACT enfatiza menos las habilidades para rendir pruebas y más el conocimiento acabado del currículo. Siguiendo la filosofía del ACT, los ítems incluidos en las pruebas tienden a ser más directos que los del SAT. La Figura 2 muestra un ítem representativo de la sección matemática del ACT.

En promedio, un automóvil rinde 27 millas por galón (11,5 kilómetros por litro). Si la gasolina cuesta \$4,04 por galón (\$1,07 por litro), ¿cuál de las siguientes alternativas se acerca más al costo de la gasolina necesaria para que el automóvil viaje 2.727 millas (4.373 kilómetros) en una autopista típica?

A) \$44,44
 B) \$109,08
 C) \$118,80
 D) \$408,04
 E) \$444,40

Respuesta: (D)

Figura 2. Ejemplo de ítem de matemática del ACT. Fuente: ACT.

Responder el ítem requiere solamente un cálculo directo. El mayor desafío para los estudiantes es lo que se conoce como *la restricción en el tiempo* de la prueba, es decir, la necesidad de responder un gran número de ítems, como este, en un tiempo acotado.

El SAT ha sido la prueba de admisión dominante en EE.UU. durante un largo tiempo, pero el ACT ha logrado avances en los últimos años; en 2012, por primera vez, hubo más estudiantes que rindieron el ACT que el SAT. El SAT predomina en las Costas Este y Oeste, así como en Texas, mientras que el ACT predomina en los estados del medio oeste. California es en su mayoría un estado SAT, aunque la Universidad de California también acepta el ACT. De todos los estados, California es el mayor usuario del SAT.

Convergencia del SAT y el ACT

A pesar de sus distintos orígenes, el SAT y el ACT han ido convergiendo con el tiempo y se han hecho más similares. El SAT ha eliminado varios de sus tipos de ítem más complicados, como las analogías verbales y las comparaciones cuantitativas. Por su parte, el ACT se ha vuelto más restringido en el tiempo, haciendo mayor hincapié en las habilidades de manejo del tiempo de los estudiantes y en el recuerdo rápido, con lo que se ha ido asemejando al SAT en ese aspecto.

Ambos han añadido también una prueba de escritura, respondiendo en parte a las investigaciones realizadas en la Universidad de California que muestran que la prueba de escritura era uno de los mejores predictores de rendimiento en la universidad. Ambas pruebas tienen aproximadamente la misma capacidad predictiva, y hoy casi todas las universidades estadounidenses aceptan las dos y ven las puntuaciones del SAT y del ACT como intercambiables. Además, el punto central es que ambas son pruebas con referencia a normas.

Pruebas con referencia a normas versus pruebas con referencia a criterios

Las pruebas con referencia a normas buscan ordenar a los estudiantes, es decir, compararlos con otros estudiantes que rinden la misma prueba. Están diseñadas para producir una distribución normal.

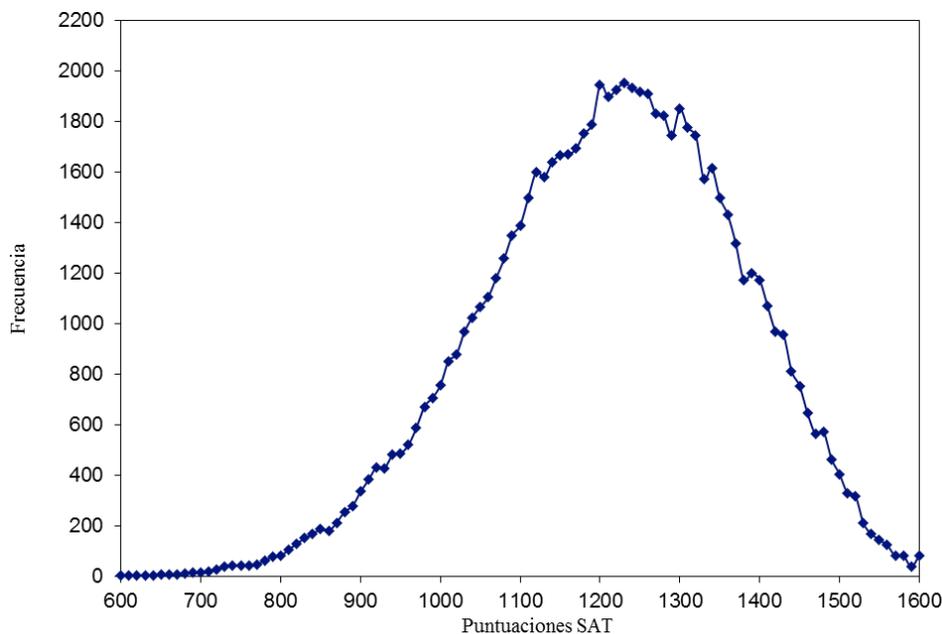


Figura 3. La distribución normal: distribución de puntuaciones SAT de estudiantes californianos.

La distribución normal es producto de la manera en que se construye la prueba y de la forma en que se crea la escala. Al desarrollar la prueba, el diseñador descarta los ítems que demasiados estudiantes pueden responder correctamente; el ítem ideal es el que divide en partes iguales a la población que rinde la prueba.

Luego, en el escalamiento de los puntajes, los puntajes originales se convierten en puntuaciones escaladas de una manera que producen «colas» con pendientes a ambos lados de la distribución.

Existe una importante diferencia entre las pruebas con referencia a normas versus las pruebas con referencia a criterios en este aspecto. Las evaluaciones con referencia a criterios miden cuánto saben los estudiantes sobre un tema dado, más que cómo se comparan con otros. En una prueba con referencia a criterios, es posible que la mayoría o incluso todos los estudiantes logren un buen desempeño, suponiendo que saben el tema sobre el cual se les pregunta. Esto no es posible en una prueba con referencia a normas debido al modo en que se califica el rendimiento en una curva.

Algunos estados en EE.UU. usan el SAT o el ACT para medir el logro académico en sus escuelas K-12 (desde kindergarten hasta el duodécimo grado). Sin embargo, debido a que las pruebas con referencia a normas están diseñadas para producir siempre una distribución normal, no son apropiadas para medir cambios en el logro académico dentro de cualquier población de estudiantes o escuelas. Los diseñadores de pruebas regularmente eliminan los ítems que demasiados estudiantes responden correctamente, por lo que es imposible que las pruebas detecten mejoras en logro académico a través del tiempo.

