# Evaluación escrita

Recursos de evaluación del Consejo Australiano para la Investigación Educativa 8



**SERIE: DOCUMENTOS TÉCNICOS** 

# EVALUACIÓN ESCRITA RECURSOS DE EVALUACIÓN DEL CONSEJO AUSTRALIANO PARA LA INVESTIGACIÓN EDUCATIVA 8





Serie Documentos Técnicos, 26

#### Consejo Directivo Ad Hoc

Carolina Barrios Valdivia, Presidenta Fabiola León-Velarde Servetto Daniel Alfaro Paredes

#### Secretaría Técnica

Haydee Chacón Cabanillas (e)

#### Cuidado de la edición

Dirección de Evaluación y Gestión del Conocimiento Verónica Alvarado Bonhote, Directora Diana Zapata Pratto, Especialista en Gestión de Publicaciones

#### Traducción

Brenda Hernández

#### Maquetación

Ángel García Tapia

Se terminó de imprimir en noviembre de 2017 en: EDITORIAL SUPER GRÁFICA E.I.R.L. Calle Luisa Beausejour 2049 Cercado de Lima

Hecho el Depósito Legal en la Biblioteca Nacional del Perú N.º 2017-14524 ISBN N.º 978-612-4322-28-0

**Tiraje:** 500 ejemplares Primera edición Lima, noviembre de 2017

#### © Sistema Nacional de Evaluación, Acreditación y Certificación de la Calidad Educativa

Calle Manuel Miota N.º 235 - San Antonio, Miraflores, Lima 18, Perú Teléfonos: (+51 1) 637-1122, (51-1) 221-4826, (51-1) 221-4807 anexo 108 E-mail: sir@sineace.gob.pe / www.sineace.gob.pe

De la edición inglesa:

© The Australian Council for Educational Research Ltd (ABN 19 004 398 1450) 2016

© El Consejo Australiano para la Investigación Educativa Ltd. (ABN 19 004 398 1450) 2016

Título original: Paper & Pen. Assessment Resource Kit

Autor: Margaret Forster y Geoff Masters

Translated by permission of the Australian Council for Educational Research Ltd (ABN 19 004 398 1450). All rights reserved. Traducido con autorización del Consejo Australiano para la Investigación Educativa Ltd. (ABN 19 004 398 1450). Todos los derechos reservados.

Se autoriza la reproducción total o parcial solo para propósitos educativos dentro del territorio nacional. Distribución gratuita. Prohibida su venta.

# ÍNDICE

Presenta	ación	7
Qué و 1.	es una evaluación escrita?	11
1.1.	Propósitos y contextos	12
1.2.	Evaluaciones del desarrollo	17
1.3.	Utilizar evaluaciones escritas para la evaluación del desarrollo	18
1.4.	Diseñar los contextos para las evaluaciones de los resultados de aprendizaje	19
1.5.	Una variedad de formatos de respuesta	24
1.6.	Calificar y registrar las respuestas de los estudiantes	25
1.7.	Estimar y reportar ubicación en el mapa de progreso	30
1.8.	Múltiples fuentes de evidencia	31
1.9.	Siguientes artículos	31
2. Una v	ariedad de formatos de respuesta	34
2.1.	Procedimiento para rellenar espacios en blanco	35
2.2.	Mapas de conceptos	36
2.3.	Respuestas construidas	37
2.4.	Formato de ensayo	37
2.5.	Investigaciones	38
2.6.	Escalas Likert	39
2.7.	Items correspondientes	39
2.8.	Formato de opciones múltiples	
2.9.	Parafraseo	43
2.10.	Formatos de respuestas cortas	44
2.11.	Autorreflexiones	46
3. Evalu	ación escrita diseñada por el profesor	48
3.1.	Planeamiento de la evaluación escrita	49

3.2.	Construyendo evaluaciones: algunos ejemplos	59
3.3.	Calificar y registrar	70
3.4.	Guía de puntaje	72
3.5.	Categorías de diagnóstico	74
3.6.	Escalas de puntuación	75
3.7.	Diferentes evaluadores	80
3.8.	La necesidad de comparar las calificaciones	81
3.9.	Estimar y reportar el logro	85
3.10.	Calificar estrategias de evaluaciones y reporte	90
4. Pruek	oas desarrolladas profesionalmente	92
4.1.	Planeamiento de medición	93
4.2.	Escribir preguntas	97
4.3.	Estandarizar condiciones de prueba	99
4.4.	Marcado estandarizado	102
4.5.	Interpretar el resultado de las pruebas	108
5. Diseñ	ando la evaluación escrita	113
5.1.	Planificación de evaluaciones escritas	114
5.2.	Calificación y registro	116
5.3.	Resumir el trabajo del estudiante	117
5.4.	Informando el trabajo del estudiante	118
5.5	Un resumen del proceso de diseño de una evaluación escrita	119

# **PRESENTACIÓN**

Entre los años 2009 y 2015, el SINEACE estuvo abocado a la elaboración de estándares de aprendizaje, con la finalidad de contribuir con el Ministerio de Educación y diversos actores a lograr mejores aprendizajes en los estudiantes de la educación básica regular; durante dicho proceso, se pudo conocer la experiencia australiana de evaluación de los aprendizajes realizada por el Consejo Australiano para la Investigación Educativa (ACER). Su trabajo inspiró la elaboración de los estándares de aprendizaje en forma de mapas de progreso, lo que significó un cambio de paradigma en la manera de enseñar, toda vez que no indican tanto qué debe enseñarse, sino qué debe aprender un estudiante.

Finalizado el encargo, nació la idea de traducir los folletos que comprendían el Kit de Recursos de Evaluación del ACER, que habían sido un valioso aporte durante todo el proceso. Para ello, el SINEACE suscribió con este un convenio que hoy permite poner al alcance de los lectores de habla castellana el octavo y último número de la serie, el cual rescata la importancia de la evaluación escrita como herramienta para recolectar evidencia que permita evaluar el desarrollo de los estudiantes.

Una evaluación escrita es un método común de evaluación que consiste en presentar a los estudiantes una serie de preguntas y utilizar sus respuestas escritas como evidencia de sus niveles de conocimiento, competencia, comprensión o actitud de forma individual. Las evaluaciones escritas son

utilizadas para una variedad de propósitos, incluyendo el diagnóstico de las fortalezas y debilidades especiales del estudiante, la evaluación del manejo de habilidades de este, la selección para cursos con vacantes limitadas, posibilidades de certificación y monitoreo de estándares educacionales.

