

## Contexto De La Evaluación El Día De Hoy: Historia, Impacto y Alternativas

Mientras que casi todas las personas externas a un salón de clase celebran las ventajas de la evaluación, estas no ven todo lo que se pierde en este proceso. Cuando en este artículo se menciona la palabra “evaluación” se hace referencia a la imposición que se le realiza a los estudiantes de realizar exámenes, que son diseñados y calificados por individuos, agencia estatales, o compañías que se encuentran por fuera del salón de clase, y sus resultados son utilizados para medir el desempeño de los profesores, estudiantes y escuelas.<sup>1</sup>

¿Por qué los exámenes a través de la historia y posiblemente más en la actualidad, llegan a definir nuestro, conocimientos y nuestra realidad?<sup>2</sup>

En vista que los exámenes estandarizados han llegado a dominar la reforma educativa en Chicago y a lo largo de todo el país y como los resultados de estos exámenes han llegado a definir el éxito y el fracaso,<sup>3</sup> es importante explorar diferentes malinterpretaciones que son aspectos claves para entender y dar respuesta a la pregunta realizada anteriormente.

### Las malinterpretaciones entre otras son las siguientes:

- *La evaluación usando exámenes es la forma en que siempre hemos medido el aprendizaje.*
- *El desempeño en un examen es una forma certera de medir el aprendizaje del individuo.*
- *la aplicación de exámenes es la mejor forma de asegurar que los maestros están enseñando correctamente.*

- *Los high-stakes testing Los exámenes con consecuencias de altos riesgo (los resultados de estos tipos de exámenes estandarizados se usan para tomar decisiones con altos riesgos como por ejemplo, retener a estudiantes en el mismo grado, cerrar escuelas, despedir a maestros o directores, perder fondos, etc...) son un mal necesario.*
- *Los exámenes están aquí para quedarse y no hay nada que podamos hacer.*

Acompañado de cada malinterpretación viene una suposición que necesita ser entendida y cuestionada para lograr alcanzar un diálogo significativo y sustantivo con el público acerca de la reforma educativa.

En esta Dirección CReAte (Investigadores y Defensores de una Educación transformadora en el área de Chicago) ha desarrollado el resumen de esta política, la cual utiliza investigaciones en el campo de la educación para desmitificar como las evaluaciones particularmente los exámenes con consecuencias de altos riesgos han llegado a ser utilizados como una herramienta de evaluación sino como una herramienta de enseñanza y aprendizaje. La forma en cómo los exámenes están siendo utilizados actualmente resulta inclusive preocupante para aquellos que los diseñan.<sup>4</sup> Este resumen dará una descripción acerca de la historia de la forma actual de abordar la evaluación educativa, la razón por la cual inclusive hasta los expertos en educación se encuentran preocupados, el impacto de este singular enfoque de cómo la evaluación ha sido definida y practicada, y la diferente variedad de alternativas de evaluación que existen, pero que han sido ignoradas en los debates actuales.

## Una Breve Historia de la Evaluación

**Malinterpretación:** *la evaluación usando exámenes es la forma en que siempre hemos medido el aprendizaje.*

### Las evidencias:

- *La era moderna de la aplicación de exámenes comenzó a finales del siglo pasado.*
- *La aplicación de exámenes está basada en la era de la fe en la ciencia y la necesidad de la industria en esa época.*
- *Los exámenes fueron diseñados para clasificar y diferenciar en vez de medir el aprendizaje.*
- *Los exámenes están siendo usados como evidencia del fracaso de las escuelas.*

Para entender el papel de la aplicación de *exámenes* en la actualidad, necesitamos mirar hacia el pasado y examinar los factores sociales, políticos y económicos que influenciaron la reforma educativa en el siglo pasado. Podemos comenzar realizando una pregunta básica: ¿por qué creemos en el valor de evaluar y en el significado de los resultados de los exámenes como base para la toma de decisiones en la educación? Algunos adultos podrían responder que “funcionó para mí” o que “esto es lo que siempre hemos hecho,” pero la historia nos dice que los exámenes se generaron en formas particulares y con objetivos particulares.

A lo largo del siglo pasado, por lo menos dos eventos jugaron un papel importante en el hecho de promover la importancia de la evaluación, uno tomó lugar a inicios del siglo 20 y la otra emergió hacia el final del mismo siglo.

La revolución industrial y científica de finales del siglo 19 y principios del siglo 20 promovió una fe en el método científico, que cuando era aplicado al campo de la educación, resultaba en una admiración de lo que el Investigador Peter Taubman<sup>5</sup> llama la “ciencia del aprendizaje.” Este cambio de paradigma fue el producto de los éxitos alcanzados por las ciencias naturales en la medida y cuantificación de datos para obtener verdades objetivas, lo cual se extendió más adelante a las ciencias sociales.<sup>6</sup> Uno de los aspectos claves del método científico era el de quitar el científico de la ciencia, para reforzar de esa manera la objetividad, bien fuera que el objetivo de la investigación era estudiar el mundo natural, las personas o la sociedad.

Uno de los primeros y más significativos desarrollos en esta área fue el nacimiento de la evaluación de la inteligencia en Francia a comienzos del siglo pasado. Desarrollado por Alfred Binet como un instrumento para identificar estudiantes cuyas capacidades cognitivas eran muy limitadas para beneficiarse de la corriente principal de la escuela, su creación evolucionó en la Escala de Medición de la Inteligencia de Stanford-Binet<sup>7</sup>, un tipo de prueba que reclamaba ser una medida científica de la inteligencia

innata. Edward Thorndike en la misma área emergente de Binet de la psicología experimental, expandió la idea de la evaluación mental en la educación y desarrolló el examen de logros estandarizados, una medición de la inteligencia innata, pero de lo que un estudiante ha aprendido. Estos exámenes consistían de un gran número de preguntas con respuestas de selección múltiple, que resultaban en un único resultado numérico que podía ser comparado con el resultado de otros que realizaran el examen. Aquí es donde comienza a lo que Taubman<sup>8</sup> se refiere como la “fe en la ciencia de la evaluación” en la educación.