A pesar de las limitaciones de las pruebas con referencia a normas como forma de medir lo que han aprendido los estudiantes (a diferencia de cómo se comparan con otros estudiantes) las instituciones de educación superior selectivas las consideran muy útiles para fines de admisión. Esto es especialmente cierto para institutos y universidades de alta selectividad que reciben muchas de postulaciones. En estos casos, las puntuaciones en estas pruebas entregan una medición breve y aparentemente objetiva de la capacidad de los estudiantes que hace fácil compararlos y ordenarlos. Debido a su diseño, las pruebas con referencia a normas destacan pequeñas diferencias en el extremo superior de la distribución de puntuaciones, dentro del cual las universidades seleccionan a sus estudiantes. Esto permite que los funcionarios de admisión establezcan diferencias sutiles dentro de grupos de postulantes donde casi todos los estudiantes han alcanzado un alto rendimiento en la educación secundaria.

La Universidad de California y el SAT

En 1996, la UC quedó sumida en una gran crisis política que la hizo reevaluar el uso de pruebas estandarizadas. Los californianos votaron para terminar con la acción afirmativa en la admisión a la UC. Hasta entonces, la UC había considerado la raza como un factor en el proceso de admisión, como una forma de aumentar el número de estudiantes matriculados pertenecientes a grupos minoritarios con tasas de asistencia a la universidad históricamente bajas. Luego de la eliminación de la acción afirmativa, el número de estudiantes latinos y afroamericanos disminuyó bruscamente en Berkeley, UCLA y otros campus de la UC, en un momento en el cual estos grupos se acercaban rápidamente a convertirse en mayoría dentro de la población de las escuelas K-12 de California. Como institución pública, cuya misión es servir a todos los californianos, la UC comenzó a recibir fuertes presiones para reexaminar sus criterios de admisión. Lo que se descubrió puso en duda muchas creencias establecidas sobre las pruebas de admisión y terminó por impulsar numerosas reformas relevantes en el campo de la admisión.

Una sorpresa fue constatar que el mejor predictor de resultados en la universidad no fueran las pruebas estandarizadas sino las calificaciones escolares. Los resultados en la universidad pueden medirse de múltiples formas. Las notas durante el primer año en la UC son la forma de medición empleada aquí, pero nuestros hallazgos fueron muy similares al emplear otros criterios, como las tasas de graduación. Sin importar el criterio de resultados empleado, las notas en la escuela secundaria representaron la mayor parte de la varianza explicada.

Predictores	Varianza explicada
HSGPA (Promedio de notas en la escuela secundaria)	15,4%
SAT	
HSGPA + SAT	13,3%
	20,8%
Incremento debido al SAT	+ 5,4%

Figura 4. Predicción de las calificaciones de los alumnos de primer año de la UC. Fuente: Geiser y Studley, 2002.

Muchas personas se sorprenden por el hecho de que las calificaciones en la escuela secundaria sean un mejor predictor que las puntuaciones en pruebas, debido a que los estándares para asignar calificaciones varían de una escuela a otra. Sin embargo, la puntuación en una prueba se basa en los resultados de una única aplicación de 3 o 4 horas. Por otra parte, las notas en la escuela secundaria se basan en una evaluación repetida del desempeño de los estudiantes a lo largo de un período de varios años.

Hubo otros dos hallazgos dignos de mencionar. En primer lugar, el nivel general de predicción es relativamente modesto. En su conjunto, las calificaciones de la escuela secundaria y las puntuaciones en las pruebas representaron solo el 21% de la varianza en el desempeño en la universidad, quedando casi un 80% sin explicación. Nuestra capacidad de predecir los resultados en la universidad es relativamente limitada. En segundo lugar, las puntuaciones en las pruebas añaden poco a la predicción del desempeño en la universidad luego de que se toman en cuenta las calificaciones en la escuela secundaria. Añadir las puntuaciones en las pruebas mejoró la predicción solo en unos cinco puntos porcentuales.

Error de predicción

Debido a su poca capacidad predictiva, las predicciones basadas en puntuaciones de pruebas están sujetas a un grado considerable de error. La figura que se muestra a continuación compara a dos estudiantes idénticos en todos los aspectos (las mismas notas en la escuela secundaria, el mismo origen socioeconómico, etc.) excepto en sus puntuaciones en las pruebas. Los datos no son hipotéticos, sino que se basan en una muestra de casi 80.000 estudiantes de la UC. Se predice que el Estudiante A, con 1200 puntos en el SAT, logrará un promedio de notas (GPA) de 3,0, una nota promedio B durante su primer año en la UC. La Estudiante B es idéntica al Estudiante A, excepto por el hecho de que alcanzó 1300 puntos en el SAT, 100 más que el estudiante A. Se predice que obtendrá un GPA de 3,13 en la UC.

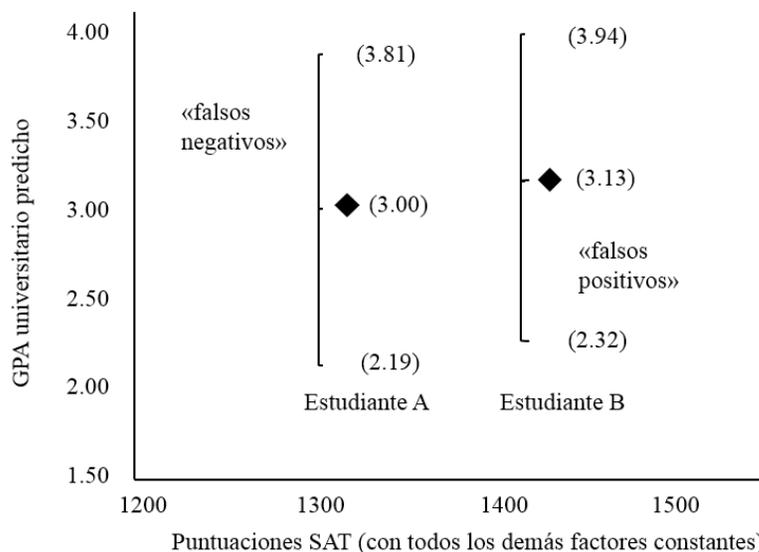


Figura 5. Errores de predicción. Fuente: Cálculos basados en un análisis de regresión de 77.893 alumnos de primer año recién ingresados a la UC entre 1996 y 1999.

La mayoría de los encargados de admisión seleccionarían automáticamente a la Estudiante B por sobre el Estudiante A. Las puntuaciones en pruebas comúnmente se usan para romper la igualdad entre estudiantes con las mismas notas en la escuela secundaria y, en este caso, la balanza se inclinaría por la Estudiante B.

Lo que este método ignora, sin embargo, son las bandas de error alrededor de las predicciones. Debido a que las puntuaciones en las pruebas representan una fracción tan pequeña de la varianza total en los resultados en la universidad, las bandas de error alrededor de las predicciones son bastante grandes. En esta muestra de 80.000 estudiantes, las bandas de error fueron más o menos 0,81 puntos de calificación en el nivel de confianza del 95%. Esto quiere decir que, para ambos estudiantes, las bandas de error son bastante anchas, y van desde un promedio A (rendimiento sobresaliente) hasta un promedio C (un rendimiento relativamente bajo en la universidad).

