

El papel de las pruebas estandarizadas en un programa de evaluación de la lectoescritura

John J. Pukulski



Hace ya cuarenta años que René Zazzo alertaba sobre un uso ciego y estático de las pruebas estandarizadas y reclamaba el "diagnóstico progresivo". El uso de las pruebas estandarizadas de lectura se ha extendido en nuestro país desde finales de los años setenta a menudo con una aplicación poco individualizada. Aunque refiriéndose a materiales no existentes (aún) en Español, la revisión del autor de las características positivas y negativas de las pruebas estandarizadas más actualizadas de lectura ayudará al lector a comprender su papel y cómo utilizarlas sin sesgos ni abusos en articulación con la evaluación personalizada del educador.

Ha sido un buen año para ser editor del departamento de evaluación de *The Reading Teacher*. El interés en la evaluación parece haber alcanzado el punto más álgido de toda su historia. Considerado un tema aburrido hasta hace poco, la evaluación es ahora un tema sobre el que se escribe y se habla mucho. Recientemente, dos importantes revistas, *Educational Leadership* (abril 1989) y *Pbi Delta Kappan* (mayo 1989) publicaron números temáticos dedicados casi exclusivamente a la evaluación. Quienes se hayan perdido estos números y estén interesados en la evaluación harán bien en leerlos.

Gran parte de lo que he leído y oído sobre la evaluación es negativo y las pruebas estandarizadas son el objetivo de muchas críticas. No he



leído un artículo ni he escuchado una disertación en donde las pruebas estandarizadas reciban elogios por el progreso que han hecho o por el bien que hacen; en cambio, son criticadas rutinariamente por sus limitaciones, si no por el daño que hacen al progreso en la evaluación y la enseñanza de la lectoescritura. El título del artículo "*Standardized Testing: Harmful to Educational Health* [Pruebas estandarizadas: Perniciosas para la salud educativa] (Neill y Medina, 1989) resume bien esta actitud. Sin embargo, a pesar de todas las críticas, las pruebas estandarizadas se utilizan ahora más que nunca. Neill y Medina estiman que se administran más de 105 millones de pruebas estandarizadas al año a los 39,8 millones de estudiantes de las aulas estadounidenses. Otros países no parecen estar menos afectados.

Aparte de la cuestión de si las pruebas estandarizadas tienen derecho a existir, a mí me parece claro, dado el empleo extraordinariamente difundido de estas pruebas y dado el clima político y social existente, que las pruebas estandarizadas no desaparecerán en un futuro inmediato. Ante esta conclusión, que creo inevitable, parece imperativo que, como profesionales responsables, reduzcamos la cantidad de tiempo que empleamos denostando las pruebas estandarizadas de una manera improductiva y que concentremos nuestros esfuerzos en tres cosas: Reducir los malos empleos y las malas interpretaciones de las pruebas estandarizadas, mejorar las pruebas existentes y proponer procedimientos de evaluación alternativos que sitúen a las puntuaciones de las pruebas estandarizadas en un contexto más amplio o empiecen a sustituirlas.

Me ha sorprendido haber recibido este año relativamente pocas pruebas para su revisión aunque escribí a todos los editores que parecían publicar pruebas de lectoescritura o de áreas afines. Esto explica, al menos en parte, por qué se han revisado tan pocas pruebas en la columna de este año. Sin embargo, la verdad es que me siento en cierta medida obligado a revisar el material que me ha sido enviado; por tanto, citaré en esta columna tres pruebas que me fueron remitidas con todos los materiales necesarios para su revisión. Puesto que cada una de estas tres pruebas tiene algunos aspectos positivos que merecen atención, las examinaré brevemente en la sección de este artículo que se centra en algunos de los métodos que podemos tener en consideración para mejorar las pruebas existentes.

