

Una perspectiva sobre la evaluación formal e informal de la lectura

Lillian R. Putnam*

En la actualidad, existe gran confusión acerca de los valores y alcances de las pruebas estandarizadas y de las pruebas informales, por ello queremos clarificar algunos principios esenciales.

Características

Las pruebas estandarizadas se basan en centenares de alumnos que representan una amplia variedad de la población geográfica, cultural y económica. Los ítems de los exámenes se escogen de manera que los porcentajes predeterminados sean fáciles, de nivel medio o difíciles y se calculan los promedios para cada grado. Se responden seleccionando la contestación correcta, de modo que los exámenes son objetivos y pueden ser evaluados por medio de una tabla o clave de calificación. (La *Prueba de Aprovechamiento de California*, la *Prueba de Lectura de Iowa* y la *Prueba de Lectura de Gates-Mac Ginitie* son ejemplos de estos exámenes.)

Por otro lado, se presentan los procedimientos informales de evaluación. No se basan en cientos de alumnos, y, de hecho, no pueden ser normalizados. Frecuentemente, los exámenes preparados por los maestros persiguen una evaluación general. Como son administrados informalmente proveen mejores oportunidades para observar con cuidado al estudiante durante la situación de examen, para un diagnóstico más detallado y para discusiones posteriores al examen con los alumnos referidas a las preguntas efectuadas.

En educación, el péndulo tiende a moverse de un extremo a otro. Actualmente estamos en una era de crítica severa a los exámenes uniformes. Sin embargo, el concepto de que "nada es del todo bueno o del todo malo" se puede aplicar aquí. Debemos reconocer las ventajas y limitaciones de ambos tipos de exámenes y no desechar totalmente a ninguno de ellos.

Ventajas y limitaciones

1. Pruebas estandarizadas

Las pruebas uniformes presentan estos **aspectos positivos**: a) En la medida en que para su normalización se tuvo en cuenta un gran número de estudiantes representativos de distintos segmentos de la población, permiten comparar clases pequeñas o determinados grados de escuelas con las normas nacionales. b) Se administran por grupos, por lo tanto, son apropiadas para una población escolar grande. c) Las preguntas representan varios grados de dificultades y no están limitadas a un ámbito. d) Debido a que el contenido ha sido seleccionado en forma general se elimina el prejuicio del maestro. Por lo

* La doctora Lillian R. Putnam es directora de la Clínica de Lectura del Kean College de New Jersey.

tanto, las pruebas estandarizadas no deben ser abolidas totalmente como muchos de sus críticos proponen.

Tienen, sin embargo, serias **limitaciones**. Los exámenes individuales varían mucho y proveen resultados muy diferentes. Aunque bastante precisos en calificaciones de grupos, frecuentemente distorsionan la conducta o actuación del individuo y las calificaciones entonces reflejan un nivel de frustración más bien que un nivel de instrucción. El proceso de selección elimina del banco de preguntas muchas que los alumnos ya conocen. Durante el año enseñamos y enfatizamos ciertos conceptos y después examinamos algo diferente. Así, el contenido del examen muy rara vez refleja el currículo. Debido a que el alumno selecciona la contestación correcta entre tres o cuatro opciones dadas, la posibilidad de adivinar siempre prevalece.

Es también importante notar el número de preguntas en una sección subsidiaria. Un examen conocido tiene una sección denominada "destrezas de estudio", pero contiene solamente dos preguntas. Otra prueba apunta a evaluar una destreza específica y tiene solamente tres preguntas. Estos son ejemplos poco deseables.

Muchas pruebas estandarizadas presumen que el proceso de lectura puede ser fragmentado en destrezas específicas a fin de ser examinadas por separado. Investigaciones sobre este tópico muestran un éxito razonable en el aislamiento de destrezas sobre el reconocimiento de palabras pero muy escasos resultados en el aislamiento de destrezas sobre comprensión. Sin embargo, Davis (1944) aisló las siguientes destrezas de comprensión, que podemos enseñar y examinar: significado de palabras, recordación de datos, habilidad para organizar, habilidad para inferir, habilidad para reconocer la intención del autor y habilidad para captar la estructura de un párrafo. Los esfuerzos por separar otras destrezas comúnmente señaladas son espurios.

Los prejuicios culturales de los exámenes uniformes han sido documentados abundantemente en la literatura. A la autora se le presentó un ejemplo de esto cuando un 80% de una clase de tercer grado seleccionó la respuesta incorrecta en un ítem. Presentaba un cuadro de una vieja bomba de agua con un brazo extenso y las opciones eran: *cuchillo, agua, mesa, casa*. Al preguntar a los alumnos acerca de la ilustración se advirtió que ellos creyeron que representaba una navaja de barbero y, por lo tanto, seleccionaron "cuchillo".

