

GLOSARIO DE TÉRMINOS QUE SE USAN FRECUENTEMENTE EN LAS PRUEBAS

Análisis de la calificación de cuartil (Quartile Score Analysis) – Proporciona el porcentaje de estudiantes cuyos rangos de percentil caen en cada uno de los cuatro campos de cuartil (1-25, 26-50, 51-75, y 76-99). En una distribución normal, se esperaría que el 25 por ciento de los estudiantes califiquen en cada uno de los campos de la gama de los cuartiles.

Artículo o Ítem (Item) – Otro término que se da a una pregunta de una prueba.

Calificación de escala (Scale Score) – Calificación sin procesar que se ha convertido a una escala estandarizada. Las escalas de calificaciones son idóneas para comparar diferentes niveles de pruebas o formularios de pruebas de la misma materia.

Calificación de la escala de desarrollo (Developmental Scale Score) – Medida que se usa para determinar el progreso anual de un estudiante de un grado al próximo. Por ejemplo, la escala de desarrollo de la *FCAT SSS* varía de 86 a 3008 para los grados comprendidos entre el tercero y el décimo.

Calificación de los grupos de contenido (Content Cluster Scores) – Número de puntos obtenidos en cada agrupación, grupo, o rama de declaraciones relacionadas que describen lo que los estudiantes debieran saber y poder hacer. Por ejemplo, en la *FCAT SSS* de Lectura se encuentran calificaciones por grupos de contenido de palabras y frases en contexto; una idea principal, argumento y objetivo; por comparaciones y causa/efecto y por referencia e investigación.

Calificación de número correcto (Number Correct Score) – Calificación que informa el número de puntos sin procesar que el/la estudiante obtuvo en cada una de las preguntas de una prueba.

Composición expositiva (Expository Writing) - Composición que proporciona información, que explica el cómo y el por qué, aclara un proceso o define un concepto.

Composición Narrativa (Narrative Writing) – Escrito que relata una experiencia personal o relata una historia basada en un evento real o imaginario.

Composición persuasiva (Persuasive Writing) – Escrito que trata de convencer al lector de que un punto de vista es válido o de que el/la lector/a tome una acción específica.

Estaninos (Stanines) – Representa la posición relativa del/de la estudiante con respecto a un grupo de referencia. Las calificaciones de los estaninos varían

entre la más baja de 1 a la más alta de 9, en la que el 5 representa el rendimiento promedio.

Fiabilidad (Reliability) – Hasta qué punto una prueba mide coherentemente lo que tiene que medir. Por ejemplo, la fiabilidad se puede medir al administrar la prueba al mismo grupo de individuos en dos ocasiones diferentes. Si la prueba es fiable, los resultados de ambas evaluaciones serán coherentes.

Introducción temática (Prompt) – Tema sobre el cual el estudiante ha de escribir una respuesta en forma de ensayo. En la FCAT de Composición+ (FCAT Writing+), la introducción temática tiene 2 partes: el tema del ensayo que van a escribir y las instrucciones de cómo escribirlo.

Materiales de pruebas protegidos (Secure Testing Materials) – Materiales sobre los que hay que dar cuenta antes, durante y después de las pruebas; mantener en un almacenamiento bajo llave cuando no estén en uso y que el personal de las escuelas y demás miembros del personal no pueden ver y/o copiar antes de su aplicación. Ejemplos de materiales de pruebas protegidos son los folletos de la prueba y las hojas de respuestas ya llenas de la *Prueba de FCAT SSS*.

Materiales de pruebas que no se protegen (Non-secure Testing Materials) – Materiales que el personal a cargo de la prueba u otros miembros del personal pueden ver antes, durante y/o después de la prueba. Ejemplos de materiales de pruebas que no se protegen son las pruebas desarrolladas por los/las maestros/as y las pruebas incluidas en los libros de texto.

Mediana (Median) – Calificación que identifica el valor central de un grupo de datos.

Modo (Mode) – Grupo de calificaciones que ocurren más frecuentemente en un conjunto de calificaciones.

Niveles de rendimiento (Achievement Levels) - Representan los niveles de habilidad basados en el juicio profesional de paneles integrados por educadores y ciudadanos. Por ejemplo, en la *FCAT – Estándares del Estado del Sol* (Sunshine State Standards o SSS, por sus siglas en inglés) hay cinco niveles de rendimiento en los que el 1 representa el nivel más bajo y el 5 representa el nivel más alto.

