



Programa de Promoción de la Reforma
Educativa en América Latina y el Caribe

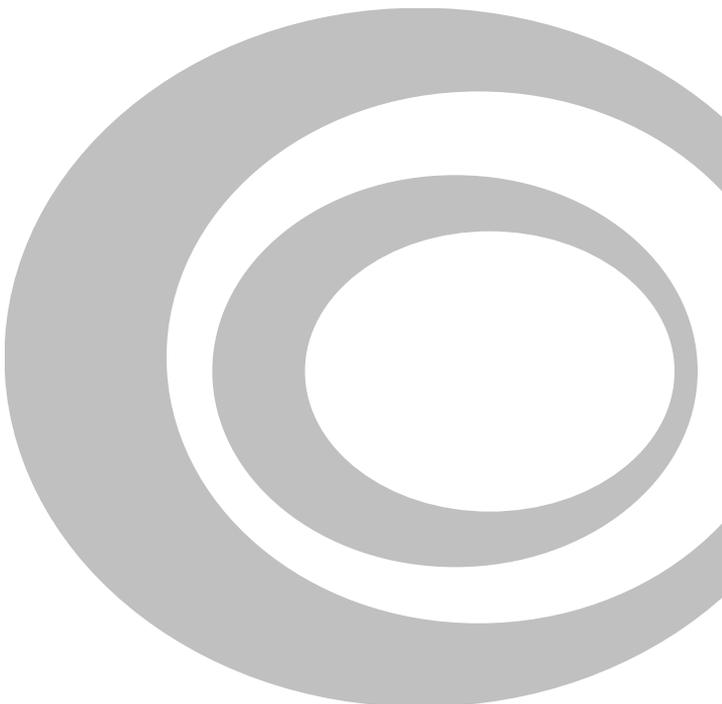
**Grupo de Trabajo
sobre Estándares y
Evaluación**

**Usos y límites de las
pruebas de desempeño**

Elliot W. Eisner



Grupo de Análisis para el Desarrollo



USOS Y LÍMITES DE LAS PRUEBAS DE DESEMPEÑO

Elliot W. Eisner

*Tomado y traducido de **Phi Delta Kappan** de mayo de 1999, con autorización de sus editores. El autor es profesor de educación y arte en la Universidad de Stanford, California y autor de "Cognición y Currícula Reconsiderados", publicado por Teacher's College Press.*

Las pruebas de desempeño son uno de los "temas candentes" en la agenda de la reforma educativa -- y por buenas razones. Las pruebas de desempeño son una medida más precisa de la capacidad de nuestros niños de lograr las aspiraciones que tenemos para ellos que las formas convencionales de la evaluación estandarizada. En efecto, nuestras aspiraciones educativas han sido influenciadas por el hecho de que nuestros niños habitarán un mundo que exigirá formas de pensamiento mucho más complejas y sutiles que aquéllas que necesitaban los niños tres o cuatro décadas atrás. Por ejemplo, nuestros niños necesitarán saber cómo plantearse problemas, cómo formular planes para encararlos, cómo evaluar resultados múltiples, cómo considerar relaciones, cómo lidiar con la ambigüedad y cómo cambiar sus objetivos a la luz de información nueva.

Estos modos de pensamiento serán críticos en una sociedad en la cual es probable que los ciudadanos cambien de carrera varias veces durante el curso de su vida laboral¹, en la cual se ha incrementado la movilidad y se requieren nuevas formas de adaptabilidad, y en la cual elegir un curso de acción requiere la consideración de información diversa y algunas veces contradictoria. La mayoría de empleos, especialmente aquéllos que son más atractivos, no exigirán más el uso de habilidades rutinarias ni de la memoria mecanizada². Según anota Edward Haertel en su artículo en este número de *Phi Delta Kappan* [mayo 1999], estas expectativas cambiantes sobre los resultados de la educación reflejan una visión no conductista sobre la naturaleza humana. "La práctica hace la perfección" podría ser utilizado como un principio guía para la enseñanza cuando el aprendizaje era concebido como la adquisición y agregación de unidades de información reforzadas. El tipo de pensamiento en que se fomenta ahora que los estudiantes se involucren requiere mucho más de lo que soñó Edward Thorndike, el padre del conexionismo psicológico norteamericano. El contexto importa, el criterio importa y la oportunidad de actuar para poner a prueba nuestras especulaciones es de importancia crítica.