La primera parte de la publicación describe la variedad de evaluaciones escritas y con qué propósitos se utilizan:

- Las evaluaciones formativas se aplican en el curso de la enseñanza diaria. Pueden ser diseñadas o adquiridas por el profesor, o ser pruebas desarrolladas profesionalmente para reportar al programa de enseñanza.
- Las evaluaciones sumativas se realizan al final de un curso o semestre para "resumir" el nivel de logro del estudiante. Estas evaluaciones pueden incluir también pruebas desarrolladas profesionalmente.
- Las evaluaciones sumativas externas se llevan a cabo al final del curso para determinar si los estudiantes pueden ganar certificaciones, obtener becas o plazas en cursos con vacantes limitadas. Las evaluaciones de este tipo usualmente son administradas por una autoridad fuera de la institución en la que se realizó el estudio.

La segunda parte explora el diseño de una evaluación escrita, el cual incluye:

- consideraciones al planear evaluaciones escritas (validación, fiabilidad y legitimidad);
- las fortalezas y debilidades de diferentes formatos escritos;
- cómo escribir una buena evaluación escrita;
- cómo calificar y registrar las respuestas de los estudiantes;
- cómo estimar los niveles de rendimientos del estudiante; y
- cómo reportar el logro.

La tercera parte explora las consideraciones comunes en el desarrollo y uso de pruebas escritas diseñadas profesionalmente, incluyendo las siguientes:

- los productos en preguntas de escritura;
- el procedimiento para condiciones de pruebas estandarizadas;
- el procedimiento para puntuación estandarizada; y
- las formas de interpretar los resultados de las pruebas.

Finalmente, se brinda una revisión completa del proceso de diseño de las pruebas escritas. Una lista de verificación resume con detalle los pasos seguidos en el proceso.

En la evaluación del desarrollo, las evaluaciones escritas son utilizadas en conjunto con otros métodos para brindar múltiples perspectivas en la comprensión de los estudiantes. El SINEACE difunde este libro con el deseo de aportar conocimientos acerca de las evaluaciones escritas como herramienta útil para evaluar el desarrollo de los estudianrtes, lo que resulta de gran relevancia para los actores educativos en su conjunto.

Consejo Directivo Ad Hoc SINEACE

# EVALUACIÓN ESCRITA RECURSOS DE EVALUACIÓN DEL CONSEJO AUSTRALIANO PARA LA INVESTIGACIÓN EDUCATIVA 8

## 1. ¿QUÉ ES UNA EVALUACIÓN ESCRITA?

Un método común de evaluación es presentar a los estudiantes una serie de preguntas o consultas y utilizar sus respuestas escritas como evidencia de niveles de conocimiento, competencia, comprensión o actitud de forma individual. Nos referimos a este tipo de evaluaciones como evaluaciones "de papel y lápiz".

Se han desarrollado varios tipos de preguntas como fuentes de evidencia sobre el logro o actitud. Estos tipos de preguntas incluyen, pero no están limitados a, preguntas de múltiple respuesta, preguntas con respuestas cortas, preguntas con respuestas largas, preguntas de verdadero y falso, pruebas para llenar espacios en blanco, preguntas de ensayo, cuestionarios estilo Likert y consultas de autorreflexión.

Al responder las preguntas, los estudiantes no siempre escriben sus respuestas. Algunas veces, marcan un círculo o formas, marcan posiciones en líneas y elaboran diagramas.

De hecho, la primera vez que se utilizó este tipo de preguntas, los estudiantes tal vez no usaban lápiz y papel, sino que probablemente registraban sus respuestas en pergaminos o pizarras utilizando crayones, lápices o tizas.

Y con la llegada de la tecnología computarizada, los tipos de preguntas discutidos en esta publicación pueden ser presentados en pantallas, con los estudiantes respondiendo a través del teclado y sus respuestas (por ejemplo, ensayos) serían impresas para la evaluación o calificadas automáticamente.

Sin embargo, como en la mayoría de los salones los estudiantes responden actualmente a preguntas escritas y consultas en papel, hemos escogido continuar refiriéndonos a las evaluaciones basadas en este rango de tipo de preguntas como evaluaciones de "papel y lápiz" o evaluaciones escritas.

La característica que define el método de evaluación discutido en este artículo es que los estudiantes respondan a una serie de preguntas o consultas.

#### Otros métodos de evaluación

Otros métodos de evaluación pueden involucrar el uso de papel, lápiz, crayones, lapiceros o computadoras:

Los proyectos de los estudiantes completados a lo largo de un periodo de tiempo y que involucran la recolección, organización, evaluación y presentación de material o información, normalmente, son más indefinidos que los métodos de evaluación discutidos aquí.

La evaluación de los proyectos es mostrada en el número de Recursos de Evaluación sobre PROYECTOS.

La recolección del trabajo del estudiante (ejemplos, portafolios de su escritura) puede ser utilizada como base para la evaluación. La evaluación de los portafolios es mostrada en el número de Recursos de Evaluación sobre PORTAFOLIOS.

Los dibujos y pinturas realizados en un papel son utilizados en la evaluación de logro de arte y dibujo técnico. La evaluación de arte producido por los estudiantes es mostrada en el número de Recursos de Evaluación sobre PRODUCTOS.

Las simulaciones basadas en computadoras (por ejemplo, simulaciones de vuelo para la evaluación de pilotos en entrenamiento) las describimos como desempeño. La evaluación del desempeño del estudiante es mostrada en el número de Recursos de Evaluación sobre EVALUACIÓN DE DESEMPEÑO.

## 1.1. Propósitos y contextos

Las evaluaciones escritas son utilizadas para una variedad de propósitos, que incluyen el diagnóstico de las fortalezas y debilidades especiales del estudiante; la evaluación del manejo de habilidades del estudiante, selección para cursos con acceso limitado, certificación de habilidades y el monitoreo de estándares educacionales.

Los contextos de evaluaciones de papel y lápiz incluyen evaluaciones formativas, evaluaciones sumativas en el salón y evaluaciones sumativas "externas".