Mientras que los exámenes estandarizados de logros no reemplazan de forma inmediata otras formas de evaluación (incluyendo los exámenes y otras formas de evaluación desarrollada por los profesores para sus cursos y estudiantes). Este fue uno de los primeros ejemplos de una norma de evaluación referenciada que rápidamente se convertiría en una práctica estándar en las escuelas a lo largo de los Estados Unidos. Una idea central de tales exámenes es la comparación de un estudiante con el resto de estudiantes que realizaron el examen. Esta norma hace referencia a dar el mismo examen a un gran número de estudiantes en diferentes clases y escuelas a través del estado y de todo el país. El interés en comparar estudiantes se aumenta en la medida que la población se va diversificando debido al incremento de la inmigración, la urbanización, la integración geográfica de diferentes regiones y la legislación para hacer la asistencia a las escuelas, libre y obligatoria. Los exámenes estandarizados de de norma referenciada, insertaron presumiblemente una herramienta objetiva para diferenciar y medir los estudiantes. A partir de este enfoque se asumen dos aspectos muy importantes: La enseñanza y el aprendizaje pueden (y deberían) ser estandarizados; y los exámenes estandarizados son una medida precisa y certera de lo que se ha aprendido (y enseñado). En el momento en que la academia estaba emergiendo, la validez de estos dos aspectos era cuestionada, pero la práctica continuaba esparciéndose.<sup>9</sup>

Iniciando en los años 1970 y 1980 y hasta el día de hoy, un segundo paradigma aparece encerrando la evaluación estandarizada, en particular los exámenes con consecuencias de alto riesgo, como una panacea educativa. Un catalizador inicial fue la Guerra Fría y la preocupación constante de los Estados Unidos de mantener su prominente posición en el mundo como un “super poder”, económica, científica y políticamente hablando. Una crisis comenzaba a aparecer, las escuelas del país estaban fallando y la nación se encontraba en riesgo de declinar, tal como lo dice el informe de 1983 “Una Nación en Riesgo” de la administración Reagan. De acuerdo a esta lógica las escuelas necesitaban rendir cuentas, y la fuente básica para medir ese progreso, debía ser la más objetiva de las evaluaciones: Los exámenes estandarizados de norma referenciada, podrían dar información confiable para la promoción de estudiantes, la retención de maestros, autonomía de las escuelas y financiación de las escuelas. Este énfasis en la estandarización, rendición de cuentas y los exámenes con consecuencias de alto riesgo, continuaron permeando la visión de la educación de

siguientes administraciones, desde los Objetivos 2000 de Clinton, No dejar ningún niño(a) atrás de Bush, hasta la carrera hasta la cima de Obama, a pesar que la ciencia de la evaluación, particularmente desde adentro del campo de la psicometría, no apoya dichas aseveraciones acerca de lo que estos exámenes nos dicen o como los resultados de dichos exámenes deberían ser utilizados. El principio fundamental y proponentes dentro del campo de la psicometría nos cuentan una historia diferente.<sup>10</sup>

## Psicometría 101

**Malinterpretación:** *El desempeño en un examen es una forma certera de medir el aprendizaje del individuo.*

### Las evidencias:

- *Los exámenes diseñados para comparar el desempeño de los estudiantes están siendo utilizados de forma incorrecta para medir el aprendizaje.*
- *Todos los resultados de los exámenes contienen errores.*
- *La validez de los exámenes o que tan correctamente miden lo que supuestamente deberían medir varía.*
- *Ninguna decisión importante debe ser tomada con base en el resultado de un solo examen o prueba.*

El primer asunto a tener en cuenta acerca de la evaluación es el propósito. Quienes desarrollan los exámenes y quienes los usan deben estar totalmente claros acerca de que se les está midiendo y de cómo los resultados serán utilizados en la toma de decisiones. El proceso de construcción de un examen es muy especializado de forma tal que un instrumento diseñado para un propósito no puede ser utilizado para medir otro.

Por ejemplo como se observó anteriormente en este resumen, las normas de referencia, trae consigo comparar el desempeño de una persona con el resto de los que presentaron el examen. La norma de referencia sirve para los propósitos de ciertas mediciones de manera correcta. Algunos podrían discutir que los exámenes de admisión a la universidad, son algunos de ellos, ya que según estos se determina que estudiantes merecen ingresar o no. Los resultados de norma referencia son a menudo reportados como un percentil, donde ubicarse en el percentil 87 significa que el 87% de quienes tomaron este examen recibieron a un puntaje inferior (un buen resultado), y un décimo percentil significa que solamente un 10% de quienes realizaron el examen obtuvieron un peor desempeño (se muestra una debilidad).

Aspectos individuales en exámenes en los cuales sus resultados son referenciados por una norma deben ser escritos y monitoreados cuidadosamente. Una razón por la cual estas preguntas son analizadas es para poder asegurar que aproximadamente la mitad de quienes realizan el examen las responde correctamente y la otra mitad las responde de forma incorrecta. Las preguntas que muchas personas fallan o que muchas personas contestan

correctamente, no contribuyen al análisis de los resultados que permitirá la clasificación del desempeño.