Y he ahí el problema: dado que los intervalos de confianza son tan amplios, las predicciones basadas en puntuaciones de pruebas están sujetas a errores de dos tipos. En primer lugar, están los falsos positivos. En una gran parte de los casos, el estudiante aceptado de hecho tendrá un desempeño inferior al que habrían tenido los estudiantes a quienes se les negó la admisión. En segundo lugar, se encuentran los falsos negativos: los encargados de las admisiones podrían rechazar a estudiantes que habrían tenido éxito en la universidad y que se hubiesen desempeñado mejor que los aceptados. Ambos tipos de error son inevitables cuando la capacidad predictiva de las pruebas es baja y las diferencias de puntuación son pequeñas, lo que suele ocurrir en los procesos de admisión en institutos y universidades de alta selectividad, cuyos postulantes casi sin excepción obtienen altas puntuaciones.

El impacto del estatus socioeconómico en las puntuaciones de las pruebas

Otra sorpresa en nuestras investigaciones en la UC fue el grado en el cual las puntuaciones en las pruebas actúan como una barrera para impedir el ingreso de estudiantes de contextos desfavorecidos. Cuando se implementó por primera vez el SAT, uno de sus atractivos fue que sus creadores afirmaban que era capaz de identificar estudiantes talentosos de origen más humilde, quienes de otro modo no serían aceptados en universidades de primer nivel. Muchos defensores del SAT siguen creyendo esto.

Sin embargo, nuestros datos muestran que las puntuaciones en el SAT se correlacionan mucho más fuertemente con el nivel socioeconómico de los estudiantes que las calificaciones en la escuela secundaria u otros criterios de admisión. Entre los postulantes a la UC, la correlación entre puntuación en el SAT e ingreso familiar fue de aproximadamente 0,3, mientras que la correlación con la educación de los padres fue de alrededor de 0,4. Por otra parte, la correlación de las calificaciones en la educación secundaria con ambas mediciones fue inferior a 0,1.

	Familia	Nivel educacional de los padres
Puntuación SAT	,32	,40
GPA educación secundaria	,04	,06

Figura 6. Correlación de puntuaciones en el SAT y GPA en la educación secundaria (HSGPA) con estatus socioeconómico. Fuente: Geiser y Santelices (2007).

Producto de esto, cuando se emplean para ordenar a los estudiantes, las puntuaciones en las pruebas tienen un efecto mucho más negativo que otros criterios sobre la admisión de estudiantes pertenecientes a minorías, quienes provienen de contextos desfavorecidos en tasas desproporcionadamente altas. Esto se demuestra en la Figura 7.

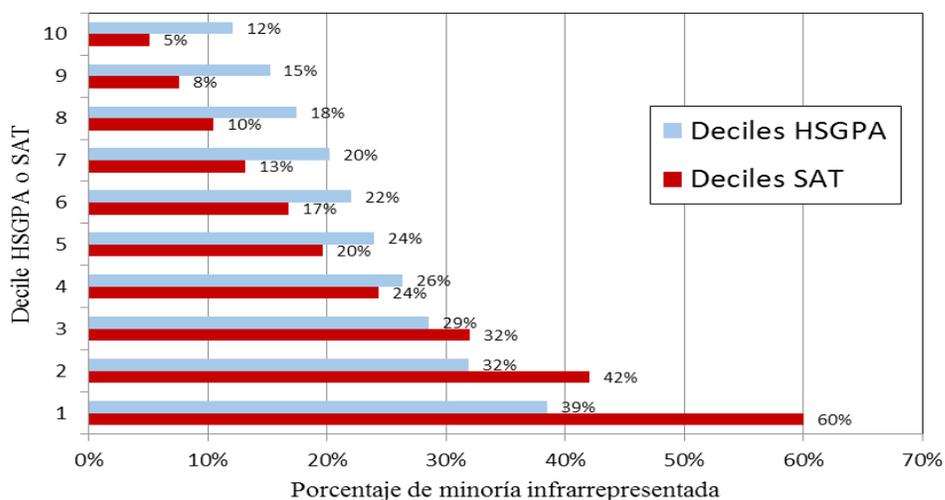


Figura 7. Distribución de postulantados pertenecientes a minorías subrepresentadas por deciles SAT versus deciles GPA en la escuela secundaria (1994-2011). Fuente: Datos del Sistema Corporativo de Estudiantes de la UC sobre postulantados de primer año residentes en California (1994-2011).

Las barras más claras muestran el porcentaje de estudiantes pertenecientes a minorías étnicas (latinos, afroamericanos y nativos americanos) dentro de cada decil de GPA en la escuela secundaria (HSGPA), desde el 10% superior hasta el 10% inferior de todos los postulantados a la UC ordenados según sus calificaciones en la escuela secundaria.

Las barras más oscuras muestran su porcentaje dentro de cada decil del SAT. Se debe tener en cuenta que estos son los mismos estudiantes, pero ordenados de acuerdo a criterios diferentes. A pesar de que los estudiantes pertenecientes a minorías tienden a agruparse en la parte baja de ambas distribuciones, la estratificación racial producida por las puntuaciones en el SAT es mucho más marcada.

Más aún, recientes hallazgos basados en la población de personas residentes en California que postularon a la Universidad de California entre 1994 y 2011, muestran que los factores asociados al origen socioeconómico (ingreso familiar, educación de los padres y raza/origen étnico) representan una gran y creciente proporción de la varianza en las puntuaciones en el SAT de los estudiantes durante los últimos veinte años. Más de un tercio de la varianza en las puntuaciones en el SAT puede predecirse hoy mediante factores conocidos al momento del nacimiento de los estudiantes, un valor que solo llegaba a un cuarto de la varianza en 1994. Además, de esos factores la raza se ha convertido en el predictor más poderoso. Más que decaer en importancia, la raza y el origen étnico son más relevantes que el ingreso familiar o el nivel educacional de los padres para explicar las diferencias en las puntuaciones obtenidas en las pruebas.

«Efectos de señalización»

Otra característica de las pruebas de admisión que es igualmente importante pero más difícil de medir es su *efecto de señalización*, en otras palabras, el mensaje que las pruebas envían a los alumnos y a los profesores en las escuelas. Por ejemplo, los estudiantes suelen ver al SAT como una prueba de inteligencia, más que como una prueba sobre lo que han aprendido en la escuela. Esto también es cada vez más frecuente en el caso del ACT dada su restricción en tiempo y su énfasis en el recuerdo rápido del material. Se interpreta que una puntuación baja en cualquiera de las dos pruebas indica que un estudiante tiene una capacidad intelectual o un potencial insuficientes.