Las evaluaciones formativas ocurren en el curso de la enseñanza diaria. Estas pueden ser diseñadas por el profesor o ser adquiridas, y estar desarrolladas profesionalmente para informar al programa de enseñanza. Las evaluaciones formativas algunas veces son llamadas evaluaciones de "bajo rendimiento".

Las evaluaciones sumativas en el salón de clase ocurren al final de un curso o semestre para "resumir" el nivel de logro del estudiante. Estas evaluaciones también pueden incluir pruebas desarrolladas profesionalmente. La audiencia para resultados de evaluaciones sumativas incluye profesores, estudiantes y padres de familia.

Las evaluaciones sumativas externas ocurren al final del curso para determinar, por ejemplo, qué estudiantes pueden ganar certificaciones, obtener becas o lograr puestos. Las evaluaciones sumativas de este tipo usualmente son administradas por una autoridad fuera de la institución en la que el estudiante ha sido enseñado. Estas evaluaciones son algunas veces llamadas evaluaciones de "alto rendimiento".

Un profesor que utiliza una prueba profesionalmente desarrollada, estandarizada para evaluar las fortalezas y debilidades de un estudiante en ortografía, para poder planear un programa de ortografía individual, está Algunos propósitos de las evaluaciones de papel y lápiz:

- diagnosticar
   las fortalezas y
   debilidades de los
   estudiantes,
- evaluar los conocimientos, habilidades y entendimientos de los estudiantes,
- certificación,
- selección,
- estándares de monitoreo.

utilizando una evaluación escrita con propósitos diagnósticos, en el contexto de evaluación formativa del salón. Un ejemplo de diagnóstico de ortografía utilizado para este propósito se muestra en la página 15.

Un profesor que administra una prueba desarrollada profesionalmente para brindar información a los estudiantes de idioma extranjero sobre sus niveles de logro en un segundo idioma, utiliza pruebas de papel y lápiz para propósitos de evaluación, en el contexto de evaluaciones sumativas en el salón de clases. En la página 16 se muestra un ejemplo de una pregunta de una prueba de lectura para los estudiantes que aprenden indonesio como un idioma extranjero.

Una universidad que administra una evaluación escrita de razonamiento moral como parte de un instrumento de filtración para seleccionar estudiantes que ingresen a la escuela de medicina, utiliza las pruebas de papel y lápiz para propósitos de selección en un contexto externo de alto rendimiento. En la página 16 se muestra un ejemplo de una pregunta de dicha evaluación.



# Utilizar evaluaciones de papel y lápiz para propósitos de evaluación formativa en clases

Los profesores que utilizan la evaluación de dictado para diagnosticar la ortografía en Reino Unido administran una prueba de dictado gradual a los estudiantes. Ellos categorizan las respuestas de los estudiantes palabra por palabra en un "cuadro" de diagnóstico. Se utilizan tres categorías de ortografía: "plausible", "inventado" y "aleatorio".

Los estudiantes cuya ortografía es en gran parte aleatoria (ortografía que no se asemeja ni en sonido o estructura al inglés) o inventada (ortografía que se inventa desde el sonido de una palabra, con un poco o nada de referencia a las secuencias de las letras en inglés) reciben clases especiales.

A continuación, se muestra una sección de "Cuatro Dictados" y las respuestas categorizadas de los estudiantes. El Dictado Cuatro está diseñado para niños de 6.ºº año en Inglaterra y Gales, P7 en Escocia e Y7 en Irlanda del Norte.

"Una silueta extraña se acercaba desde el valle sureño. Los niños que estaban ahí notaron un aeroplano de diseño avanzado dando vueltas. La máquina aterrizó con precisión en la región montañosa sin siguiera arañar la superficie...".

PLAUSIBLE		INVENTADO	AL AZAR	
Legible	l legible			
Palabras que se ajustan a la ortografía en castellano y que son legibles en el contexto del pasaje.	Palabras con algún parecido estructural a la palabra de estímulo, que pueden no ser legibles en el contexto del pasaje.	Palabras inventadas del sonido de la palabra, con poca o ninguna referencia a secuencias de letras en castellano.	Palabras que no tienen semejanza, ya sea en sonido o estructura, con el castellano.	
aero o plano región (región)	precion   (precisión)   sileta (silueta)   máqina   (máquina)	montosa (montañosa) supricia (superficie) aere llano (aeroplano)	nons (niños) plasino (precisión) arinmó (aterrizó) mulsralas (montañosa)	

Peters, M. y Smith, B. (1993). Spelling in Context Strategies for Teachers and Learners, pp. 70 y 79. Berkshire: NFER-Nelson.

# Utilizar evaluaciones de papel y lápiz para propósitos sumativos del salón de clases

En el programa de Certificación del Banco de Lenguaje Nacional de Australia, los estudiantes de Australia y Nueva Zelanda completan una serie de tareas de lectura en uno de tres niveles: Principiante, Inicial o Intermedio.<sup>2</sup> El nivel inicial está diseñado para estudiantes de nivel secundario que han estudiado la lengua meta entre 80 a 200 horas. Esta evaluación sumativa es administrada por los profesores del salón y permite certificarse para animar a los estudiantes en su aprendizaje de otro idioma. Las evaluaciones son calificadas por personal capacitado. Los estudiantes reciben certificados indicando su nivel de logro.

La pregunta a continuación se toma de una evaluación de nivel Inicial en indonesio.

Vestimenta de las mujeres de Minangkabau

Kalau di Jawa para ibu memakai kain dan kebaya di Sumatera meeka biasanyamemakai sarung dan kebaya panjang. Sarung mereka adalah "kain songket" dengan sedikit benang perak atau emas.

"Kain son	gket" es bonito porque está decorado con:
	batik
	lentejuelas
	bordados
	hilos dorados o plateados

# Utilizar evaluaciones de papel y lápiz para evaluaciones externas de alto rendimiento

Obtener un alto puntaje académico no garantiza que los estudiantes obtengan un lugar en la escuela de medicina en las Universidades de Newcastle y Melbourne.<sup>3</sup> Tanto como el logro académico, los estudiantes también necesitan obtener un buen puntaje en la prueba de "perfil personal" en el que responden de manera escrita a preguntas que solicitan consideraciones éticas.

A continuación, se muestra un ejemplo de una respuesta corta de una evaluación de perfil personal.