Pregunta con respuesta en cuadrícula (Gridded Response Item) – Pregunta que requiere que el/la estudiante responda llenando una respuesta numérica en una cuadrícula con números.

Pregunta de respuestas múltiples (Multiple Choice Item) – Pregunta que presenta al estudiante varias opciones de las cuales escoger la respuesta correcta.

Promedio (Mean) – Promedio aritmético de un grupo de calificaciones que se calcula al sumar todas las calificaciones juntas y dividir por el número de calificaciones de un grupo.

Prueba estandarizada (Standardized Test) – Prueba en la cual las instrucciones, los límites de tiempo, los materiales y los procedimientos de calificación están formulados a fin de que se mantengan constantes cada vez que se aplique la prueba con el objeto de garantizar la comparabilidad de las calificaciones. Las pruebas estandarizadas pueden estar relacionadas con criterios o relacionadas con normas.

Prueba estatal obligatoria (State Mandated Test) – Prueba que requiere el estado de la Florida. En algunos casos, el estado requiere una prueba en particular, por ejemplo, la *FCAT SSS*. En otros casos, el estado requiere un tipo de prueba y el distrito selecciona o elabora una prueba en particular, por ejemplo, la carpeta de lectura del estudiante de tercer grado, que se aplica a los estudiantes de tercer grado durante el segundo semestre de cada curso escolar, así como una prueba relacionada con la normas, la Evaluación Alternativa para la Promoción al 3er Grado, que se aplica en el verano de cada curso escolar a todos los estudiantes de tercer grado a los que se les ha retenido.

Prueba obligatoria nacionalmente (Nationally Mandated Test) – Prueba que el gobierno federal requiere. Un ejemplo de prueba obligatoria nacionalmente es la Evaluación Nacional del Progreso Educativo (National Assessment of Educational Progress, NAEP, por sus siglas en inglés).

Prueba relacionada con criterios (Criterion Referenced Test) – Prueba que tiene el propósito de medir hasta qué punto el estudiante ha aprendido destrezas específicas relacionadas con áreas de contenido. Ejemplos de pruebas relacionadas con criterios, también conocidos como CRTs incluyen las *FCAT SSS* de lectura, matemáticas, ciencias, y composición.

Prueba relacionada con normas (Norm Referenced Test, o NRT por sus siglas en inglés) – Prueba que compara el desempeño del/de la estudiante con el modo en que otros estudiantes de un grado normativo se desempeñan en la prueba. Un ejemplo de una prueba relacionada con normas, es la *Prueba de Aprovechamiento Stanford (Stanford Achievement Test, SAT, por sus siglas en inglés)*.

Punto de referencia (Benchmark) – Declaración específica, como parte de los Estándares del Estado del Sol, que describe lo que el estudiante debiera saber y poder hacer.

Rango de la mediana del percentil (Median Percentile Rank) - Punto medio de un grupo de calificaciones de rango percentil.

Rango de percentil (Percentile Rank) – Indica el porcentaje de un grupo de referencia que tiene calificaciones iguales o inferiores a la calificación obtenida por un individuo. Dicho rango indica la posición relativa de un estudiante en comparación con las de estudiantes del mismo grado que hayan tomado la prueba. Las clasificaciones de percentil pueden utilizarse para informar el desempeño del estudiante en todas las asignaturas.

Rúbrica (Rubric) – Instrumento para calificar o una serie de criterios utilizados a fin de evaluar la respuesta de un/a estudiante a un artículo (o ítem) de respuesta estructurada.

Tarea de desempeño (Performance Task) – Pregunta que requiere que los estudiantes escriban una respuesta en lugar de seleccionar una respuesta entre varias opciones o llenar una respuesta en una cuadrícula con números. En la *FCAT SSS* se emplean dos tipos de tareas de desempeño que difieren en términos de la extensión o complejidad de la respuesta esperada (preguntas de respuesta breve y preguntas de respuesta extensas).

Validez (Validity) – Grado al cual una prueba mide en realidad lo que se propone medir. Por ejemplo, la prueba *FCAT SSS* debe medir la competencia en las Normas del Estado del Sol directa y fiablemente para ser una medida válida del programa de estudios. Para que una prueba sea válida, primeramente debe ser fiable.