Por otra parte, las pruebas con referencia a criterios envían un mensaje muy diferente. Una baja puntuación en una prueba con referencia a criterios simplemente significa que el estudiante no ha aprendido el contenido incluido en la prueba. Esto puede deberse a una amplia variedad de factores, como un bajo nivel de enseñanza, falta de recursos instruccionales o incluso falta de trabajo duro por parte del estudiante. Las pruebas con referencia a criterios dirigen la atención hacia determinantes del desempeño que son alterables y que, por lo tanto, son más adecuados para estimular mejoras y reformas educacionales.

Crecimiento de las pruebas con referencia a criterios

La evaluación con referencia a criterios ha crecido enormemente en Estados Unidos durante las últimas dos décadas como parte de los esfuerzos para establecer estándares curriculares en las escuelas K-12 de cada estado. Las pruebas con referencia a criterios también se conocen como evaluaciones basadas en estándares, y ambos términos se usan indistintamente. La evaluación basada en estándares se asocia con los esfuerzos a nivel de cada estado para establecer estándares curriculares referidos a lo que se espera que aprendan los estudiantes en cada área de enseñanza.

La idea clave que subyace a la evaluación basada en estándares es el alineamiento. La meta es alinear la enseñanza, el aprendizaje y la evaluación estableciendo estándares claros de lo que se espera que los estudiantes aprendan, enseñando de acuerdo a los estándares y luego evaluando a los estudiantes con respecto a estos estándares. Las puntuaciones de los estudiantes reflejan su nivel de dominio en una asignatura dada, en lugar de compararlos con otros. El movimiento de los estándares ha sacudido la educación secundaria en EE.UU. durante los últimos 20 años, pero aún no logra establecerse en el nivel universitario, donde las pruebas con referencia a normas aún predominan. Las escuelas en Estados Unidos están mucho más avanzadas que los institutos y las universidades en este aspecto.

A continuación se muestra un ejemplo de los estándares curriculares del Estado de California, en este caso, de la asignatura de Biología. Los estándares definen lo que se espera que sepan los estudiantes sobre el código genético y la secuencia de aminoácidos codificados en el ADN y el ARN. Como evidencia este ejemplo, los estándares requieren un conocimiento profundo del tema.

Los genes son un conjunto de instrucciones codificadas en la secuencia de ADN de cada organismo, que especifican la secuencia de aminoácidos en las proteínas características de cada organismo.

Como base para comprender este concepto:

- a. *Los estudiantes conocen* la vía general mediante la cual los ribosomas sintetizan proteínas, empleando ARNt (ARN de transferencia) para traducir información genética presente en el ARNm (ARN mensajero).
- b. *Los estudiantes saben* cómo aplicar las reglas de codificación genética para predecir la secuencia de aminoácidos basándose en una secuencia de codones en el ARN.

Figura 8. Ejemplo de estándares de contenido para la asignatura de Biología/Ciencias de la Vida en el Estado de California. Fuente: Departamento de Educación de California.

A nivel universitario en EE.UU., los dos ejemplos más prominentes de pruebas basadas en estándares son las pruebas de asignaturas del SAT [SAT Subject Tests] y los Advanced Placement Exams. A diferencia del SAT o del ACT, estas pruebas son específicas para cada asignatura y miden el conocimiento de los estudiantes sobre biología, química, idiomas extranjeros, historia, física y otros temas.

Las pruebas de asignaturas del SAT se implementaron por primera vez en la década de 1930 y hoy se ofrecen para alrededor de 20 asignaturas. Son pruebas de opción múltiple y duran una hora. Las pruebas de asignaturas del SAT actualmente se exigen para ingresar principalmente a universidades privadas de primer nivel como Harvard.

Los exámenes AP se usan más ampliamente y fueron implementados por primera vez en 1955. Su propósito original era permitir que los estudiantes obtuvieran créditos en la universidad mientras asistían a la escuela secundaria. Los exámenes Advanced Placement (AP) duran de 2 a 3 horas e incluyen una combinación de preguntas de opción múltiple, respuesta libre y ensayo. Actualmente se ofrecen en más de 30 asignaturas. Aunque su meta original era entregar créditos universitarios, los exámenes AP se han convertido en una parte cada vez más importante en las admisiones universitarias, especialmente en las instituciones de primera línea.

A continuación se presenta un ítem de muestra perteneciente al examen AP de biología. Comparados con el SAT o el ACT, los exámenes AP son mucho más rigurosos y requieren un conocimiento más profundo de las materias.

Si un segmento de ADN es 5'-TAC GAT TAG-3', el ARN resultante de la transcripción de este segmento será:

- (A) 3'-TAC GAT TAU-5'
- (B) 3'-ATG CTA ATA-5'
- (C) 3'-UAC GAU UAG-5'
- (D) 3'-AAC GAU UAA-5'
- (E) 3'-AUG CUA AUC-5'

Respuesta: (E)

Figura 9. Ejemplo de ítem del examen AP de biología. Fuente: College Board.

Y, a continuación se presenta una distribución típica de las puntuaciones en un examen AP, siguiendo con el ejemplo de la biología. Esta claramente no es la distribución normal asociada con el SAT o el ACT.

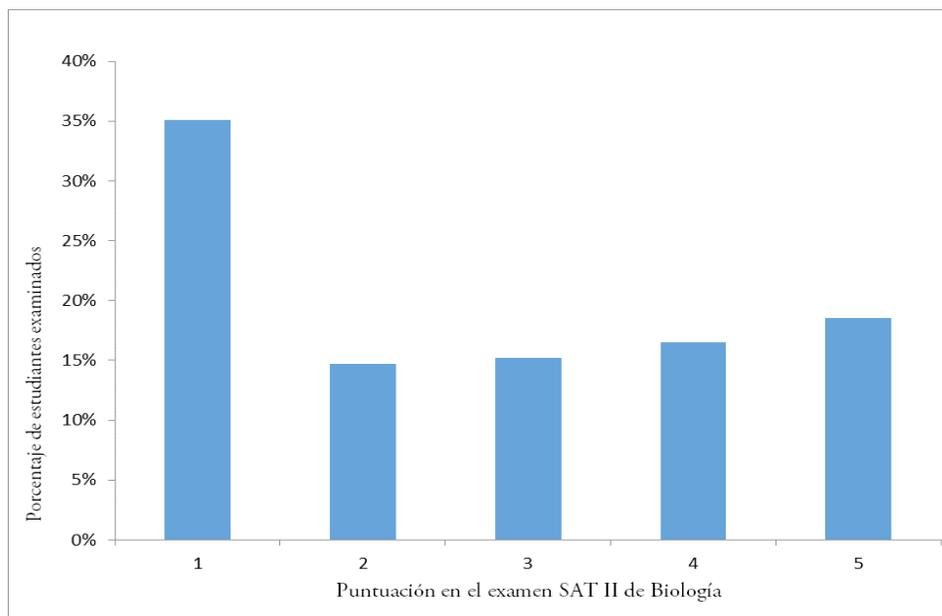


Figura 10. Distribución de las puntuaciones en el Examen AP de Biología. Fuente: College Board, 2010.