Un hombre roba dinero para comprarle a su esposa una máquina de diálisis que ella necesita desesperadamente. Usted es un doctor. ¿Cuál es su punto de vista de este crimen?

<sup>&</sup>lt;sup>2</sup> Australian Council for Educational Research (1998). Natural Australia Bank Language Certificates Beginners Level Indonesian. Camberwell: ACER.

<sup>&</sup>lt;sup>3</sup> Australian Council for Educational Research.



Es útil pensar sobre estos diferentes contextos y propósitos para las evaluaciones escritas a lo largo de un "propósito" continuo, como se muestra aquí. La evaluación formativa del salón (incluyendo la evaluación de diagnóstico) se encuentra a la izquierda del continuo; "alto rendimiento", evaluación externa, a la derecha. Las evaluaciones sumativas del salón de clases se encuentran cerca al centro del continuo.

#### 1.2. Evaluaciones del desarrollo

Las evaluaciones del desarrollo consisten en el proceso de monitorear el progreso de los estudiantes a través de un área de aprendizaje, para que se puedan tomar decisiones sobre las mejores formas de facilitar el aprendizaje.

El centro de la evaluación del desarrollo es el mapa de progreso. Un mapa de progreso describe el conocimiento, habilidad, comprensión, actitudes o valores que el estudiante desarrolla en un área de aprendizaje, en el orden en el que ellos se desarrollan normalmente. Las descripciones a lo largo del mapa son llamadas algunas veces "resultados de aprendizaje".

En la evaluación del desarrollo, el progreso del estudiante en un área de aprendizaje es monitoreado sobre el resultado de aprendizaje descrito en un mapa de progreso.

La evaluación del desarrollo utiliza un rango de métodos para recolectar evidencia del logro del estudiante, incluyendo portafolios, proyectos,

productos, desempeño y evaluaciones de papel y lápiz. Cada método es apropiado para recolectar tipos particulares de evidencia, esto es, para obtener algunos resultados específicos. Por ejemplo, la evaluación del desempeño se utiliza particularmente para evaluar los resultados de "esfuerzo", resultados para los cuales es necesario observar a los estudiantes en acción. La evaluación del proyecto, la evaluación de trabajos extensos completados a lo largo de un periodo de tiempo (involucrando la recolección, organización, evaluación y presentación de material o información) es utilizada para evaluar conocimiento y habilidades específicas del sujeto, y habilidades generales como escribir un reporte y trabajar como miembro de un grupo.

## 1.3. Utilizar evaluaciones escritas para la evaluación del desarrollo

Las evaluaciones escritas son útiles para recolectar evidencia sobre un amplio rango de resultados de aprendizaje en diferentes áreas curriculares. En el cuadro inferior se muestran los ejemplos de resultados de logros en los que se puede evaluar utilizando evaluaciones escritas.

# Ejemplos de resultados de logros que pueden ser obtenidos utilizando evaluaciones escritas

- Comprende y aplica propiedades básicas y avanzadas de funciones y álgebra (matemáticas).
- Demuestra competencia en habilidades generales y estrategias del proceso escrito (lenguaie).
- Comprende las premisas y paradojas de la segunda mitad del siglo 20 (historia global).
- Utiliza procesos básicos de la palabra, hojas de cálculo, bases de datos y programas de comunicación (conocimientos básicos).
- Conoce el tipo de fuerzas que existen entre objetos y dentro de los átomos (ciencias físicas)
- Conoce los aspectos de uso y abuso de sustancias (salud).4

<sup>&</sup>lt;sup>4</sup> Kendall, J. y Marzano, R (1996). *Content Knowledge: A Compendium of Standards and Benchmarks for K-12.* Aurora CO: Education Mid-continent Regional Laboratory.

# 1.4. Diseñar los contextos para las evaluaciones de los resultados de aprendizaje

Las evaluaciones escritas pueden ser utilizadas para ensamblar evidencia sobre los logros en un mapa de progreso. Debido a que los resultados normalmente están formulados de forma general (por ejemplo, "escribir historias o ensayos que muestran conocimiento del propósito previsto" 5), es posible diseñar un numero casi ilimitado de tareas de evaluación para el propósito de recolectar evidencia.

Las tareas a ser evaluadas brindan contextos específicos para recolectar evidencia sobre resultados generales. Nosotros utilizamos el desempeño de un estudiante en un contexto específico para recolectar evidencia sobre el resultado general. Lo que significa, contextos particulares en los cuales las observaciones realizadas no son nunca de interés en sí mismas: estas son solo de interés en el sentido de que proveen información sobre la dimensión a ser evaluada. Como indica Ramsden (1993): "Los educadores están interesados en qué tan bien los estudiantes comprenden la velocidad, distancia y tiempo, no en cuánto conocen sobre los corredores o botes o las personas que caminan en los corredores. Sin embargo, paradójicamente, no hay otra forma de describir y probar la comprensión más que a través de estos ejemplos específicos".6

En las páginas 21-22 se muestran algunos ejemplos de contextos específicos para recolectar evidencia sobre resultados generales.

El primer ejemplo en la página 21 ilustra un contexto específico para evaluar un resultado general "comprensión de relación entre velocidad, distancia y tiempo para objetos móviles". En esta tarea se pide a los estudiantes demostrar

<sup>&</sup>lt;sup>5</sup> Kendall, J. y Marzano, R. (1996: 294).

<sup>&</sup>lt;sup>6</sup> Ramsden, P.; Masters, G.; Stephanou, A.; Walsh, E.; Martin, E.; Laurillard, D. y Marton, F. (1993). Phenomenographic research and the measurement of understanding: An investigation of students' conceptions of speed, distance and time. *International Journal of Educational Research*, 13, pp. 301-316.

su comprensión analizando un gráfico que muestra la distancia recorrida durante una carrera a lo largo del país con medición de tiempo.

El segundo ejemplo en la página 22 ilustra un contexto para evaluar la comprensión del estudiante de cómo un puntaje adicional se relaciona con el promedio de un conjunto de puntajes. En esta tarea se pide a los estudiantes ayudar al "Dr. Matemáticas" a responder una pregunta numérica desconcertante. La respuesta a la pregunta ayudará a "David el curioso" a evitar completar una parte del trabajo.