La caducidad del conductismo y el surgimiento del constructivismo en nuestra visión sobre la naturaleza humana no son las únicas fuentes de nuestro cambiante concepto sobre los niños y la educación. Hemos llegado a comprender que el significado importa y que no es algo que pueda impartirse de maestro a alumno. Hasta cierto punto, todo lo que los maestros pueden hacer es "introducir ruido en el entorno". Con esto quiero decir que en educación no tenemos un canal directo hasta los cerebros de nuestros estudiantes. Somos modeladores del medio ambiente, estimuladores, motivadores, guías, consultores, recursos. Pero en última instancia, lo que los niños hagan con lo que

¹ Comisión de la Secretaría sobre Adquisición de Habilidades Necesarias; *Lo que el Trabajo Demanda de los Colegios: un reporte SCANS para América 2000*. Washington, D.C., Departamento de Trabajo de los Estados Unidos, 1991.

² Ibid.

les proporcionemos es función de lo que ellos deriven de lo que nosotros brindamos³. Los significados no están dados, se hacen. Y estamos interesados en facilitar que los alumnos hagan significativas sus actividades en la escuela, no sólo por las notas que reciben si no, lo que es más importante, por las satisfacciones y percepciones que sus esfuerzos hacen posibles.

Ahora comprendemos que los tipos de significados que pueden establecer nuestros estudiantes están relacionados con las formas de representación que ellos mismos pueden utilizar o pueden decodificar cuando otros las han utilizado⁴. Cada una de las formas de representación que existen en nuestra cultura -- formas visuales en el arte, formas auditivas en la música, formas cuantitativas en las matemáticas, formas proposicionales en ciencias, formas coreográficas en la danza, formas poéticas en el lenguaje -- son medios a través de los cuales se conceptualiza y expresa el significado. Una vida impulsada por la búsqueda de sentido se enriquece cuando los significados que se buscan y encuentran son múltiples.

Además de estas consideraciones, también hemos empezado a reconocer que el propósito de la educación no es simplemente que a nuestros niños les vaya bien en la escuela. Las implicancias o costos son considerablemente más altos. Lo que buscamos es que les vaya bien en la vida fuera de la escuela; los puntajes generados por los tipos de pruebas que hemos estado utilizando son aproximaciones pero, desafortunadamente, hemos descubierto que, como aproximaciones, su mayor utilidad es para hacer deducciones sobre los puntajes que los estudiantes podrían obtener en otras pruebas⁵.

Queremos más. Lo que queremos es una aproximación a la evaluación que posea lo que los psicometristas llaman *validez concurrente o predictiva*. Esto es, queremos que los puntajes de las pruebas nos digan de qué manera los estudiantes se enfrentan a tareas más allá del salón de clase.

Estos factores -- la casi caducidad del conductismo, la emergencia del constructivismo, la importancia del sentido, el deseo de obtener validez concurrente y predictiva -- han proporcionado la base para el interés en las pruebas de desempeño. A pesar de que no existe una definición única, las pruebas de desempeño intentan trasladarse desde las prácticas evaluativas que exigen a los estudiantes seleccionar la única respuesta correcta entre una serie de cuatro o cinco distractores hacia una práctica que requiere que los alumnos generen evidencia, mediante su propio desempeño, que permita a los examinadores realizar juicios válidos sobre "lo que ellos saben y lo que pueden hacer" en situaciones que importan. Las pruebas de desempeño constituyen el desarrollo más importante en la evaluación desde el invento de las pruebas de respuesta corta y su uso extendido durante la Primera Guerra Mundial.

³ Desde John Dewey hasta Jean Piaget y Jerome Bruner, los principales psicólogos y filósofos han enfatizado la naturaleza constructiva de la cognición. Los seres humanos *hacen* sentido de los eventos del mundo en el cual habitan. Cómo se construyen esos eventos está influenciado por los recursos con los cuales trabaja la gente. El modelo transmisivo de la enseñanza estuvo fundamentado en la premisa de que la transmisión es posible.