En la mayoría de las universidades, un estudiante necesita una puntuación AP de 4 o 5 para ser considerado calificado o diestro en una asignatura dada, aunque a veces las instituciones de primer nivel exigen un 5. Esto usualmente se denomina *puntuación de corte*. Pero una puntuación de corte en una prueba con referencia a criterios tiene un significado distinto a una puntuación de corte en una prueba con referencia a normas. En estas últimas, la puntuación de corte representa el percentil en el que se ubica un estudiante en comparación con otros examinados. En pruebas con referencia a criterios, la puntuación de corte representa el nivel de destreza en una asignatura dada.

Validez predictiva de las pruebas con referencia a criterios versus las pruebas con referencia a normas

La Universidad de California es una de las pocas instituciones en EE. UU. que ha usado tanto los Exámenes AP como las Pruebas de Asignaturas SAT, además del SAT y el ACT, por lo que posee una amplia base de datos para comparar los dos tipos de pruebas. Una de las mayores sorpresas que nos

entregaron nuestras investigaciones fue que las pruebas con referencia a criterios predicen los resultados en la universidad al menos igual de bien que las pruebas con referencia a normas. Aunque los Exámenes AP y las Pruebas de Asignaturas SAT no están diseñados para predecir, estuvieron entre los mejores predictores del rendimiento de los estudiantes en la universidad.

Variables <u>de predicción</u>	Peso <u>estandarizado</u>
HSGPA (Promedio de notas en la escuela secundaria)	,29
Exámenes AP	,14
Pruebas de asignaturas SAT	,09
Nivel educacional de los padres	,09
SAT (o equivalente en el ACT)	,06
Ingreso familiar	,03
<hr/>	
Varianza explicada	25,1%

Figura 11. Peso relativo de diferentes variables en la predicción de las calificaciones de estudiantes de primer año de la UC. Fuente: Geiser y Santelices, 2006.

Las calificaciones en la educación secundaria fueron el mejor predictor del desempeño en la universidad. El peso predictivo del promedio de notas en la escuela secundaria fue de ,29. Esto significa que por cada aumento de una desviación estándar en el promedio de notas de la enseñanza secundaria, el promedio de notas en la universidad aumentó ,29 desviaciones estándar, manteniéndose constantes todos los demás factores.

Los exámenes AP fueron el segundo mejor predictor, seguidos de las Pruebas de asignaturas del SAT. A continuación se ubicó la educación de los padres. Finalmente, las puntuaciones en el SAT y el ACT se ubicaron tras el nivel educacional de los padres y solo un poco por delante del ingreso familiar en la predicción de resultados en la universidad.

La principal lección que nos dan estos hallazgos no es que las pruebas con referencia a criterios son preferibles a las pruebas con referencia a normas porque son mejores predictores. De hecho, la capacidad de predecir el éxito de los estudiantes en la universidad es limitada. En su conjunto, todos estos predictores representaron únicamente un cuarto de la varianza de las calificaciones de los alumnos de primer año de la UC, con lo que quedan tres cuartos sin explicación. Las diferencias estadísticas entre las distintas pruebas son demasiado pequeñas para dictar la adopción de una prueba por sobre otra.

Más bien, la lección que debemos sacar es la siguiente: debido a que las diferencias estadísticas son tan pequeñas, las pruebas seleccionadas para la admisión universitaria deben basarse en consideraciones distintas de la predicción. Es aquí que las ventajas de las pruebas con referencia a criterios son más evidentes.

Características esenciales de las pruebas de admisión universitaria

Sobre la base de las investigaciones de la UC, las siguientes características son las más importantes que debieran exigirse a las pruebas de admisión:

- El foco principal no debiera ser cómo se compara un postulante con otros, sino más bien si ha logrado dominar el conocimiento y las habilidades básicas necesarias en la universidad.
- Las pruebas de admisión debieran exhibir tanto validez aparente como validez predictiva, de modo que el vínculo entre el conocimiento y las habilidades puestas a prueba y las necesarias para la universidad, sea transparente.
- Las pruebas de admisión debieran alinearse con el trabajo realizado en la educación secundaria y reforzar la enseñanza y el aprendizaje de un riguroso currículo en nuestras escuelas.

- Las pruebas de admisión debieran minimizar la necesidad de prepararse, premiando el conocimiento de los contenidos por sobre las habilidades para rendir pruebas, de modo tal que la mejor preparación para la prueba sea la instrucción normal en el aula.
- Finalmente, las pruebas de admisión debieran enviar una señal clara a los estudiantes: trabajar duro y dominar las asignaturas en la educación secundaria son el camino más directo a la universidad.

Las pruebas basadas en estándares o con referencia a criterios tienen ventajas educacionales significativas por sobre las pruebas con referencia a normas en todos estos aspectos.

Para resumir las conclusiones de la investigación en la Universidad de California, los hallazgos muestran que los criterios de admisión que miden el dominio de los contenidos curriculares (como las calificaciones en la educación secundaria y las pruebas basadas en estándares) predicen el éxito de los estudiantes en la universidad al menos igual de bien que las pruebas con referencia a normas como el SAT o el ACT. Estos indicadores (especialmente el promedio de calificaciones en la enseñanza secundaria [HSGPA]) también son más justos y más inclusivos para los postulantes de bajos ingresos y pertenecientes a minorías.

Otro punto igualmente importante es que estos criterios tienen un efecto de señalización más amplio y más beneficioso durante todo el proceso educativo. Ellos envían un potente mensaje a los estudiantes: trabajar duro y dominar las asignaturas académicas en la educación secundaria (no la capacidad innata) es la clave para ingresar a la universidad. De este modo, ayudan a alinear la enseñanza, el aprendizaje y la evaluación durante el camino entre la escuela secundaria y la universidad.