El tercer ejemplo en las páginas 23-24 ilustra un contexto para evaluar el resultado de escritura "producción de textos de forma fluida y con estilo legible, y uso de tecnología para presentarlos efectivamente en una variedad de formas". Se pide a los estudiantes desarrollar una explicación escrita sobre un aspecto del tema estudiado y redactarla en la pantalla de la computadora.



# Desarrollar un contexto para direccionar un resultado de aprendizaje general

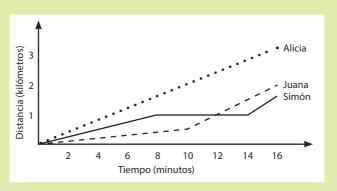
## **Ejemplo 1: Ciencia**

Esta tarea brinda contextos específicos para recolectar evidencia sobre resultados generales. Esta tarea de ciencia dirige el siguiente resultado: "describe la relación entre velocidad, distancia y tiempo para objetos móviles".

Contexto para observación:

Una carrera a campo traviesa

El siguiente gráfico muestra la distancia recorrida por los estudiantes durante una carrera a lo largo del país.



1. Complete la siguiente tabla para mostrar qué tan lejos ha viajado cada estudiante en 10 minutos.

#### Nombre

#### Distancia recorrida

Alicia

Juana

Simón

2. Para cada estudiante, describa los cambios en velocidad durante el recorrido.

Alicia

.

Simón\_

3. Realice su propio grafico de distancia sobre tiempo para dos dueños de mascotas que pasean a sus perros en el parque. Muestre y etiquete a los dueños a) desplazándose despacio, b) desplazándose rápido, y c) en descanso. Utilice el reverso de la hoja.

Andrews, C. et al. (1998). Effective assessment for Science, p. 115. Carlton y South Melbourne: Board of Studies and Addison Wesley Longman Australia Pty Limited.

## **Ejemplo 2: Matemáticas**

La tarea brindada provee contextos específicos para recolectar evidencia sobre resultados generales. Esta tarea de matemáticas dirige el resultado: "muestra comprensión de como un puntaje adicional se relaciona al promedio de un conjunto de puntajes".8

Contexto para observación:

Ayude al Dr. Matemáticas a responder la siguiente pregunta:

Estimado Dr. Matemáticas:

Todos nuestros proyectos de matemáticas fueron puntuados con 20. Luego de cuatro proyectos, mi puntaje promedio es 13. Si puedo subir mi promedio hasta 15 con un buen puntaje en mi quinto proyecto, no tendré que hacer un sexto. ¿Qué puntaje necesito obtener? Por favor explíqueme cómo obtuvo el resultado.

Sinceramente,

David el curioso.

Estimado David el curioso:



Beesey, C. et al. (1998). Effective assessment for Mathematics, p. 99. Carlton y South Melbourne: Board of Studies and Addison Wesley Longman Australia Pty Limited.

## **Ejemplo 3: Escritura**

La tarea brindada provee contextos específicos para recolectar evidencia sobre resultados generales. Esta tarea de escritura dirige el resultado: "produce textos en un estilo fluido y legible y utiliza la tecnología para presentarlo de forma efectiva en una variedad de formas".

Contexto para observación:9

Los estudiantes discuten el formato y estructura de una explicación en relación a una unidad en "agua". Luego de investigar sobre el tema con un compañero ellos construyen una explicación independientemente en la pantalla.

A continuación, se muestra el borrador final de un estudiante.

#### ¿Cómo ocurren las inundaciones?

Por Miguel

Las inundaciones son un fenómeno natural que ha ocurrido por cientos de años. Las inundaciones son conocidas por causar devastación a los ambientes naturales y construidos. Muchos han causado millones de vidas y han dejado miles de personas sin hogar, también han destruido plantaciones de maíz.

Una de las fechas más comunes para las inundaciones es primavera. En primavera, por supuesto, la nieve se derrite. Cuando la nieve se vuelve agua, baja por las montañas causando inundaciones. Esta es una de las causas principales para una crecida, una inmensa cantidad de agua corriendo a una gran velocidad. Estas inundaciones han causado daño devastador, pero afortunadamente desaparecieron hace 2 a 3 años.

Otra causa común de una inundación es mucha cantidad de agua. Esto ocurre cuando un río o canal es bloqueado por una represa o peñasco. Si hay mucha lluvia el agua no puede fluir y pasar la represa y en ese momento el agua sobrepasa su límite. Ese es otro momento cuando ocurren las inundaciones.

Algunas inundaciones ocurren simplemente por una fuerte Iluvia. Cuando la Iluvia cae incesantemente no hay tiempo para el agua se seque y el agua simplemente inunda ciudades y campos. Por ejemplo, a mediados de 1996 hubo una caída de Iluvia continua en la costa norte de Nueva Gales del Sur y la ciudad de Grafton se cubrió con un metro de agua.

<sup>&</sup>lt;sup>9</sup> Board of Studies NSW (1998). English K-9 Work Samples, p. 143.

Los huracanes y ciclones también son fuentes de inundaciones. Cuando un huracán llega trae a las nubes que prácticamente llenan de agua de lluvia a los campos. Otra forma de que el huracán pueda causar inundaciones es trayendo el agua de los lagos y ríos. La lanza al cielo y luego cae al piso con una fuerza poderosa, esta es otra causa de inundaciones.



En algunos países las inundaciones son anuales. El año se divide entre estaciones húmedas y secas; sin embargo, las inundaciones son esperadas en algunas ocasiones y son de alguna forma beneficiosas. Cuando la inundación de los ríos termina, deja un suelo fresco y fértil. Esto ayuda a los granjeros en países desiertos (por ejemplo, Egipto) a plantar maíz para alimentar al país.

#### 1.5. Una variedad de formatos de respuesta

Se puede utilizar una variedad de formatos de respuesta en las evaluaciones escritas. Por ejemplo, la comprensión de los estudiantes puede ser evaluada a través de preguntas de selección múltiple, preguntas con respuestas cortas, ensayos y mapas de conceptos, y sus actitudes y valores a través de cuestionarios del tipo Likert.

Algunos formatos de respuestas encajan mejor con la evaluación de tipos particulares de logros que otros. A continuación, se muestran los ejemplos de logros y de formatos de respuestas escritas que pueden ser utilizados en sus evaluaciones.