⁴ Para una discusión sobre los efectos de formas de representación sobre la cognición, ver **La cognición y el currículo reconsiderados** de Elliot W. Eisner, segunda edición (Nueva York, Teachers College Press, 1994).

⁵ David C. McClelland, "Midiendo Competencia en Lugar de Inteligencia", **American Psychologist**, Enero 1973, pp. 1-14.

La prueba de inteligencia Alfa del ejército norteamericano⁶ era, entre otras cosas, una representación de nuestra visión sobre el aprendizaje, una visión promulgada por Thorndike, el padre de la teoría del estímulo - respuesta. Esta forma de evaluación recibió fuerte ímpetu con la creación de dispositivos ópticos para la calificación y fue considerada como un procedimiento sólido para determinar qué habían aprendido los estudiantes. Las pruebas estandarizadas de logros encajaban cómodamente no sólo dentro de nuestro concepto sobre el aprendizaje sino también dentro de nuestros conceptos sobre la educación escolar. Considérense las maneras en que están organizadas las escuelas, los supuestos tácitos subyacentes a esa organización, y cómo encajan las pruebas de logros convencionales en la organización escolar.

En 1847 se inventó la primera escuela dividida por grados en Norteamérica⁷. Los supuestos sobre el curso del aprendizaje en base a los cuales se construyó la escuela eran sencillos: los alumnos serían agrupados por edades, y cada edad sería asignada a un grado. La distribución por grados según las edades se convirtió en la estructura organizacional dominante en nuestras escuelas. Se asumía además que ya que los alumnos estaban agrupados por edades, el contenido y los objetivos para cada grado deberían ser los mismos para todos los niños en dicho grado. Se definía una enseñanza efectiva como la capacidad de hacer posible que todos los niños de un grado logaran los objetivos para ese grado. Los alumnos, como un ejército marchando en tándem, dejarían la escuela al final de un período de ocho o doce años habiendo dominado el contenido asignado a cada uno de los grados anteriores.

Estos supuestos sobre el aprendizaje de los seres humanos y estas características de la organización escolar siguen vigentes hoy en día en las escuelas norteamericanas. En efecto, quizás más que en el pasado, la especificación de los estándares por grado y nivel es más que una aprobación tácita de la distribución por grados según la edad. Si siguiéramos el consejo del presidente Clinton, tomaríamos tan en serio los estándares por grado y nivel que ningún niño sería promovido al grado siguiente sin haber cumplido las expectativas para ese grado, a pesar de que las investigaciones indican que generalmente la repitencia no es un buen remedio⁸. Las raíces del problema son más profundas.

Hemos descubierto que el desarrollo humano no procede de una manera ordenada. [...] Los niños difieren no sólo en la tasa de su desarrollo sino también en las áreas de trabajo particulares que se espera desempeñen. Algunos alumnos tienen un alto nivel de aptitudes en artes; otros en ciencias. Algunos niños están dotados de habilidades sociales; otros, en el uso del idioma. Si por arte de magia la enseñanza pudiera hacerse óptima para todos los alumnos de una clase de veinticinco, en cada grado subsiguiente se incrementaría la variabilidad del desempeño estudiantil de esa clase. Las formas óptimas de enseñanza para aquéllos con talento para el arte les permitiría avanzar en actividades artísticas más lejos y más rápidamente que aquéllos talentosos en otras áreas; los que tienen talento para las matemáticas podrían desplazarse en las áreas cuantitativas más lejos y más rápidamente que aquéllos no tan talentosos en esa área. Con una enseñanza óptima, la variabilidad, con toda seguridad, se incrementaría. Sin embargo, mucho de lo que intentamos hacer en educación está basado en la estandarización y la uniformidad, en

⁶ Clarence S. Yoakum y Robert M. Yerkes, editores, **Pruebas Mentales del Ejército** (Nueva York: Henry Holt & Co., 1920).

⁷ David Tyack, **El Mejor Sistema** (Cambridge, Mass.: Harvard University Press, 1974).

⁸ John M. Goodlad, **La Escuela Elemental Multigrado** (edición revisada, Nueva York, Teachers college, Universidad de Columbia, 1987).

la homogenización y en un modelo de aprendizaje representado por una procesión ordenada de alumnos marchando al unísono a través de los grados hacia objetivos fijos que han sido definidos bastante de antemano. La evaluación estandarizada siempre ha encajado bien en este modelo.