REDUCIR EL MAL EMPLEO Y LA MALA INTERPRETACION DE LAS PRUEBAS ESTANDARIZADAS

El principal mal empleo de las pruebas estandarizadas puede ser su *abuso*. Aunque, ciertamente, respeto la necesidad de alguna forma de racionalización, uno de los abusos más patentes de las pruebas estandarizadas se encuentra en los programas estatales de evaluación en los que cada estudiante es sometido a las pruebas cada año. Esta racionalización podría satisfacerse de manera fácil y adecuada administrando las pruebas a los estudiantes solo en determinados cursos, preferiblemente *no antes del tercer o cuarto curso*, y un número de estados cada vez mayor parece estar moviéndose en esta dirección. Además, debería con-



siderarse la alternativa de administrar las pruebas a una muestra de la población total de estudiantes en vez de administrarlas a cada niño de cada curso. Los resultados de administrar pruebas a gran escala, pocas veces *o ninguna* son significativos para interpretar el rendimiento individual de los estudiantes; por tanto, tiene sentido adoptar alguna forma de muestreo aleatorio o aleatorio estratificado. Si todos los estados adoptaran estos procedimientos, la economía que ello representaría -no solo en dólares sino también en tiempo de estudiantes, enseñantes y otro personal educativo- sería enorme. Otra ventaja destacada de administrar pruebas a menos estudiantes sería permitir a los estados adoptar pruebas con elementos más abiertos y otras formas de pruebas de diseño innovador que son difíciles de puntuar en grandes cantidades.

Un ejemplo grave de mala interpretación de las pruebas estandarizadas es la posibilidad de que la puntuación de todo el mundo se encuentre por encima de la media. Con frecuencia me he encontrado con una declaración de objetivos indicando que todos los estudiantes, o la mayoría de ellos, puntuarán por encima de la media. Para muchas poblaciones de estudiantes y muchos distritos escolares, este objetivo carece por completo de realismo y solo se puede alcanzar si los enseñantes emplean enormes cantidades de tiempo enseñando de cara a la prueba o incluso a enseñar la prueba en sí. En esta época de gran importancia para las pruebas, cunde la desesperación cuando se establecen expectativas poco realistas. Por definición deberíamos esperar que, en grupos grandes de estudiantes distribuidos normalmente, el 50 por ciento de los que se sometan a una prueba estandarizada puntuarán por encima de la media y el otro 50 por ciento puntuará por debajo de ella. Si esperamos que todo el mundo puntúe por encima de la media, estamos fijando un objetivo ridículo e imposible.

MEJORAR LAS PRUEBAS EXISTENTES: ALGUNOS EJEMPLOS PROMETEDORES DE PRUEBAS RECIENTES

De las pocas pruebas que me han sido remitidas este año, he elegido tres como representativas de algunas tendencias prometedoras. Proceden de tres países angloparlantes distintos y todas tienen limitaciones. Quien considere utilizarlas debería realizar una revisión más detallada que la que puede incluirse aquí. Sin embargo, para los fines de este artículo, deseo destacar algunas de las características positivas de estas pruebas que podrían ser emuladas por otras.

El *Test of Early Reading Ability-2* (TETRA-2) (Reid, Hresko y Hammill, 1989), publicado en los Estados Unidos, intenta utilizar la investigación y los desarrollos teóricos recientemente generados en el campo de la aparición de la lectoescritura, para apartarse de las pruebas tradicionales de facilidad de lectura que fueron diseñadas para "predecir" si un niño tendría o no éxito en aprender a leer. Además, TETRA-2 intenta apartarse del concepto de aptitudes subyacentes como requisito previo para aprender a leer y, en cambio, intenta tomar una muestra de la capacidad que tienen los niños para abordar actividades que forman parte de los inicios de la lectura y la escritura. Por ejemplo,



en vez de comprobar la percepción visual o la discriminación auditiva, trata de ver si los niños pueden reconocer la escritura en comparación con dibujos o si están familiarizados con los conceptos de letra y palabra.

Puesto que esta prueba se administra mediante un único cuaderno de material encuadrado en espiral, algunos de los materiales y tareas no son auténticos y, en consecuencia, pueden conducir a conclusiones potencialmente engañosas. Los autores de la prueba hacen un sincero reconocimiento que les honra: "El usuario debe comprender que los elementos de TETRA-2, al estar presentados con un formato de prueba formal, necesariamente se encuentran fuera de contexto y no capturan el conocimiento de los niños de la misma manera que las observaciones naturalistas de sucesos relacionados con la lectoescritura, donde los intentos de comunicación social tienen una importancia extrema" (Manual de TETRA-2, p.4). El manual prosigue indicando que los usuarios de la prueba "pueden considerar conveniente" complementar los datos de esta prueba con observaciones; yo diría que los datos de la prueba *deben* ser complementados.