Recientemente, un colega reportó una pregunta en la cual el 70% de los alumnos eligieron la respuesta incorrecta. El ítem exhibía un cuadro de un perro con la palabra "perro" debajo. Las distintas opciones que se presentaron eran: *diversión, correr, perro, alrededor*. Mediante preguntas a los alumnos se observó que ellos conocían la palabra "perro", pero seleccionaron la palabra "diversión" porque un perro es una diversión; "correr" porque el perro puede correr; "alrededor" porque el perro puede correr alrededor. Por lo tanto ésta no fue una prueba válida de lectura. A menudo no entendemos el proceso de pensamiento de los niños o ellos no captan el propósito de la prueba con más frecuencia de lo que nosotros nos damos cuenta.

Las preguntas relacionadas con hechos son las más fáciles de componer, de contestar, de corregir y de evaluar. No es, por consiguiente, nada sorprendente que la mayoría de las preguntas de una prueba estandarizada sea de este tipo. Niveles altos de reconocimiento, inferencia y lectura crítica son examinados menos corrientemente.

La mayor crítica a los exámenes uniformes no debe recaer sobre los exámenes mismos, sino sobre nosotros, los maestros que los usamos. Teniendo en cuenta que la calificación se acerca más al nivel de frustración que al nivel instruccional, no debemos usar estas calificaciones para ubicar a los niños en determinado nivel de lectura. En su lugar, tenemos que recordar que hay un margen de error uniforme en un examen y darnos cuenta de que una calificación de 4.5 puede fluctuar entre 4.0 y 4.9 ó más. Así, no se puede aceptar la calificación 4.5 como exacta sino tan sólo como un indicador, en un día, sobre un examen, con referencia a la conducta de un alumno. La mayoría de las pruebas uniformes son indicadores dudosos para ubicar a los alumnos en materiales de instrucción. En trabajo clínico, es usual presumir que el alumno está funcionando, probablemente, cerca de un año por debajo de la calificación de la prueba estandarizada.

Como los exámenes uniformes son calificados por medio de máquinas o, al menos, por medio de tablas o claves, los maestros raramente saben por qué el alumno escogió la contestación incorrecta, por qué cometió esos errores y por qué recibió tal calificación. Nosotros únicamente vemos el error en el producto; no vemos el error en el proceso, que es tan vital en la instrucción. Además, aunque todos los alumnos de una clase recibieran una calificación del 70%, diferirían indudablemente en cuanto a la necesidad de instrucción porque cada uno ha adquirido esa calificación a partir de diferentes destrezas.

Es también práctica habitual en muchas escuelas registrar las evaluaciones en un expediente acumulativo que se archiva en las oficinas. Por lo común quedan allí hasta asentar el otro examen y así los maestros dejan de usar esas evaluaciones por completo o fallan al emplearlas incorrectamente.

Debemos recordar asimismo que en años recientes los exámenes uniformes se han hecho más difíciles, aunque el nivel de grado permanece igual. Requiere un "mayor aprovechamiento" mantener el mismo nivel de grado escolar.

2. Pruebas informales

Estas pruebas presentan también ventajas y limitaciones. Generalmente consisten en evaluaciones fonéticas, listas graduadas de palabras e inventarios informales de lectura. Las **evaluaciones fonéticas** indican un conocimiento de las relaciones del fonema y el grafema en forma aislada. No tienen en cuenta el contexto. Las **listas graduadas de palabras** atienden al conocimiento de palabras ya vistas, y la síntesis de destrezas fonéticas al deletrear. Tales listas nos ayudan a indicar el nivel de comienzo para un inventario informal de lectura. La limitación obvia es que están aisladas, no en un contexto.

El **inventario informal de lectura** es el mejor de todos los instrumentos. Como está compuesto de párrafos con dificultades que van en aumento, el formato es menos amenazador psicológicamente que el típico examen uniforme. También ofrece la oportunidad de un éxito inicial. Los maestros pueden seleccionar inventarios publicados o crear sus propios inventarios basados en las series de lectura en uso.

La mayoría de los inventarios informales publicados contienen tipos de preguntas sobre hechos, inferencias y crítica. Dado que el niño tiene que redactar la respuesta, en vez de seleccionar una entre diferentes opciones, la posibilidad de adivinar se reduce. El maestro tiene una excelente oportunidad de observar durante el examen y aprende a ser un *observador del alumno* en ambos dominios, cognoscitivo y afectivo. Pero lo más positivo es que el inventario informal da la oportunidad de preguntar al alumno porqué escogió una respuesta particular o de explicar sus ideas. Muchas veces en mi experiencia, esto ha sido un proceso muy valioso. Una anécdota ilustra este punto. Ante la pregunta: "Si estás cuidando seis vacas y dos se van, ¿cuántas quedarán para llevar a tu madre?", el niño contestó: "Si dos vacas se van, yo no me atrevería a regresar a mi casa". En otro caso, el problema aritmético decía: "Si dos lápices cuestan 10 centavos, ¿cuánto cuesta uno? y el alumno contestó: "Bueno, yo no sé". Al principio, el maestro pensó que el alumno no sabía contestar, después de interrogarlo nuevamente el niño respondió que podían costar, 6 y 4, 5 y 5 centavos, etcétera. El diálogo reveló una evaluación más amplia que la simple respuesta de 5 centavos.