Cuando el propósito no es clasificar, como cuando se pretende determinar el grado en el cual los estándares son alcanzados, no es difícil ver porque la norma de referencia es altamente inapropiada. Las mediciones de referencia estándar deberían estar midiendo el desempeño de un estudiante en relación a estándares de currículo, y no en relación al desempeño de otros estudiantes. Utilizar puntajes de norma de referencia para estos propósitos podría ser matemáticamente imposible para que todos los estudiantes alcancen el éxito, y para que una gran cantidad de estudiantes de nuestros estudiantes pueda ser retenido en el umbral de la evaluación. Después de todo el 50% de los estudiantes será siempre calificado por debajo del percentil 50, aún cuando todos los estudiantes alcancen un dominio sin precedentes del material.

Con la intención de direccionar estas equivocaciones a propósito, algunos exámenes de gran escala, tales como el utilizado para reportar el Progreso Adecuado del Año (AYP por sus siglas en Inglés) por las escuelas elementales de Illinois (el ISAT), también reporta el desempeño en relación al material con un sistema de puntaje llamado “referencia de criterio.” Sin embargo, el hecho de que el examen fue desarrollado para ser norma de referencia significa que cada aspecto del examen esta ajustado para obtener la mitad de las respuestas correctas y la otra mitad incorrecta, e ítems que con cuentan que miden de mejor manera los estándares fueron rechazados. Algunos académicos postulan que la variación de las respuestas en exámenes sin norma de referencia es solamente alcanzada por la adquisición de conocimiento, que no es parte de la información estándar la cual los estudiantes ganan a través de su experiencia por fuera de la escuela y que además no es aprendido en el salón de clase, sin importar que duro el maestro o los estudiantes trabajen.<sup>11</sup>

Otro principio fundamental en la psicometría es la fiabilidad la cual es definida más simplemente como “la consistencia con la cual un examen mide lo que esta midiendo.”<sup>12</sup> Los desarrolladores de exámenes trabajan muy fuerte para asegurar que sus productos sean confiables. Sin embargo, esta es aspecto no tiene discusión ya que no hay un examen que sea completamente confiable. Dicho en otras palabras, “algún grado de inconsistencia está presente en todos los procedimientos de medición.”<sup>13</sup> Los profesionales de la psicometría conceptualizan esto como la diferencia entre el resultado verdadero de quien presenta el examen (una reflexión perfectamente precisa de la capacidad, necesariamente hipotética dadas las limitaciones de la evaluación) y los resultados inmersos en el examen. La diferencia es llamada “error de medición”, y es un componente de *cada examen*. Este es el porqué los resultados son acompañados por una clasificación representando los errores estándar de medición, o un estimado de cuanto el resultado de una persona que elabore un examen puede variar. Una decisión podría ser psicométricamente muy injusta y poco representativa de la verdad – a no ser que de que esta fuera sustentada por un puntaje en una clasificación o rango.

Aún más importante que la confiabilidad para un examen es su validez, o que tan bien mide lo que este fue diseñado para medir. De acuerdo a Popham este es “uno de los conceptos más importantes en la evaluación.”<sup>14</sup> Una analogía tomada de Thorndike,<sup>15</sup> puede ayudar a ilustrar la diferencia entre la confiabilidad y la validez, y explica porque el último es más importante. Imagínese que el instrumento es cuestión es una pesa de baño. Si usted se pesara varias veces seguidas, usted encontraría una variación en los resultados cada vez, usted inmediatamente concluiría que la pesa es inconsistente (poco confiable) y por ende usted diría que se encuentra dañada (no válida). Si, sin embargo el resultado de la pesa fuera consistentemente 5 libras por debajo o 5 libras por encima de su peso actual, el instrumento tendría confiabilidad pero no validez. La pesa continúa dañada, aún si el descubrimiento de este hecho viene acompañado de una comparación con otras pesas (lo cual generaría una sorpresa agradable o desagradable). Un examen que es inválido puede ser confiable (aunque continúe sin ser útil para su propósito), pero un examen que no es confiable nunca podrá ser válido.

La confiabilidad es solo uno de muchos factores que acompañan la validez de un examen. Tradicionalmente los expertos en evaluación, piensan acerca de la validez en 3 formas diferentes (técnicamente, tres tipos de evidencia): contenido, criterio, y construcción. El contenido está asociado con que también los exámenes reflejan el contenido curricular. En otras palabras, el examen concuerda con los estándares del curso y sus objetivos? El criterio trae consigo una comparación de un resultado de un examen con los resultados de alguna otra medición de la misma habilidad. La evidencia del criterio positivo relacionado, mostraría como una fuerte correlación entre el resultado del examen y los resultados de la otra medición, la cual se llama criterio.

Finalmente, construir la validez es una preocupación asociada a que tan bien el examen mide los “intereses de la construcción.” Una construcción es una idea, o algo que no es concreto. Instrumentos psicométricos, tales como medidas mentales, son por definición exámenes de construcción. Aquí nos hacemos la pregunta de si todas formas es este un buen examen. Los tres tipos de validez enunciados anteriormente, así como otros factores tienen influencia en llegar a una conclusión. Este es el mayor reto de validez que debe ser conceptualizado, pero además el más exhaustivo y significativo.

Debido a que los expertos en medición entienden las limitaciones de los exámenes que han sido descritos, hay un acuerdo universal de que deben ser usados con cuidado. Ninguna decisión importante debe ser basada en el resultado de un solo examen. Este principio es expuesto por quienes desarrollan estos exámenes, cuerpos académicos y asociaciones profesionales como el Concejo Nacional de Investigación, y el Concejo Nacional de Evaluación en la Educación.<sup>16</sup> El hecho de anexar toma de decisiones a los exámenes rompe uno de los principales principios de la psicometría.

Actualmente, la mayoría de los expertos en evaluación,

continúan teniendo los exámenes bajo estrictos y altos estándares de propósito, confiabilidad, validez y uso limitado. Desafortunadamente algunos usuarios no lo hacen. Depende de nosotros que estas instituciones rindan cuentas a los estándares de la evaluación.