Francamente, preferiría basarme en las observaciones de un enseñante experto en aptitudes iniciales para la lectoescritura que en los resultados de una prueba como TETRA-2, pero al menos esta prueba intenta realmente sondear comportamientos iniciales de lectura que presentan alguna correlación con la lectura.

La prueba *Tests of Reading Comprehension* (TORCH) (Mossenson, Hi y Masters, 1987) procede de Australia y evita, como mínimo, seis importantes limitaciones de muchas pruebas estandarizadas utilizadas normalmente: a) dependencia excesiva de preguntas que requieren un recuerdo literal; b) abuso de elementos de opción múltiple y el requisito de una sola respuesta correcta; c) el empleo de trozos de texto breves y artificiales; d) requerir de los estudiantes que intenten leer y responder a elementos excesivamente difíciles; e) utilizar las puntuaciones por "nivel de curso" o "nivel de edad" tan fáciles de malinterpretar; y f) el empleo de límites de tiempo que pueden afectar negativamente a lectores lentos.

La prueba está diseñada para ser administrada a estudiantes con tres a diez años de escolarización y para que los enseñantes seleccionen pasajes que, a su entender, puedan ser leídos y respondidos por los estudiantes. Los alumnos leen pasajes, un equilibrio entre género novelístico y no novelístico, cuya longitud oscila entre 200 y 900 palabras. La comprensión se evalúa mediante la capacidad de transcribir lo leído por escrito (aunque se pueden dictar respuestas si es necesario) en forma de un procedimiento de palabras incompletas modificado. Los ejercicios de frases incompletas tradicionales requieren que los estudiantes rellenen palabras solas, pero como esta prueba requiere rellenas frases u oraciones, se incluyen espacios en blanco mucho más grandes y se insta a los estudiantes a escribir en el anverso de la hoja de la prueba o a pedir ayuda al enseñante si la respuesta que les gustaría dar a la prueba no cabe en el espacio disponible. Fue delicioso ver una clave de respuestas con varias respuestas "correctas". Las transcripciones constituyen una potente manera de llegar al



significado que los lectores construyen a partir del texto; hasta que no encontremos maneras de fomentar un empleo más extendido de transcripciones libres (orales o escritas), el formato utilizado por TORCH ofrece unas interesantes posibilidades para evitar algunas de las dificultades más comunes de las pruebas estandarizadas.

Las *Effective Reading Tests* (Vincent y de la Mare, 1989), que se publican en el Reino Unido, se encuentran entre las pruebas estandarizadas más interesantes que nunca he visto. También ellas evitan muchos de los inconvenientes evitados por la prueba TORCH, con la excepción de la excesiva dependencia de preguntas literales y del abuso de un formato de opción múltiple. Para la revisión solo recibí dos niveles de las pruebas: el Nivel 0 es adecuado para ser utilizado con niños de cursos equivalentes a segundo o tercer curso de EGB, y el Nivel 3A es adecuado para niños de cursos equivalentes a segundo, tercer y cuarto cursos de BUP. Lo más intrigante de estas pruebas es que los materiales que leen los niños no se parecen en nada a una prueba. Los materiales se encuentran en un libro separado e ilustrado a todo color que tiene el aspecto de un breve libro de bolsillo. Los estudiantes leen los relatos del libro y, después, en un cuadernillo separado, anotan sus respuestas a una serie de preguntas asociadas al relato, pudiendo consultar el texto cuando lo crean necesario.

El segundo aspecto destacado de esta prueba estandarizada es el tono relajado del Manual del Enseñante. Por ejemplo, el libro del Nivel 0 contiene dos fábulas y se sugiere que, como los enseñantes del Reino Unido tiende a organizar la lectura en grupos relacionados temáticamente, el enseñante "puede encontrar conveniente integrar la prueba con un programa de trabajo temáticamente relacionado, por ejemplo, cuentos y fábulas populares u otras historias de animales". El Manual prosigue expresando la esperanza de que esto contribuya a hacer de la prueba un suceso más natural. Imaginémosnos tratar de integrar la mayoría de las pruebas estandarizadas disponibles con un tema educativo en lectoescritura!