Según señala Linda Mabry en su artículo publicado en este número de *Phi Delta Kappan* [mayo 1999], si bien las prácticas de las pruebas de desempeño que emplean rúbricas restrictivas para calificar pueden ser congruentes con una aproximación a la enseñanza estandarizada y graduada por edades, las prácticas de las pruebas de desempeño que ofrecen oportunidades para respuestas abiertas y que permiten a los jóvenes demostrar sus fortalezas chocan con los supuestos sobre la uniformidad. Se puede buscar y evaluar la uniformidad si las prácticas evaluativas limitan significativamente la manera en que se obtienen y analizan los datos. La evaluación de resultados estandarizados es mucho más compleja cuando las tareas de las pruebas de desempeño tienen calidad de abiertas y hacen así posible la expresión de la individualidad. El problema es aun más complicado.

Una de las motivaciones detrás del movimiento a favor de los estándares es el deseo de responsabilizar a las escuelas, y esa responsabilización se hace más fácil si se pueden comparar escuelas, salones de clase y alumnos. Si los alumnos se desplazan entre diferentes rutas u opciones curriculares o si las prácticas evaluativas proporcionan tareas abiertas, la capacidad para comparar se debilita; las escuelas, los distritos y estados pueden compararse más fácilmente cuando las prácticas evaluativas, los objetivos y el contenido curricular son uniformes y cerrados. Pero las comparaciones son más que engañosas cuando, por ejemplo, la cantidad de tiempo asignada al estudio de un tema varía de escuela en escuela, cuando el contenido de un tema enseñado varía de una escuela a otra o cuando las metas de un campo de estudio varían en distintas escuelas. De este modo, si bien las prácticas evaluativas abiertas hacen posible la evaluación de individuos, ellas pueden no proporcionar el tipo de “toma de temperatura” general y comparativa que parece desear un público preocupado por la productividad educativa de sus escuelas⁹. ¿Hay alguna posibilidad de desarrollar una aproximación a las pruebas de desempeño que haga visibles los logros particulares de estudiantes individuales y al mismo tiempo proporcione información sobre una clase o escuela que sea útil para compararla con otras clases o escuelas?

Claro que una opción es emplear dos tipos diferentes de pruebas. Una de éstas sería el uso continuado de las pruebas de “toma de temperatura” a gran escala con el fin de proporcionar datos comparativos sobre el desempeño de las escuelas o los distritos escolares. Tal práctica no permitiría la descripción de formas características del desempeño del estudiante individual. Se necesitaría diseñar un segundo tipo de prueba para revelar los talentos distintivos de estudiantes individuales y los efectos que tiene la práctica escolar en su desarrollo. Una forma de evaluación se focalizaría en lo general; la otra, en lo particular.

Y, sin embargo, no puedo evitar preguntarme si una prueba orientada a la revelación de talentos individuales podría sobrevivir en una “carrera” con una prueba destinada a que el público pueda hacer comparaciones entre escuelas y distritos

⁹ El concepto de “toma de temperatura” es particularmente aplicable al enfoque de la Evaluación Nacional de Progreso Educativo (NAEP), en sus muestras de rendimiento estudiantil.

escolares. Por ejemplo, un aspecto de la propuesta más reciente del Presidente Clinton para la reforma escolar es la elaboración de una “libreta de notas” que clasifique el desempeño de las escuelas.

Cada vez que hablamos de escuelas “de primer nivel mundial” está implícito el deseo de comparar. El supuesto básico es que no sólo podemos comparar escuelas, distritos escolares y estados, sino que también podemos hacer comparaciones significativas sobre la calidad educativa de las escuelas en distintas naciones. Si no podemos hacer tales comparaciones de manera significativa y al mismo tiempo realizar los tipos de pruebas que describirían el desempeño del alumno individual con respecto a sus aspectos distintivos, ¿es probable que decidamos abocarnos a la segunda aproximación a la evaluación y renunciemos a la primera? ¿Dejaremos de lado nuestra inclinación a buscar comparaciones, a formar categorías, a establecer jerarquías, a posicionarnos en relación a los demás? Después de todo somos una meritocracia y tendemos a utilizar los puntajes de las pruebas en educación para identificar a los meritorios. El uso de las pruebas estandarizadas favorece estos fines.