## El Impacto de la Evaluación en el Salón de Clases y las Escuelas

**Malinterpretación:** *la aplicación de exámenes es la mejor forma de asegurar que los maestros están enseñando correctamente.*

### Las evidencias:

- *el desarrollo y preparación de exámenes han hecho que el currículo se haya reducido.*
- *La presión de mejorar los resultados de los exámenes ha hecho que algunos maestros abandonen la profesión.*
- *Los resultados asociados con la lectura no han crecido en Chicago, y la brecha de logros ha aumentado entre los grupos raciales (Latinos y Afro Americanos en comparación con blancos no-Latinos).*
- *Las escuelas con los mejores resultados en los exámenes, son aquellas que poseen las tasas más bajas de estudiantes que viven en situación de pobreza.*

La evaluación en las escuelas no es un fenómeno nuevo. Sin embargo, hasta mediados de los años 90, las evaluaciones estandarizadas que fueron realizadas por estudiantes en Chicago (y en la mayoría de los Estados Unidos) ni implicaron muchas decisiones. Eso cambió en 1995, cuando al entonces Gobernador Richard Daley se le dio control del distrito de escuelas por la legislación del estado de Illinois. Daley rápidamente marco el comienzo de un nuevo CEO de las Escuelas Públicas de Chicago (CPS por sus siglas en Inglés) – su anterior director de presupuesto, Paul Vallas – junto con un mandato de llevar una estricta “rendición de cuentas” a las escuelas y puntajes más altos en los exámenes en los distritos más grandes del estado.<sup>17</sup> Aparentemente de la noche a la mañana exámenes estandarizados fueron tenidos como la nueva medida de progreso para los estudiantes y escuelas en la ciudad y todo se trasladó hacia una nueva estructura, el currículo prescriptivo estaba en camino.

En los años siguientes el modelo de los exámenes con consecuencias de alto riesgo en Chicago se vería reflejado en el programa de George W. Bush de No dejar ningún niño(a) atrás (NCLB por sus siglas en Inglés), Escuelas a lo largo de todos los Estados Unidos, comenzarían a sentir sus impactos. La retórica de NCLB sonó prometedora para muchos: un enfoque nacional renovado de esfuerzos para encontrar estudiantes con bajo rendimiento, un

llamado para todas las escuelas – inclusive para aquellas conocidas por su excelencia académica-para examinar que tan bien ellos encuentran las necesidades del niño. Pero para los maestros, fue imposible ignorar el daño colateral de la visión estrecha. En muchas escuelas los exámenes eran ahora el currículo, y el impacto tanto en estudiantes como es profesores era palpable. Para los estudiantes, esto significaba un currículo que se había reducido a sólo un poco más de lectura y matemáticas, y con más tiempo dedicado para la preparación de exámenes. Para los profesores este significaba una reducción de la autonomía, un aumento de los niveles de estrés al igual que un decrecimiento en la satisfacción de su trabajo.

Por ejemplo el siguiente correo electrónico, escrito en 2006 por un participante de un programa de certificación de profesores para aquellos que desean cambiar de carrera, ofrece una idea del clima de presión en las escuelas de Chicago, en gran parte, por el enfoque de elevar los resultados de los exámenes. El autor de este correo electrónico había llegado a ser un profesor con grandes expectativas y aparentemente lleno de energía, estaba determinado a crear una experiencia significativa para sus estudiantes y de permanecer en el salón de clase por un largo periodo. Pero sólo tres años después de comenzar a enseñar su primera clase en el CPS, ella iba abandonando todo eso.

Este es un mensaje difícil de escribir. Después de meses de sopesar las opciones y de luchar con la complejidad de todo esto, he decidido renunciar de las *Escuelas Públicas de Chicago*. Cuando llegué a la enseñanza, no era ingenuo a los desafíos en el CPS, comprendí que la aventura que iba a emprender iba a ser difícil. [Pero] he luchado intensamente durante este año para justificar el enfoque de los resultados en las pruebas, en lugar del aprendizaje de los estudiantes. ... Créanme, esto no fue una decisión fácil ... pero el año continuó, y se acercó la temporada de pruebas, y me di cuenta que *mis prioridades* son con la administración del CPS. Mi personalidad no me dejará solo cerrar mi puerta y enseñar.<sup>18</sup>

Otro maestro que dejó el CPS en el 2010 y más tarde tomó por un largo plazo un *puesto de sustituto en una escuela suburbana* se sorprendió por el ambiente diferente *que encontró allí*:

Estoy impresionado por las diferencias significativas en las dos experiencias. La principal es la falta de conversaciones, publicidad, desarrollo profesional, e incluso de notas semanales mencionando *las pruebas estandarizadas*. Un completo extraño me dijo *en mi primer día [en la escuela de los suburbios] que yo era de confianza, que yo era un profesional*. Casi lloré. Los exámenes, especialmente en la ciudad, me han hecho ver que no soy de confianza o que no tengo valor.<sup>19</sup>

Muchos maestros que han enseñado o trabajado a lo largo de la última década han cuestionado el *énfasis excesivo en los exámenes*. Algunos vieron como sus escuelas no alcanzan el AYP a pesar de enfocarse en las pruebas durante todo el año. Incluso los maestros cuyos estudiantes aumentaron sus puntajes *a veces se preguntaban si valía la pena todo lo que sacrificaban en el proceso*. En la primavera de 2011, un docente de tercer grado en el lado suroeste de Chicago, describió sus sentimientos encontrados de esta forma:

Recibimos nuestros resultados del ISAT el viernes pasado. 21 de mis 26 alumnos aprobaron. Las escuelas Públicas de Chicago y la administración diría que importó. Sin embargo, honestamente, no puedo estar de acuerdo. A pesar de que dejó escapar un suspiro de alivio cuando vi los resultados, todavía me siento como si engañara a *mis* estudiantes. Pasaron una prueba que les permite continuar en el 4 ° grado, pero sienten no que han aprendido nada. De enero a marzo todo lo que se enseñó fue la preparación del ISAT. Era una tortura para mí y para mis alumnos. Estoy seguro de que hay algunas otras cosas que importan a mis alumnos, pero debido al *gran énfasis en los ISATs*, la mayoría de mis estudiantes diría que pasar *el ISAT era lo que importaba*. En mis cortos cinco años en el CPS, la educación se ha ido cuesta abajo, y por desgracia lo que creo que importa no puede ni siquiera ser enseñado.<sup>20</sup>

Aunque muchos celebran las altas puntuaciones y sus ganancias, cuando se examina más de cerca, a menudo son mínimas. Hay creciente evidencia que las políticas de la última década de una educación centrada en las pruebas no han sido eficaces incluso en sus propios y estrechos objetivos. En las escuelas públicas de Chicago, una sucesión de CEO desde finales de 1990 se ha impactado por las calificaciones de los estudiantes de todo el sistema. Pero en un informe de 2011,<sup>21</sup> que sintetiza 20 años de investigación sobre la reforma de la escuela en la ciudad, el Consorcio de Investigación Escolar de Chicago encontró que las estadísticas utilizadas por el distrito para las comparaciones de puntuación de las pruebas *año a año* se encontraban defectuosas. Cuando estas fueron correctamente ajustadas, las puntuaciones de lectura en las escuelas primarias de Chicago mostraron un crecimiento casi nulo en un período de 20 años. Incluso peor aún, la brecha en los puntajes en lectura entre Estudiantes *blancos* y Estudiantes afro-americanos aumentó en las últimas dos décadas.

Por supuesto, estos resultados van en contra de la información que la oficina central ha presentado en los últimos años. Pero no es una sorpresa para Chicago, que se haya leído entre líneas los *comunicados de prensa*. Incluso un rápido vistazo a la lista de escuelas con *mayor puntuación en las pruebas anuales* revela grandes disparidades distribuidas en *la raza* y en las diferencias de clase social. En la primaria 15 de Chicago la escuelas con los puntajes más altos de prueba en 2010, sólo el 19% de los estudiantes,

en promedio, eran de bajos ingresos. Comparado eso con el distrito en su conjunto, donde el 86% de los estudiantes son pobres.<sup>22</sup>

Y es en las escuelas con los mayores índices de pobreza donde los profesores, estudiantes y padres de familia sienten la presión de mejorar los resultados de las pruebas de forma más aguda. Un maestro de séptimo grado en una escuela de inmigrantes predominantemente mexicanos *dice que el énfasis en los exámenes con consecuencias de alto riesgos ha llevado a una visión distorsionada de la enseñanza y el aprendizaje.*

Estos exámenes han transformado mi conversación con los padres y miembros de la comunidad. Los resultados son vistos como una finalidad. Una valoración errónea se está asignando a mis estudiantes en la cual se está definiendo el desarrollo, como una especie de marca de ganado o un número de identificación criminal. Ya sabes, “el juego se termino” si no obtienes un buen puntaje. Los padres escuchan *la explicación del distrito acerca de las puntuaciones, y en vez de mirar el completo paisaje educativo y la situación de sus hijos, me preguntan: “¿algo anda mal con mi hijo?”*<sup>23</sup>

El punto más importante, sin embargo, es el siguiente: La obsesión por la las pruebas estandarizadas en las escuelas públicas es algo más que educativamente *ineficaz. No es democrático. La estrecha mirada en los resultados de las pruebas producidas por la ley NCLB y el CPS ha erosionado el control comunitario de las escuelas, disminuido la voz del maestro, y-en muchos casos ha extinguido, el enfoque humanista al currículo. El enfoque en los exámenes con consecuencias de alto riesgos juegan un papel de distracción. Este anima a los educadores a dedicar todo su tiempo en el cumplimiento de las apariencias, y «datos» en lugar de enfocarse en preguntas que han sido perdidas en los exámenes con consecuencias de alto riesgos: Educación con qué fin? ¿Cuál después de todo, son los propósitos de las escuelas públicas en una democracia?*

## ALTERNATIVAS Y SOLUCIONES

**Malinterpretación:** los exámenes con consecuencias de alto riesgos *son un mal necesario.*

### Las evidencias:

- *Finlandia, uno de los países con puntajes más altos en medidas internacionales, ha abandonado casi completamente los exámenes estandarizados.*
- *El pensamiento complejo y las aptitudes son medidas muy pobremente en un examen de selección múltiple.*
- *las herramientas basadas en el desempeño evalúan y fomentan un aprendizaje más válido y apropiado.*

Algunas personas aún preguntan: ¿Aún con los defectos e inconvenientes, no necesitamos algún tipo de prueba

estandarizada? ¿No son la equidad y la calidad asegurados por medidas uniformes y estándar? ¿No son los *exámenes* una muestra de un sistema educativo globalmente competitivo? Las investigaciones muestran de forma convincente que la respuesta a estas preguntas es no. La evaluación es una parte necesaria en la enseñanza, en la planeación curricular, y en la elaboración de políticas educativas, pero esto no quiere decir que los exámenes con consecuencias de alto riesgos, son la única herramienta que puede ser utilizada con propósitos evaluativos.