# Ejemplos de formatos de respuestas que pueden ser utilizados para evaluar resultados seleccionados

Resultado	Formato de respuesta
Produce textos claramente, efectivamente y de manera acertada, utilizando la estructura de la oración, características gramaticales y puntuación del tipo de texto (Inglés, Australia).	ensayo
Interpreta y compara la información mostrada (Matemáticas, Ontario)	respuesta corta
Utiliza ideas abstractas sobre el fenómeno físico en las explicaciones (Ciencia, Reino Unido).	respuesta corta
Identifica eventos que puedan afectar el balance en un ecosistema (Ciencia, Australia).	opción múltiple
Demuestra discernimiento de temas éticos y reconoce la necesidad de veracidad e integridad (Valores, Australia Occidental).	cuestionario estilo Likert

# 1.6. Calificar y registrar las respuestas de los estudiantes

Dependiendo del formato de respuesta utilizado para recolectar evidencia del logro, las respuestas a los exámenes escritos pueden ser puntuadas de manera computarizada, a mano o utilizando una guía de evaluación o escala de puntuación.

Las dos consideraciones importantes en la calificación de las respuestas del estudiante son:

- El método de calificación; y
- · La comparación de calificaciones.

## **Puntuación parcial**

La puntación parcial permite a los evaluadores reconocer y registrar la variación de niveles de logro del estudiante, su comprensión y estrategias parcialmente correctas. En este ejemplo, los evaluadores acreditan a sus estudiantes con un punto si demuestran comprensión parcial del proceso descrito. Ellos son evaluados por su comprensión de lectura y no por su habilidad al dibujar.<sup>10</sup>

Marconi obtiene la grasa de lagarto de las personas que vivieron en los arbustos. Una forma que utilizaron para extraer la grasa fue asar el lagarto sobre el fuego. Ellos colocaron una pieza de acero corrugado en una ligera pendiente junto al fuego. La grasa caliente bajaba por las canaletas en la plancha de acero y se recogió al final en latas.

Dibuje una imagen mostrando este método

**2 puntos** el lagarto cocinado sobre el fuego y la intención de mostrar la recolección de grasa de una forma plausible.



**1 punto** solo el lagarto cocinado sobre el fuego o recolección poco factible.

**0 puntos** dibujo mostrando ninguno de los anteriores; por ejemplo, solo el fuego o el lagarto

<sup>&</sup>lt;sup>10</sup> Forster, M.; Mendelovits, J. y Masters, G. (1994). Developmental Assessment Resource for Teachers DART English. Camberwell: Australian Council for Educational Research.

#### Método de calificación

Algunos formatos de repuesta de evaluaciones escritas (por ejemplo, preguntas de elección múltiple, verdadero/falso) no requieren que los evaluadores utilicen su juicio cuando evalúan las respuestas de los estudiantes. La calificación sobre la respuesta correcta se ha realizado durante la construcción de la pregunta. Otro formato de respuesta requiere de los evaluadores juzgar la calidad de la respuesta del estudiante utilizando una guía de puntación o una escala de calificaciones.

Las guías de puntuación algunas veces son utilizadas para calificar la calidad de las respuestas de los estudiantes a preguntas que requieran respuestas cortas. Por ejemplo, las guías de puntuación parcial permiten a los evaluadores reconocer y registrar las variaciones de niveles de éxito de los estudiantes, su comprensión y estrategias parcialmente correctas. En la página anterior se muestra un ejemplo de una guía de calificación parcial para una pregunta de comprensión de lectura.

Las escalas de calificación se utilizan normalmente para ayudar a los evaluadores a emitir juicios sobre respuestas amplias. Al utilizar escalas de puntaje, los evaluadores juzgan la calidad del trabajo de los estudiantes sobre criterios específicos. Los evaluadores realizan calificaciones holísticas de trabajo cuando ellos realizan una sola calificación a partir de una impresión general del trabajo, y ellos realizan una calificación analítica cuando califican diferentes aspectos del trabajo.

Una pieza de escritura, por ejemplo, puede ser calificada por la calidad de ideas tanto como por el control sobre las características de la escritura (ortografía, puntuación y gramática). Un ejemplo de una escala de calificación holística para la evaluación de temas abiertos en ciencia se muestra en la página 28. Un ejemplo de la calificación analítica de una secuela narrativa se muestra en la página 29.

## Escala de puntuación holística

Al utilizar escalas de puntuación holísticas, los evaluadores califican la calidad del trabajo del estudiante sobre criterios especificados. Los evaluadores hacen calificaciones holísticas del trabajo cuando ellos realizan una sola calificación basada en una impresión general del trabajo.

Rúbrica de calificación para la Prueba de desempeño Académico de Connecticut (CAPT, por sus siglas en inglés) para temas sin desarrollo preestablecido.<sup>11</sup>

3

La respuesta es una respuesta excelente a la pregunta. Es correcta, completa y apropiada, y contiene elaboración, y extensión y/o evidencia de un pensamiento más amplio y conocimiento previo relevante. No hay evidencia de mala interpretación de conceptos. Errores mínimos no bajarán el puntaje necesariamente.

2

La respuesta es una respuesta práctica a la pregunta. Es generalmente correcta, completa y apropiada, a pesar de inexactitudes menores que puedan aparecer. Puede haber evidencia limitada a la elaboración, extensión, pensamiento de orden más amplio y conocimiento previo relevante, o puede haber evidencia significativa de estas cualidades, pero otras fallas (ej., inexactitud, omisiones, impropiedades) pueden ser más que menores.

1

La respuesta es una respuesta marginal a la pregunta. Mientras pueda contener algunos elementos de una respuesta competente, es imprecisa, incompleta, y/o inapropiada. Hay poca, si lo hay, evidencia de elaboración, extensión, pensamiento de orden más amplio o conocimiento previo relevante. Puede haber evidencia de ideas equivocadas significativas.

0

La respuesta, a pesar del tema, es una respuesta insatisfactoria a la pregunta. Puede fallar en responder la pregunta, o puede responder la pregunta en una forma muy limitada. Puede no haber evidencia de elaboración, extensión o pensamiento de orden más amplio, o conocimiento previo relevante. Puede haber evidencia de serias ideas equivocadas.