Postdata 2014: Lecciones aprendidas a raíz de la implementación de políticas

La sección anterior presentó las conclusiones principales de mis investigaciones hasta el año 2001. Esos hallazgos ayudaron a iniciar uno de los períodos más notables y sostenidos de reformas en las políticas de la Universidad de California en toda su historia. En primer lugar, las reformas incluyeron el Plan Top 4% de la UC, que garantizaba el ingreso a los mejores estudiantes de cada escuela secundaria de California. El ranking de los estudiantes en la escuela se determinaba únicamente por sus calificaciones, reduciendo así la importancia de las puntuaciones en las pruebas de admisión como criterio de ingreso. En segundo lugar, la UC introdujo la revisión holística, un proceso de admisión que toma en cuenta una amplia gama de factores como las desventajas socioeconómicas y las oportunidades de aprender. Esta reforma también fue diseñada para reducir el rol de las pruebas. En tercer lugar, la UC inició un amplio programa de vinculación con la comunidad en escuelas de bajo rendimiento en todo el estado de California, dirigido a mejorar la preparación para la universidad de los estudiantes pertenecientes a minorías y de bajos ingresos. Y, finalmente, con el liderazgo del Presidente de la UC Richard Atkinson, la universidad organizó una potente ofensiva para hacer frente a la prueba que entonces predominaba a nivel nacional: el SAT.

Algunas de esas reformas tuvieron éxito, mientras que otras fueron menos positivas. En la parte final de este artículo, y con la visión panorámica que es posible tener el año 2014, reflexionaré brevemente sobre cada una de las reformas y las lecciones que aprendimos durante la implementación de las políticas impulsadas.

El Plan Top 4%

La UC inició el Plan Top 4% en 2001. Esta política logró bastante éxito. Buscaba aumentar el ingreso de estudiantes provenientes de escuelas que históricamente habían enviado pocos estudiantes a la universidad. Como se muestra a continuación, tuvo precisamente ese efecto.

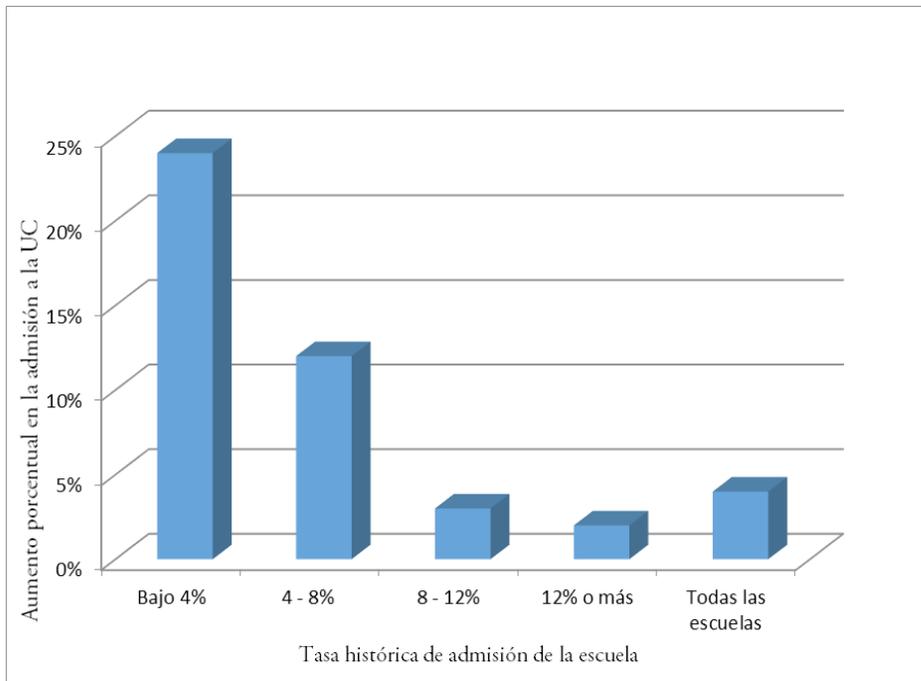


Figura 12. Impacto del Plan Top 4% en escuelas con tasas de ingreso a la UC históricamente bajas. Fuente: Universidad de California.

Otra fortaleza del Plan Top 4% es lo que podría llamarse la narrativa de sus políticas. La narrativa o historia usada para explicar una nueva política al público puede ser un factor importante en su éxito o fracaso final. En el caso del Plan Top 4%, la idea de recibir estudiantes de primer nivel de cada una de las escuelas de California, fue una combinación poderosa de valores tanto igualitarios como meritocráticos. Fue algo igualitario puesto que extendió la posibilidad de admisión a los estudiantes de todas las escuelas del estado. Pero también fue meritocrático ya que limitó esta posibilidad solo a los mejores estudiantes de cada escuela. Esta política resultó ser extremadamente popular por esta razón, lo que ilustra la importancia de tener una narrativa fuerte para implementar una política de manera efectiva.

La principal debilidad de esta política fue su escala limitada: generó sólo unas 2.500 nuevas admisiones porque la mayoría de los estudiantes dentro del 4% superior de sus escuelas ya podían ingresar según los criterios de admisión regular de la UC. Para abordar este problema, la UC expandió esta política desde el 4% superior al 9% superior de cada escuela secundaria en 2012.

Revisión holística

La revisión holística fue una segunda gran reforma. Su objetivo es evaluar a los postulantes considerando un conjunto de factores más amplio que solo las pruebas y las calificaciones. La idea clave es evaluar el logro en contexto, es decir, el grado en el cual los estudiantes han aprovechado sus oportunidades de aprender. Por ejemplo, que un estudiante de una escuela secundaria pobre alcance 1400 puntos en el SAT es un logro mayor que si un estudiante de una escuela de primer nivel obtiene la misma puntuación. Hasta 2001, la revisión holística solo se había usado en instituciones privadas altamente selectivas. Una de las primeras ocasiones en que se usó en una universidad pública con un gran volumen de postulantes fue cuando se adoptó en la UC.

La revisión holística ha causado mucha más controversia que el Plan Top 4%. Dentro de este enfoque, los revisores son entrenados para considerar todo el material incluido en el expediente de un postulante y luego asignarle una puntuación general. Es una tarea que requiere mucho trabajo y que, por lo tanto, es costosa. Necesita que los evaluadores hagan al menos dos revisiones de cada postulación, y a veces tres o más en caso de que las dos primeras difieran. Los funcionarios de admisión deben ser continuamente

entrenados y re-entrenados para asegurar resultados consistentes entre revisores, un proceso denominado *normativa*. Aparte del costo y del esfuerzo involucrado, a los críticos de la revisión holística no les gusta la falta de transparencia que esta introduce al proceso de admisión, y además sospechan de posibles sesgos favorables o desfavorables a diferentes grupos raciales.

Otra crítica se refiere a la metodología de evaluación. Debido a que el revisor combina los distintos factores de la admisión en su mente para producir una puntuación general, el peso asignado a los factores individuales puede variar entre un caso y el siguiente. Sin embargo, el enfoque holístico no es el único método para realizar una revisión integradora de los expedientes de los postulantes. Una alternativa es el *sistema de pesos fijos o de puntos* que se muestra a continuación, tomado de la UC en San Diego.