Si bien el énfasis en las pruebas puede considerarse un movimiento de dimensiones mundiales, hay un montón de ejemplos de los sistemas educativos de todo el mundo, donde la evaluación se ha estructurado de manera muy diferente a lo que vemos en muchas de nuestras escuelas en los Estados Unidos. Un ejemplo es Finlandia. Las escuelas de este país tienen Recientemente llamaron la atención internacional debido a su alto desempeño de sus estudiantes en el Programa para la Evaluación de Estudiantes Internacionales (PISA por sus siglas en Inglés). Si bien es importante tener cuidado con lo que las lecciones y las características pueden ser “llevadas” de un lugar a otro, algunas de las características que lograr hacer de la educación finlandesa un éxito en el logro de la equidad, el desarrollo individual de los estudiantes, y la competitividad económica nacional, incluye elementos, que investigadores educativos de los Estados Unidos han documentado como componentes esenciales de escuelas exitosas. Estos incluyen: la inversión en profesores altamente capacitados, cuyo tiempo de trabajo incluye la planificación y la preparación, un equilibrio entre la descentralización y la centralización en la gobernanza, la gestión y el diseño curricular, la provisión de recursos para aquellos que más lo necesitan; ayudas para estudiantes con necesidades especiales; así como la confianza y el respeto como en la sociedad finlandesa por el trabajo de los educadores.<sup>24</sup>

Lo que quizás más llama la atención sobre el sistema educativo finlandés son las reformas que han estado en curso desde la década de 1970 en las cuales se evidencia el abandono casi total de los exámenes con consecuencias de alto riesgos. En cambio, Finlandia utiliza tareas basadas en la escuela y en el estudiante, las cuales se encuentran *están* incorporadas en el currículo. Los maestros proveen informes formativos verbales y por escrito, con el principal objetivo de fomentar el aprendizaje en los alumnos. *Sin duda, Finlandia debe tener algunas pruebas estandarizadas? Sí, hay una* (1) un examen que los estudiantes pueden tomar antes de la universidad, el examen de matrícula. Este examen no es necesario para graduarse de la escuela secundaria o para ingresar en una universidad, sin embargo, es una práctica común que los estudiantes tomen este examen que hace *énfasis* en la resolución de problemas, análisis y la escritura.

Finlandia, sirve como un recordatorio útil de que los exámenes con consecuencias de alto riesgos no son la única manera de controlar y mejorar las escuelas. Como se señaló anteriormente, las investigaciones *muestran* que los exámenes con consecuencias de alto riesgos son

contraproducentes. Es por esto que escuelas, distritos y países de todo el mundo son cada vez más conscientes de que:

- *Las pruebas de selección múltiple son poco adecuadas para la medición de las aptitudes complejas y analíticas y otros tipos de pensamiento crítico que niños y jóvenes necesitan para un verdadero compromiso cívico y éxito económico en el mundo de hoy.*
- *En vez de realizar pruebas para medir la capacidad de responder preguntas preestablecidas, las escuelas deben evaluar a los estudiantes « sus habilidades para sintetizar, para colaborar, para deliberar, para gestionar proyectos, para resolver problemas e innova de manera creativa.*
- *En lugar de pruebas, obligatorias de fin de grado, los estudiantes necesitan evaluaciones que fomenten la auto-reflexión, el establecimiento individual de metas y habilidades de aprendizaje activas.*

Hace un siglo, los Estados Unidos *adoptó la realización de exámenes en la educación* buscando la uniformidad y la normalización. Si bien era dudoso que esto era apropiado inclusive para el industrialismo de principios y mediados del del siglo 20, es muy claro que en el contexto mundial contemporáneo, instrumentos más complejos tales como las evaluaciones de desempeño, son viables y vitales como una alternativa a los exámenes con consecuencias de alto riegos.<sup>25</sup>

## Hacia la Acción

**Malinterpretación:** *Los exámenes están aquí para quedarse y no hay nada que podamos hacer.*

### Las evidencias:

- *Una variedad de alternativas de evaluación pueden ser usadas en vez de los exámenes con consecuencias de alto riegos.*
- *Escuelas con alto desempeño están utilizando métodos de evaluación originales.*
- *La unión de los maestros puede cambiar las políticas de evaluación.*
- *El público en general debe manifestarse en contra de los exámenes con consecuencias de alto riegos.*

La dura realidad de los exámenes con consecuencias de alto riegos ha permeado nuestra comunidad y nuestro subconsciente en las últimas décadas. Con el Acto de Educación Primaria y Secundaria (ESEA por sus siglas en Inglés) para la reautorización, éste llegó a ser críticamente importante para las comunidades, de forma tal que tomaran actitudes proactivas en contra de los exámenes con consecuencias de alto riegos. CReate está de acuerdo con el Sindicato de Maestros

de Chicago de que las pruebas de que estos exámenes se siguen promoviendo, a pesar de los resultados de investigaciones *en contra de su uso*, debido a la rentabilidad de las industrias relacionadas con la elaboración de estos exámenes (*fabricantes de pruebas, proveedores, y así sucesivamente*), y que constituyen *parte de un gran intento por parte de los intereses corporativos de “alinear los propósitos de la educación pública con las necesidades de sistema económico altamente polarizado y desigual.”*<sup>26</sup> Sin embargo, no todo está perdido. A lo largo de la ciudad y la nación, se están llevando a cabo esfuerzos para interrumpir la tendencia deshumanizadora de los exámenes con consecuencias de alto riegos como única medida de logro académico.