Hibbard, M. et al. (1996). A Teacher's guide to Performance-Based Learning and Assessment, p. 24. Alexandria: Association for Supervision and Curriculum Development Alexandria Virginia.

## La necesidad de comparar calificaciones

En el contexto del salón de clases, si otro profesor realiza la misma calificación del trabajo de los estudiantes utilizando la misma guía de evaluación usualmente no es de gran preocupación. En contextos de altos rendimientos, sin embargo, la fiabilidad del interevaluador es crucial.

Usualmente, mientras mayor sea el requerimiento para comparación, más ajustadamente el criterio de evaluación será especificado. Los evaluadores son capacitados y el proceso de evaluación es cuidadosamente monitoreado para asegurar un alto nivel de acuerdo interevaluadores. En algunos escenarios de alto rendimiento, los productos son evaluados por varios evaluadores.

Escala de pun	Escala de puntuación analítica			
Al utilizar escalas de calificación, los evaluado sobre criterios específicos. Los evaluadores diferentes aspectos del trabajo. Guía de puntuación para la escritura de una s	hacen calificacione		•	
Criterio	No muestra	Bajo	Medio	Alto
<ol> <li>La historia es una continuación lógica del original.</li> </ol>				
2 El escenario se describe vívidamente.				
3 Los personajes se describen bien.				
4 Hay un diálogo a través de la historia.				
5 Hay un argumento lógico que incluye exposición, acción en crecimiento, punto culminante, acción en descenso, resolución.				
<b>6</b> La pieza es mecánicamente correcta.				
<b>7</b> La pieza es especialmente interesante, creativa y cautivadora.				

<sup>&</sup>lt;sup>12</sup> Adaptado de Hibbard, m. *et. al.* (1996: 63).

## 1.7. Estimar y reportar ubicación en el mapa de progreso

En la evaluación del desarrollo, los profesores monitorean el progreso de los estudiantes sobre un mapa preconstruido de desarrollo de conocimiento, habilidades y comprensión (ver *Evaluación del desarrollo. Recursos de Evaluación del Consejo Australiano para la Investigación Educativa 2*). Los profesores realizan observaciones relevantes a los resultados en un mapa de progreso y utilizan estas observaciones como "evidencia" para estimar los niveles de logro del estudiante.

La evaluación del desarrollo requiere de una decisión equilibrada (inferencia) sobre la ubicación del estudiante en un mapa de progreso. Las evaluaciones escritas brindan un tipo de evidencia. La forma en la que esta evidencia es utilizada para inferir un nivel de rendimiento es normalmente determinada por el propósito de la evaluación: mientras más alto sea el rendimiento, mayor el requerimiento de comparación y más precisamente el "proceso de inferencia" será especificado.

## Estimación subjetiva

Cuando la estimación del nivel de logro de un estudiante es realizada subjetivamente, puede haber una pérdida de conexión entre la evidencia disponible (calificaciones y registros de desempeño en tareas escritas) y el resultado estimado de la ubicación de los estudiantes en un mapa de progreso.

## Estimación objetiva

En situaciones de alto rendimiento, donde son deseables altos niveles de comparación interevaluadores, la forma en la que los registros y calificaciones del desempeño del estudiante en las tareas escritas son utilizados para estimar la ubicación del estudiante en un mapa de progreso puede ser prescrita precisamente para asegurar que el estimado se realiza objetivamente. Por

ejemplo, la inferencia puede ser de forma numérica basada en un patrón de calificaciones del evaluador sobre un criterio definido cuidadosamente.

#### 1.8. Múltiples fuentes de evidencia

En la evaluación del desarrollo, las evaluaciones escritas son utilizadas en conjunto con otros métodos de evaluación para brindar múltiples perspectivas acerca de la comprensión de los estudiantes. El extracto en las páginas 32-33 es tomado de un documento de apoyo del profesor. Este brinda un modelo de diseño de evaluación que dirige el enfoque curricular y los resultados en una unidad particular de trabajo. La unidad "placas tectónicas" hace explícita la relevancia curricular y el enfoque del resultado del aprendizaje. Note que la evaluación incluye una presentación grupal (evaluación del desempeño) un modelo con explicaciones escritas (evaluación del producto) y una prueba de conceptos (evaluación escrita). Al hacer una evaluación del nivel de logro del estudiante en un mapa de progreso, puede ser posible utilizar la evidencia de estas tres fuentes.

## 1.9. Siguientes artículos

Los siguientes segmentos exploran evaluaciones escritas más detalladamente.

El primero describe la variedad de evaluaciones escritas. ¿Cuáles son las diferentes formas de evaluaciones escritas y para qué propósitos se utilizan? El segundo artículo explora el diseño de evaluación escrita:

- consideraciones al planear evaluaciones escritas, incluyendo temas de validación, fiabilidad y legitimidad;
- las fortalezas y debilidades de diferentes formatos escritos;
- cómo escribir una buena evaluación escrita;
- cómo calificar y registrar las respuestas de los estudiantes;

- cómo estimar los niveles de rendimientos del estudiante; y
- cómo reportar el logro.

El tercer artículo explora las consideraciones comunes en el desarrollo y uso de pruebas escritas profesionalmente diseñadas incluyendo las siguientes:

- problemas en la elaboración de preguntas justas;
- procedimiento para condiciones de prueba estandarizadas;
- procedimiento para puntuación estandarizada; y
- formas de interpretar los resultados de las pruebas.

El artículo final brinda una revisión completa del proceso de diseño de las pruebas escritas. Una lista de verificación resume con detalle los pasos en el proceso.

# Utilizar evaluaciones escritas en conjunto con otros métodos de evaluación

Este extracto proviene de un documento de respaldo de un profesor. Este provee un modelo para diseñar tareas a ser evaluadas que dirige el enfoque curricular y los resultados en una unidad particular de trabajo.<sup>13</sup> La evaluación escrita (una prueba de conceptos) es un método de evaluación sugerido.

Muestra de una tarea de diseño a ser evaluada que dirige el enfoque curricular y el resultado en una unidad particular de trabajo.

<sup>&</sup>lt;sup>13</sup> Board of Studies Victoria (1996). *Using the CSF Assessment and Reporting*, p. 33. Carlton: Board of Studies.