Factores académicos:	<u>puntos</u>	Factores adicionales:	<u>puntos</u>
Promedio de calificaciones en la educación secundaria	4.500	Escuela desfavorecida	300
Puntuaciones en pruebas	3.200	Ingreso familiar bajo	300
Clases adicionales	500	Primera generación en la universidad	300
Top 4%/Elegibilidad en el Contexto Local (ELC)	300	Liderazgo demostrado	300
		Talentos especiales	300
		Servicio a la comunidad	300
		Participante en programas de vinculación con la comunidad de la UC	300
		Circunstancias especiales	300

Figura 13. Una alternativa a la revisión holística: sistemas de «peso fijo». Fuente: UC en San Diego.

Los sistemas de puntos toman en cuenta los mismos factores que considera una revisión holística. Sin embargo, cada factor se evalúa por separado y tiene un peso fijo; posteriormente, las evaluaciones se suman para producir una puntuación general. Existe una masa considerable de investigaciones que indican que un enfoque de peso fijo es más confiable y transparente, así como también menos costoso que la revisión holística (Grove, 2005; Meehl, 1954).

En el análisis final, la pregunta clave es si alguno de los dos tipos de revisión integradora (holística o de peso fijo) produce decisiones de admisión diferentes y mejores. Un primer estudio piloto en Berkeley reveló que la revisión integradora cambió el resultado sólo de un 5 a un 6% de todas las decisiones de admisión. Esto ha tenido un impacto medible en la diversidad del cuerpo de estudiantes, pero ese efecto es relativamente limitado y probablemente permanezca así mientras las calificaciones y las puntuaciones en pruebas sigan siendo los criterios primarios para la admisión universitaria.

El ascenso y la caída de los programas de vinculación con la comunidad de la UC [Outreach]

La tercera gran iniciativa de la UC fue una considerable expansión de los programas de vinculación con la comunidad dirigidos a estudiantes y escuelas desfavorecidas de California desde el año 2001. Para 2003, los programas de vinculación con la comunidad de la UC entregaban servicios a cerca de 300.000 estudiantes y a 70.000 profesores y directores. La universidad había firmado acuerdos con 300 de las escuelas de peor rendimiento en el estado, ubicadas principalmente en zonas urbanas pobres cercanas a cada uno de los campus de la UC.

La expansión de los programas universitarios de vinculación con la comunidad nació del hecho de reconocer que las reformas de los procesos de admisión, en sí mismas, solo podrían tener un efecto marginal en el incremento del acceso para los estudiantes pertenecientes a minorías y de bajos ingresos. Permitir el ingreso de estos grupos seguiría teniendo un efecto limitado mientras estos estudiantes continuaran concentrados en las escuelas de peor rendimiento de California. Se necesitaba un esfuerzo enorme para mejorar la enseñanza y el aprendizaje en esas escuelas.

La UC llevaba largo tiempo entregando tutorías y programas de enriquecimiento académico individuales para estudiantes en los niveles K-12. Lo que era nuevo era el esfuerzo para intervenir dentro de las escuelas mismas. La asamblea legislativa del estado entregó financiamiento sustancial que permitió que cada campus de la UC formara sociedades con escuelas de bajo rendimiento en su región. La UC comenzó a ofrecer programas de desarrollo profesional para los profesores y directores de esas escuelas. La UC también ofreció programas sobre asignaturas específicas con el fin de actualizar los currículos. Se daba por entendido que este sería necesariamente un proyecto a largo plazo. Algunos de los socios de la UC eran escuelas primarias e intermedias (5° a 8° grado), por lo que pasaría algún tiempo antes de que se observaran los resultados.

Pero los plazos de la asamblea eran mucho menores. Ellos querían un aumento inmediato de los ingresos a la UC de estudiantes pertenecientes a minorías. Cuando esto no ocurrió, se retiró el financiamiento y los programas de acercamiento a la comunidad centrados en las escuelas se detuvieron casi completamente hacia 2005. Mirando hacia atrás una década más tarde, los programas de vinculación con la comunidad de la UC deberían verse no tanto como el fracaso de una política, sino más bien como un experimento sin terminar. La pregunta clave sigue sin respuesta. ¿Es posible que las universidades intervengan para mejorar la enseñanza y el aprendizaje en las escuelas más pobres?

Desafío a las pruebas nacionales de admisión

La última gran iniciativa de la UC en el campo de las políticas de admisión fue su desafío a la principal prueba de ingreso a las universidades en ese momento, el SAT. (Desde entonces, el SAT ha sido sobrepasado por el ACT como la prueba más rendida en EE.UU.). Con respecto a esa iniciativa, el veredicto final es ambiguo. Los intentos de cambiar las pruebas resultaron ser parcialmente exitosos en varios aspectos. En 2005, tanto el SAT como el ACT añadieron una prueba de escritura en respuesta a los hallazgos de varios investigadores, y también míos, que indicaban que, entre todas las pruebas sobre asignaturas escolares, la de escritura estaba dentro de los mejores predictores del rendimiento en la universidad. El SAT eliminó las analogías verbales y las comparaciones cuantitativas, las que habíamos criticado por ser un remanente de la tradición de medición del CI de esa prueba. Y, en general, el Consejo de Universidades (College Board) ha tomado medidas para acercar el contenido del SAT al currículo de la educación secundaria. Por ejemplo, las comparaciones cuantitativas (otro tipo de ítem complicado que normalmente no se enseña en las escuelas) también fueron eliminadas en 2005 y reemplazadas con ítems más directos que cubren la matemática que se enseña en segundo y tercer año. Todos estos cambios son pasos en la dirección correcta. Y tanto el SAT como el ACT recientemente anunciaron cambios que desplazarán aún más esas pruebas en la misma dirección.

Sin embargo, el problema es que ambas pruebas siguen siendo evaluaciones con referencia a normas. Están fundamentalmente fuera de sincronización, en ese sentido, con respecto al movimiento basado en estándares del resto de la educación estadounidense durante las últimas dos décadas. Probablemente sea inevitable que las pruebas de admisión de EE.UU. evolucionen hasta convertirse en evaluaciones basadas en estándares, aunque lo más seguro es que esto tome algún tiempo. Hasta entonces, el SAT y el ACT seguirán alimentando la obsesión con las pruebas de alto impacto (y a toda la industria de preparación para las pruebas) en la admisión a las universidades estadounidenses. La importancia de pequeñas diferencias entre puntuaciones en pruebas que tienen poca relación con los resultados en la universidad seguirá exagerándose.