Los diagnósticos y evaluaciones deben constituir un proceso continuo, constante. El inventario informal con varias formas alternadas se presta más a este fin. Estas formas alternadas pueden ser para exámenes orales, lectura silenciosa y exámenes de comprensión sobre lo que se escucha.

Los inventarios de lectura informales tienen sus propias limitaciones. La mayor de estas limitaciones es la presunción de que las calificaciones son finales, terminantes y de que se puede depender de ellas, cuando no es así. Sin embargo, proveen ejemplos de comportamiento en un tiempo determinado, y por ende, pueden ser repetidos a intervalos.

La mayoría de los inventarios informales son diseñados para ser administrados oral e individualmente. Esto reduce su uso cuando se está frente a una población escolar grande. No obstante, los maestros con frecuencia los adaptan a exámenes de grupo, usando solamente la lectura silenciosa y respuestas escritas sobre preguntas relativas a comprensión. Como varios inventarios publicados se refieren a diferentes niveles de grado, deben ser coherentes y usar el mismo de manera que puedan compararse los resultados.

Se ha discutido considerablemente respecto al criterio a adoptar para los diferentes niveles. El viejo criterio establecido por Betts (1954) ha sido criticado como muy estricto. Él prefirió el 95% en reconocimiento de palabras y el 75% en comprensión al nivel de instrucción. Muchos de los presentes inventarios adoptan criterios menos severos. Como maestros, debemos decidir y seleccionar individualmente; pero es un imperativo el ser coherentes.

Cuando elegimos un inventario informal de lectura para ser usado en una escuela, tenemos que examinar cuidadosamente el criterio establecido para los niveles de instrucción y de frustración. Para el reconocimiento de palabras, simplemente hay que contar el número de errores "permitidos" en una selección a ese nivel de instrucción y dividirlo por el número total de palabras. Por ejemplo: 10 errores en un párrafo de 100 palabras es un 10% de errores o un 90% correcto. En la comprensión, debe dividirse el número de contestaciones incorrectas por el total de preguntas. Por ejemplo: dos errores en 10 preguntas son el 20% incorrecto o el 80% correcto. Como no existe una información estadística que pruebe la "corrección" de cualquier conjunto de criterios, el maestro debe usar su propio juicio. Sin embargo, es importante que sepa qué criterio adopta su examen y también si es severo o es fácil. Aún mejor, ¿el criterio del inventario informal muestra niveles de grado que ayudan para ubicar a los alumnos en materiales instruccionales? Es importante lograr tal cosa.

Tradicionalmente, los inventarios informales contaban las repeticiones como errores; algunos modelos actuales no lo hacen. Esto se debe a la creencia de muchos lingüistas de que la repetición ocurre cuando el alumno está reprocesando información para relacionarla con la historia. En otras palabras, es una clase de juego en falso o desacertado que él corrige. Mi preferencia personal es contar las repeticiones como errores si ocurren frecuentemente, ya que indican dificultades para procesar la información. Cuando ocurren a menudo, y no son contadas, el nivel de instrucción mostrado es falsamente alto, con el resultado de que el alumno es ubicado en material de lectura a un nivel de frustración.

¿Qué pasa con la lectura oral y silenciosa de los inventarios informales? Habitualmente el alumno lee en forma oral, a primera vista, un pasaje. Sin embargo, si la lectura silenciosa del mismo pasaje precede a la lectura oral el resultado es, a menudo, mayor (Brecht, 1976).

Uno de los principios cardinales de la educación es **motivar** al alumno. El examen constituye, sin embargo, una de las circunstancias en que se viola este principio. Los niños no son motivados para ser examinados. Las pruebas estandarizadas y los procedimientos informales son culpables de esto, aunque algunos inventarios informales proveen una oración motivadora o un propósito que precede a cada selección.

Es interesante notar que las correlaciones entre las calificaciones de las pruebas estandarizadas y los inventarios informales varían mucho dependiendo del nivel del grado y de la prueba usada específicamente. Los exámenes uniformes observados se relacionan entre un 70% y un 85% con los inventarios informales, mientras que otros tipos de pruebas estandarizadas difieren de un modo significativo.

Para el maestro de clase que se enfrente al dilema de la evaluación constante, un consejo juicioso es que obtenga el máximo de información de un mínimo de examen, aprovechando las ventajas de las pruebas estandarizadas y de las evaluaciones informales. La ventaja principal de las primeras es que permiten una comparación con las normas nacionales y la ventaja del segundo

es la oportunidad de obtener un diagnóstico mejor y una información más amplia a través de una observación cuidadosa.