Bajo estas circunstancias, ¿qué se requiere para dar un espacio destacado a un sistema de pruebas que facilite la enseñanza y el aprendizaje y también revele los logros intelectuales distintivos de los alumnos individuales? Desde mi punto de vista, lo que necesitamos es un cambio en el concepto que tiene el público sobre la misión de las escuelas. Por supuesto que producir un cambio así no es tarea fácil. Sin embargo, se necesita generar un cambio desde el concepto de enseñanza como una carrera de caballos o como una especie de olimpiadas educacionales hacia un concepto de las escuelas como lugares que fomentan los talentos distintivos de los alumnos. Como he sugerido, una buena escuela no disminuye las diferencias individuales; las incrementa. Eleva el promedio e incrementa la varianza.

Producir un cambio tal en el concepto de la escolaridad no se puede lograr, desde mi punto de vista, sin cambios en la estructura educacional norteamericana y en los criterios empleados por las instituciones de educación superior para tomar decisiones sobre sus admisiones. Las universidades afectan de manera muy conservadora a las escuelas secundarias, intermedias e incluso, a las primarias. Necesitamos tener criterios considerablemente más flexibles y abiertos para las admisiones que no continúen privilegiando una serie restringida de aptitudes que son fundamentalmente reflejo de las ventajas de la clase social. Si esos criterios no cambian, los padres, siempre preocupados por la movilidad social y educacional de sus hijos, continuarán adaptándose a los ya existentes. Nuestras prioridades para las escuelas norteamericanas no han surgido de un profundo debate público sobre la misión de la educación; han surgido del deseo por mejorar las puntajes de las pruebas. Los puntajes de las pruebas son ampliamente considerados como aproximaciones a la calidad de la educación. Pero, según declaran todos los autores en este número de *Phi Delta Kappan* [mayo 1999], son totalmente inadecuados.

Mi propósito en este artículo es situar las pruebas de desempeño dentro de un contexto educacional y social amplio. En principio, las pruebas de desempeño nos ofrecen una ocasión para desarrollar maneras de revelar los rasgos característicos de estudiantes individuales. Nos brindan una oportunidad de obtener información sobre el aprendizaje que pueda ayudar a mejorar la calidad, tanto de la currícula como de la enseñanza. En síntesis, nos da una oportunidad para utilizar la evaluación de manera

formativa y tratarla como un recurso educacional. Pero si no cambian las actitudes del público y las expectativas sobre la enseñanza, es poco probable que se materialicen estas oportunidades y es poco probable que cambien sin la revisión de las políticas que afectan la movilidad educacional y social de los alumnos en las escuelas. Nuestra forma de organización escolar brinda bienestar y apoyo a un modelo de educación comparativo y competitivo: la curva en forma de campana es un encomio a la competencia. Nuestra economía está construida sobre un modelo competitivo y hay demasiados reformadores educativos deseando hacer la enseñanza similar a los negocios y formarse un concepto de los factores que ellos creen se deben infundir en la educación de manera similar a aquéllos que dinamizan el mundo de los negocios.

Mis propias inclinaciones son completamente opuestas. Creo que para proporcionar condiciones educativas que favorezcan el desarrollo de nuestros niños, no necesitamos ser motivados a través de la competencia. La mayor satisfacción la recibimos no de la competencia sino de la calidad de la experiencia que proporciona el trabajo significativo. Alfred North Whitehead comentó alguna vez que la mayoría de personas creen que los científicos investigan con el fin de conocer. Señaló que era todo lo contrario. Los científicos conocen con el fin de investigar. El punto de Whitehead era que, para los científicos -- para los mejores, naturalmente -- el gozo se encuentra en el trayecto. Esa observación no es un mal ideal para la educación. Está por verse si nuestra sociedad meritocrática adoptará dicho ideal y realizará cambios en la enseñanza y la política educativa que sean congruentes con la observación de Whitehead. Pero no debemos engañarnos sobre qué es lo que está en juego. A fin de cuentas, lo que está en juego no es sólo la calidad de vida que nuestros hijos puedan gozar, sino también la calidad de la cultura en la que ellos vivirán.