Se insta a los enseñantes a interactuar con los estudiantes mientras se administra la prueba, por ejemplo recordándoles que vuelvan a consultar la prueba en busca de respuestas cuando vean que no lo hacen. Aunque se advierte a los enseñantes que no lean el material a los alumnos, se les insta a prestar ayuda en cuanto a la manera de responder a las preguntas, por ejemplo recordando a los alumnos que no todas las respuestas se expresan de una manera directa en el texto. El Manual llega al extremo de destacar tres preguntas de la prueba que parecen provocar en los niños alguna dificultad inicial y propone maneras de clarificar los elementos de la prueba para los alumnos. Este manual ofrece una sensación muy distinta a la mayoría de los manuales de pruebas estandarizadas. Por ejemplo, incluye fotografías de niños que están realizando la prueba y de enseñantes y estudiantes interactuando durante su desarrollo. Esta prueba ofrece otras características innovadoras y positivas que se deberían utilizar en mayor medida en las pruebas estandarizadas.

SITUAR LAS PUNTUACIONES DE LAS PRUEBAS ESTANDARIZADAS EN UNA PERSPECTIVA MAS AMPLIA O BUSCAR ALTERNATIVAS

La columna dedicada a la evaluación de *The Reading Teacher* correspondiente a los meses de enero y febrero de este año incluía dis-



cusiones sobre un método de evaluación de la lectoescritura basado en carpetas. Durante los últimos meses he tenido la oportunidad de hablar con centenares de enseñantes y administradores sobre este concepto. Virtualmente ninguno de ellos creía factible reemplazar completamente las pruebas estandarizadas obligatorias existentes por carpetas, pero casi todos estuvieron de acuerdo en que los datos que se pueden recopilar en las carpetas pueden servir para complementar y situar en una perspectiva mejor los resultados de pruebas estandarizadas. Además de la puntuación percentil media, ¿no sería informativo para un distrito escolar y los ciudadanos de un distrito el hecho de saber, por ejemplo, la media de libros que los estudiantes han leído por su cuenta, el porcentaje de niños de tercer curso que han sido capaces de escribir un buen resumen de un libro que han leído, o el número de niños de sexto curso que han sido capaces de preparar con éxito un informe científico basado en recursos de biblioteca?

Ciertamente, debemos seguir destacando lo que tienen de malo las pruebas estandarizadas, pero también debemos empezar a dar pasos más activos para mejorar las pruebas que tienen la posibilidad de seguir siendo usadas, para situar sus resultados en una perspectiva más adecuada y para proponer mejores alternativas.

Referencias

- Mossenson, L., Hill, P. y Masters, G. (1987). *Tests of reading comprehension*. Melbourne: Australian Council for Educational Research.
- Neill, D. y Medina, N. (1989). Standardized testing: Harmful to educational health. *Pbi Delta Kappan*, 70, 688-702.
- Reid, D., Hresko, W. y Hammill, D. (1989). *Test of early reading ability-2*. Austin, TX: Pro-Ed.
- Vincent, D. y de la Mare, M. (1989). *Effective reading tests*. Londres: MacMillan Assessment.

El papel de las pruebas estandarizadas en un programa de evaluación de la lectoescritura

John J. Pikulski

CL&E, 1993, 19-20, 45-50

Datos sobre el autor: John J. Pikulski trabaja en la Universidad de Delaware

Artículo original: The role of tests in a literacy assessment program. En *The Reading Teacher*, Mayo de 1990, pp. 686-688. Traducción de Genis Sánchez. Reproducido con autorización de John J. Pikulski y de la International Reading Association (La I.R.A. no se responsabiliza de la adecuación de la traducción)

Dirección: Department of Educational Development, University of Delaware, Newark, DE 19716, USA

© De todos los artículos deberá solicitarse por escrito autorización de CL&E y de los autores para el uso en forma de facsímil, fotocopia o cualquier otro medio de reproducción impresa. CL&E se reserva el derecho de interponer acciones legales necesarias en aquellos casos en que se contravenga la ley de derechos de autor