Organizaciones como Proyecto Appleseed *han comenzado a operar* como un centro para nuevas alternativas de evaluación que las escuelas, los distritos, las escuelas individuales, y los profesores en las aulas puedan utilizar en lugar de los exámenes con consecuencias de alto riegos.<sup>27</sup> Expertos en psicometría y otros investigadores *han reconocido que la principal función de las pruebas estandarizadas es medir qué tan bien los estudiantes toman un examen; en cambio existen otros que pueden proporcionar información mucho más útil para los educadores sobre el desarrollo de sus alumnos, sus conocimientos y habilidades,*<sup>28</sup> “*incluyendo el proceso mediante el cual ellos resuelven problemas.*”<sup>29</sup> Requiriendo estudiantes que sean participantes activos, los profesores son capaces de diseñar proyectos que les permitan desarrollar las habilidades necesarias para interpretar, analizar y tomar decisiones. Las evaluaciones basadas en el desempeño fomentan las competencias que les serán de uso en el mundo real a los estudiantes, por ejemplo:

- *Los proyectos del Grupo de habilitan a los estudiantes a trabajar juntos en un problema complejo el cual requiere planificación, investigación, discusión interna, y presentación al grupo.*
- *Los ensayos evalúan la comprensión de los estudiantes de un tema a través de una descripción escrita, análisis, explicación o de un resumen.*
- *Los experimentos prueban qué tan bien los estudiantes entienden conceptos científicos y pueden llevar a cabo procesos científicos.*
- *Las demostraciones dan a los estudiantes las oportunidades para mostrar su dominio de los contenidos de las asignaturas.*
- *Los portafolios permiten a los estudiantes proporcionar un amplio retrato de su desempeño a través de los archivos que almacenan a lo largo del tiempo.*<sup>30</sup>

En Chicago, este tipo de evaluación y currículo es visto con mayor frecuencia en escuelas con alto rendimiento. Mientras que algunas escuelas afirman que se centran en

el aprendizaje basado en proyectos, la gran mayoría de las escuelas que han sido puestas en observación terminan limitándose o eliminando este tipo de instrucción con el fin de aumentar significativamente el tiempo y los recursos para la preparación de estos exámenes, alejándose de los tipos de enseñanza que pueden aumentar el aprendizaje y el éxito.

Es importante señalar que las comunidades no han tomado esto sin hacer nada. Chicago, con sus profundas raíces en el activismo educativo, ha sido el hogar de varios intentos de interrumpir la locura de los exámenes actuales. Un ejemplo es un grupo de profesores de la Escuela Secundaria Curie en el lado suroeste conocida como la “Curie 12,” que se negó a administrar el examen de distrito de escuelas secundarias prueba conocida como el Examen Estándar Académico de Chicago (CASE por sus siglas en Inglés).<sup>31</sup> Ellos argumentaron que el examen era “pobremente desarrollado, el cual realizaba preguntas al azar a los estudiantes no relacionadas con el currículo”.<sup>32</sup> También actuaron estratégicamente:

- Se recopilaron datos concretos para apoyar su afirmación de que el examen CASE no mejoraba la educación, sino que en realidad impedía los logros de los estudiantes.
- Permitieron que la administración de la escuela supiera lo que iban a hacer con anterioridad.
- Eligieron a dos maestros veteranos como portavoces para el grupo, para evitar que los nuevos maestros fueran atacados de forma individual.<sup>33</sup>

Poco después, el CPS retiró el caso de cada escuela secundaria en las CPS. Katie Hogan, uno de los “Curie 12.” refleja que,

No, nosotros no éramos doce profesores enojados que querían decirle a CPS que la retiraran. Éramos doce maestros regulares en diferentes extremos del activismo que quería poner fin a un mal examen. Nosotros no esperábamos ganar, sólo esperamos luchar. ¿No es eso lo que debemos enseñar a nuestros estudiantes todos los días?<sup>34</sup>

Otros esfuerzos impulsados por la comunidad han desafiado abiertamente la eficacia de los exámenes con consecuencias de alto riesgo. Por ejemplo, Padres Unidos por una Educación Responsable (PURE por sus siglas en Inglés) tiene desarrollado un conjunto de acciones de incidencia pública para impulsar los pasos para una evaluación responsable ante los problemas y preocupaciones de los padres y estudiantes, entre ellos:

- Aprender acerca de algunos de los problemas con los exámenes. El sitio web FairTest ([www.fairtest.org](http://www.fairtest.org)) es un buen lugar para comenzar.

- Piense en cómo las pruebas han afectado las vidas de usted, su hijo y / o alguien que usted conozca.
- Elabore un mensaje acerca de las pruebas para compartir en los diversos contextos, tales como cartas al editor.
- Informe a los demás.
- Dígale a sus legisladores locales, estatales, y federales lo que piensa acerca de la obsesión nacional actualmente con los exámenes.
- Únase a otros para aprender más acerca de los problemas, compartir lo que sabes, y asumir una responsabilidad conjunta.<sup>35</sup>

También sugieren que cuando las personas están en contacto con los legisladores o los medios de comunicación se debe utilizar “una sola frase clara de porqué que se oponen a la evaluación”.<sup>36</sup> Estos pequeños pasos colectivos son fundamentales para escuchar la voz de las familias que pueden sentir que algo está mal con el currículo basado en la preparación de exámenes, pero que de alguna forma no pueden expresar sus preocupaciones. La clave de este proceso es el reconocimiento de los esfuerzos en toda la ciudad frente a la excesiva dependencia de los exámenes con consecuencias de alto riesgo. Numerosos grupos de activistas como la Organización de la Comunidad de Kenwood-Oakland, la Federación Grand Boulevard, y Cuadras Unidas han adoptado una postura colectiva sobre los peligros de los currículos basados en la preparación de exámenes.

Mediante el desarrollo de documentos en esta área, CReATE busca apoyar el movimiento de los estudiantes, los maestros, familias y miembros de la comunidad para oponerse a los exámenes con consecuencias de alto riesgo. Mientras que la investigación tiene el potencial de beneficiar las vidas de los grupos antes mencionados, sólo puede servir de manera efectiva si se combina, la acción comunitaria.