Pero existe una alternativa clara. En 2008, la Asociación Nacional para la Orientación en la Admisión a las Universidades [National Association for College Admissions Counseling] organizó una comisión de alto nivel dedicada a las pruebas de admisión en Estados Unidos. Dirigida por William Fitzsimmons, decano de admisión en Harvard, la comisión llegó a una conclusión muy similar a la presentada en este artículo:

Al emplear el SAT y el ACT como una de las herramientas de admisión más importantes, muchas instituciones están obteniendo lo que podría ser una capacidad marginal de identificar talento académico más allá del que indican las calificaciones escolares, las recomendaciones y las puntuaciones en pruebas de logro. Por el contrario, el uso de [...] Pruebas de Asignaturas del Consejo de Universidades y pruebas AP, o exámenes del Bachillerato Internacional, crearía un poderoso incentivo para que las escuelas secundarias estadounidenses mejoraran sus

currículos y su enseñanza. Las universidades perderían poco o nada de la información que necesitan para tomar buenas decisiones sobre el ingreso a las clases, al tiempo que beneficiarían a millones de estudiantes estadounidenses que no ingresan a universidades selectivas y tendrían un impacto positivo sobre la enseñanza y el aprendizaje en las escuelas del país (National Association for College Admission Counseling, 2008).

Ciertamente, las pruebas con referencia a currículos o a estándares no son la panacea que logrará hacer que la admisión universitaria se vuelva completamente justa y racional. Las desigualdades en nuestras escuelas secundarias son demasiado profundas y generalizadas como para que eso sea posible. Sin embargo, las pruebas basadas en estándares pueden ayudar a hacer el proceso más justo y más racional de lo que es actualmente. En primer lugar, a nivel universitario, pueden ayudar a nivelar el campo de juego para quienes no pueden pagar costosos servicios de preparación para pruebas: incluso en las escuelas más pobres, los estudiantes que trabajan duro y tienen un buen desempeño en sus estudios escolares normales, podrían mejorar sus probabilidades de admisión.

Y, en segundo lugar, a nivel de las escuelas, las pruebas con referencia a criterios pueden reforzar la enseñanza y el aprendizaje de un riguroso currículo preparatorio para la universidad. El refuerzo que entregan puede ser incluso más importante en las escuelas de bajo rendimiento que en las demás. La experiencia en la implementación de estándares curriculares estatales en EE.UU. sugiere que la estrategia de establecer estándares de contenidos claros, enseñar teniendo en cuenta los estándares y evaluar a los estudiantes con respecto a esos estándares, podría producir los mayores beneficios en las escuelas más desfavorecidas. Al adoptar evaluaciones basadas en estándares, nuestras instituciones de educación superior selectivas pueden tener un efecto más amplio y más beneficioso en todo el campo educacional.

Discusión

La experiencia en la UC puede ser importante para otras instituciones selectivas, tanto en EE.UU. como en el extranjero, que dependan en gran medida de puntuaciones en pruebas estandarizadas como parte de sus procesos de admisión. Las pruebas con referencia a criterios o basadas en estándares tienen ventajas claras por sobre las pruebas con referencia a normas para propósitos de admisión universitaria, y desde un punto de vista más general, para el mejoramiento escolar. Además, los hallazgos sobre errores de predicción, la relación entre puntuaciones en las pruebas y el nivel socioeconómico y los efectos de señalización de los indicadores de admisión sugieren que las puntuaciones en pruebas estandarizadas debieran usarse solo como uno de muchos criterios de admisión. Las puntuaciones en pruebas debieran complementarse con otras fuentes de información sobre perfiles de los estudiantes, de modo similar a la revisión holística implementada por la UC. Los programas de vinculación con la comunidad siguen siendo una alternativa potencial para aumentar la diversidad del cuerpo de estudiantes de la institución, pero deben entenderse como una estrategia a largo plazo que requiere financiamiento y compromiso institucional constante. La experiencia de la UC al implementar el Plan Top 4% sugiere que las admisiones basadas en el ranking en la educación secundaria tendría un alcance considerable para diversificar las admisiones a la educación superior sin descensos significativos en las tasas de graduación universitaria. Deben considerarse nuevas y diferentes opciones en el ámbito de las políticas de admisión si la meta es aumentar la representación de los grupos minoritarios en la educación superior, especialmente en instituciones altamente selectivas.

El artículo original fue recibido el 15 de marzo de 2015

El artículo revisado fue recibido el 21 de enero de 2016

El artículo fue aceptado el 14 de marzo de 2016

Referencias

- Geiser, S. (2009). Back to the basics: In defense of achievement (and achievement tests) in college admissions. *Change: The Magazine of Higher Learning*, 41(1), 16-23.
- Geiser, S. (2014). *Back to the future: Freshman admissions at the University of California, 1994 to the present and beyond*. Research and Occasional Paper Series: CSHE.4.14. Berkeley, CA: Center for Studies in Higher Education, University of California, Berkeley. Recuperado de <http://www.cshe.berkeley.edu/publications/back-future-freshman-admissions-university-california-1994-present-and-beyond>
- Geiser, S. (2015). *The growing correlation between race and Sat Scores: New findings from California*. Research and Occasional Paper Series No. 10.15. Berkeley, CA: Center for Studies in Higher Education, University of California, Berkeley. Recuperado de <http://www.cshe.berkeley.edu/publications/growing-correlation-between-race-and-sat-scores-new-findings-california-saul-geiser>
- Geiser, S., & Atkinson, R. (2013). Beyond the master plan: The case for restructuring baccalaureate education in California. *California Journal of Politics and Policy*, 4(1), 67-123.
- Geiser, S., & Caspary, K. (2005). 'No show' study: College destinations of University of California applicants and admits who did not enroll, 1997-2002. *Journal of Educational Policy*, 19(2), 396-417.
- Geiser, S., & Santelices, M. V. (2006). The role of Advanced Placement and honors courses in college admissions. En P. Gandara, G. Orfield, & C. Horn (Eds.) *Expanding opportunity in higher education: Leveraging promise* (pp. 75-114). Albany, NY: State University of New York Press.
- Geiser, S., & Santelices, M. V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes*. Research and Occasional Paper Series #CSHE 9.07. Berkeley, CA: Center for Studies in Higher Education, University of California, Berkeley.
- Geiser, S., & Studley, R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, 8(1), 1-26.
- Grove, W. M. (2005). Clinical versus statistical prediction: The contribution of Paul E. Meehl. *Journal of Clinical Psychology*, 61(10), 1233-1243.
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota.
- National Association for College Admissions Counseling (2008). *Report of the commission on the use of standardized tests in undergraduate admissions*. Arlington, VA: Autor.