## Notas

- 1 McCracken, M., & McCracken, H. T. (2001). Teaching in the time of testing: What have you lost? *The English Journal*, 91(1), p. 30.
- 2 Ronell, A. (2005). *The test drive*. Urbana: University of Illinois Press, p. 17.
- 3 Taubman, P. M. (2009). *Teaching by numbers: Deconstructing the discourse of standards and accountability in education*. New York, NY: Routledge.
- 4 Wainer, H. (2011). *Uneducated guesses: Using evidence to uncover misguided education policies*. Princeton, NJ: Princeton University Press.
- 5 Ibid.
- 6 Kliebard, H. (2004). *The struggle for the American curriculum: 1983-1958*. New York, NY: Routledge/Falmer; Spring, J. (2012). *Education networks: Power, wealth, cyberspace and the digital mind*. New York, NY: Taylor & Francis; Pinar, W., Reynolds, P., Slattery, P., & Taubman, P. (2008). *Understanding curriculum*. New York, NY: Peter Lang.
- 7 Thorndike, R. M. (2011). *Measurement and evaluation in psychology and education* (8th ed.). Upper Saddle River, NJ: Prentice Hall.
- 8 Taubman, 2009.
- 9 Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools*. Boston, MA: Heinemann; Sacks, P. (1999). *Standardized minds: The high price of America's testing culture and what we can do to change it*. Cambridge, MA: Perseus.
- 10 Wainer, 2011.
- 11 Kohn, A. (2004). *What does it mean to be well educated?* Boston, MA: Beacon Press.
- 12 Popham, W. J. (2011). *Classroom assessment: What teachers need to know*. Boston, MA: Pearson, p. 61.
- 13 From Thorndike, 2011.
- 14 Popham, 2011, p. 83.
- 15 Thorndike, 2011.
- 16 National Research Council. (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, D.C.: National Academy Press; American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for psychological and educational testing*. Washington, D.C.: AERA; Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, D.C.: American Psychological Association.
- 17 Lipman, P. (2004). *High stakes education: Inequality, globalization, and urban school reform*. New York, NY: RoutledgeFalmer.
- 18 Personal communication, 2006.
- 19 Personal communication, 2012.
- 20 Personal communication, 2011.
- 21 Luppescu, et al. (2011). *Three eras of Chicago school reform*. Chicago, IL: Consortium on Chicago School Research.
- 22 Illinois Interactive Report Card data, 2011.
- 23 Personal communication, 2012.
- 24 Sahlberg, P. (2011). *Finnish lessons: What can the world learn from educational change in Finland?* New York, NY: Teachers College Press; Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future*. New York, NY: Teachers College Press; Laukkanen, R. (2008). *Finnish strategy for high-level education for all*. In N. C. Soguel & P. Jaccard (Eds.), *Governance and performance of education systems* (pp. 305-324). Dordrecht, the Netherlands: Springer.
- 25 See also Davidson, C. N. (2011). *Now you see it: How the brain science of attention will transform the way we live, work and learn*. New York, NY: Viking.
- 26 Chicago Teachers Union. (2010). *CTU position paper: Debunking the myths of standardized testing*.
- 27 [www.projectappleseed.org/assessment.html](http://www.projectappleseed.org/assessment.html)
- 28 Bryk, A. S. (2003). *No Child Left Behind, Chicago style*. In P. E. Peterson & M. R. West (Eds.), *No Child Left Behind?* Washington, D.C.: Brookings Institution Press.
- 29 Ibid.
- 30 Ibid.
- 31 Hogan, K. (2008). *The Curie 12: A case for teacher activism*. In W. Ayers, G. Ladson-Billings, G. Michie, & P. Noguera (Eds.), *City kids city schools: More reports from the front row*. New York, NY: The New Press.
- 32 Ibid.
- 33 Ibid.
- 34 Ibid.
- 35 Parents United for Responsible Education. (2010). *Action steps for better assessment: What you can do to help win the testing battle*.
- 36 Ibid.

Para mayor información acerca de las actividades, ser parte de CReATE o para leer la versión completa de este y otros artículos, por favor visite <http://createchicago.blogspot.com> o envíe un correo electrónico a [createchicago@gmail.com](mailto:createchicago@gmail.com)

Este resumen fue elaborado por Isabel Nuñez, Peter Hilton, Eileen Quinn Knight, Kevin Kumashiro, Jeff Kuzmic, Greg Michie, Brian Schultz, Noah Sobe, and David Stovall. Este artículo fue traducido por Felipe Agudelo.

### CReATE

La educación pública en una sociedad democrática está basada en el principio de que cada niño y niña tienen el mismo valor, además de ser incalculable. Este principio requiere el desarrollo completo de cada miembro de este país. Se necesitan escuelas públicas más efectivas para permitir que cada miembro de los Estados Unidos alcance su más alto potencial. Las escuelas en una democracia forman la próxima generación para que sean ciudadanos y residentes informados; para que usen el pensamiento de una manera crítica; para que solucionen problemas creativamente; para que estén preparados a contribuir positivamente en sus comunidades, lugares de trabajo y sociedades caracterizadas por la diversidad y las inequidades; y para ser saludables, felices y preparados para apoyar el bien común con compasión y valor. Los niños y juventud de Chicago no merecen menos.

CReATE, es un grupo de investigadores académicos voluntarios que estudian e investigan la educación en el área de Chicago. CReATE conduce, revisa y distribuye los resultados de estudios, los cuales se enfocan en los problemas de nuestros estudiantes, padres de familia, y escuelas, además de promover a lo largo de la ciudad, conocimiento y diálogos acerca de asuntos educativos a través de eventos públicos gratuitos. CReATE se asocia con educadores, grupos de interés público y organizaciones comunitarias. Nuestros miembros académicos están disponibles para comentarios y foros en temas relacionados en el presente resumen y en nuestra política, la cual se encuentra disponible en: <http://createchicago.blogspot.com>