Resumen de enfoque curricular relevante	Resultados de aprendizaje relevantes	Tareas a ser evaluadas		
En este nivel, el estudiante tiene conocimiento de: Enfoque en las causas de pliegues y fallas en relación a las placas y a la teoría tectónica; los tipos de investigación de pliegues y fallas en las rocas y las causas de temblores y volcanes; evidencia de teoría de placas tectónicas.	<ul> <li>La Tierra cambiante</li> <li>Argumenta sobre pliegues y fallas y plataformas asociadas.</li> <li>Argumenta sobre la evidencia que respalda la teoría de las placas tectónicas.</li> </ul>	1 Los estudiantes desarrollan un mapa rompecabezas de Gondwana, 120 millones de años atrás, que se separa en la actual masa terrenal sureña; representando y explicando las asociaciones entre el giro continental, tipos de roca en conjunto y distribución de especies.  2 Los estudiantes construyen un modelo de trabajo de un pliegue o falla, representando y explicando cómo ocurre el fenómeno.		
¿Qué es lo que hago?	¿Qué evalúo? ¿En qué facetas?	¿Cómo evaluaré?		
Procedimiento  1 ¿Dónde se encuentran? Volcanes, terremotos, crestas y zanjas.  2 Ejercicio de expansión de la cresta marítima del Atlántico medio.  3 Evidencia fósil para placas tectónicas.  Fuentes brindadas  1 Mapa mundial  2 Papel, tijeras  3 Escala de tiempo geológico y ejemplos de tipos de fósiles (actual o foto), trilobites,	<ul> <li>Tierra y más allá</li> <li>Reconocimiento de tipos de pliegues y desperfectos.</li> <li>Comprensión de causas de pliegues y fallas, actividad volcánica y tectónica con relación a la teoría de placas tectónicas.</li> <li>Comprensión del uso de evidencia fósil para respaldar la teoría de placas tectónicas.</li> </ul>	<ul> <li>Presentación de grupo del modelo y explicación de pliegues y fallas.</li> <li>Modelo y explicación escrita de Gondwana.</li> <li>Prueba de conceptos de deposición fósil, pliegues, fallas y placas tectónicas.</li> </ul>		

#### 2. UNA VARIEDAD DE FORMATOS DE RESPUESTA

La característica que define la evaluación escrita es que los estudiantes responden a una serie de preguntas o apuntes. Las respuestas a preguntas o apuntes brindan evidencia del conocimiento, habilidad, comprensión o actitudes del estudiante.

#### Ejemplo de formato de respuestas escritas

- Rellenar espacios en blanco
- Mapas de concepto
- Ensayos
- Investigación
- Unir términos
- Opción múltiple
- Cuestionarios estilo Likert
- Autorreflexiones
- Respuestas cortas
- Parafraseo escrito

Una variedad de formatos de respuesta puede ser utilizada con preguntas escritas. Por ejemplo, se le puede pedir a los estudiantes que respondan preguntas que miden conocimiento, habilidad y comprensión utilizando formatos de opción múltiple, de respuestas cortas o largas. Se les puede pedir que respondan a cuestionarios que midan actitudes y valores utilizando términos no concluyentes, escala Likert y diferenciales semánticas.

Este artículo describe un rango de formatos de respuestas comunes de papel y lápiz y discute el tipo de información que están diseñados para recolectar.

#### Ejemplo: Evaluar la comprensión de texto de los estudiantes

#### Extracto de texto<sup>1</sup>

"¡Culebra!" Exclamó la madre de Tony.

Era el primer día de las vacaciones de campamento de la familia.

Con una mano, su madre levantó al bebé del pasto cerca a la carpa. Con la otra alargó el palo de la sombrilla de playa. "¡Aléjate!" gritó mientras Tony se acercaba a la carpa.

Tony rio y se acercó a la carpa. "Es solo una lagartija, mamá." La vi caminando por el pasto.

Era el primer día de las vacaciones de campamento de la familia. Algo se arrastró a la carpa y \_\_\_\_\_\_ pensó que era una culebra. La madre levantó al bebé porque ella pensó que la culebra podría \_\_\_\_\_. Tony rio porque él sabía que el animal era una \_\_\_\_\_.

# 2.1. Procedimiento para rellenar espacios en blanco

El procedimiento para rellenar espacios en blanco requiere que los estudiantes completen un texto cuyas palabras han sido borradas.

En el formato tradicional, se coloca una sección introductoria del texto sin modificarse a fin de fijar un contexto para el lector. Luego, cada cinco o diez palabras se elimina una palabra. En el formato modificado, se retira una palabra o frase para evaluar, por ejemplo, la comprensión o conocimiento de las estructuras sintácticas del lenguaje.

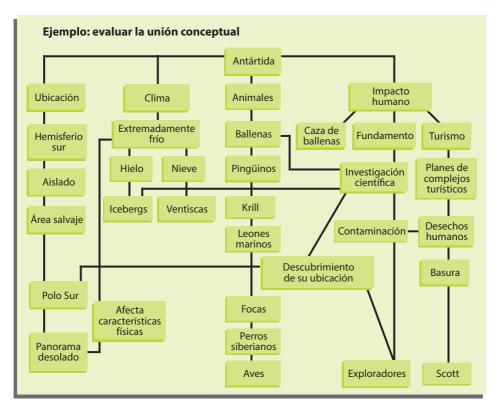
En el ejemplo anterior de un formato modificado, se ha evaluado la comprensión de lectura.

Mossenson, L.; Hill, P. y Masters, G. (1987). Tests of Reading Comprehension (TORCH). Camberwell: Australian Council for Educational Research.

#### 2.2. Mapas de conceptos

Los mapas de conceptos son representaciones diagramáticas de conceptos clave que están relacionados a un área de investigación. Las relaciones entre los conceptos se muestran con líneas o flechas. Otros nombres para los mapas de conceptos pueden ser mapas conceptuales, red semántica, y organizadores gráficos.

Los mapas de conceptos se utilizan para evaluar el nivel y complejidad de la comprensión del estudiante. El criterio de puntaje, normalmente, incluye el número de ideas generadas, la clasificación de información y la calidad de unión de conceptos. A continuación, se muestra un ejemplo de un mapa de conceptos sobre la Antártida.<sup>2</sup>



<sup>&</sup>lt;sup>2</sup> Clarke, J. Monitoring Learning in English, p. 83. Brisbane: Mount Coot-tha School Support Centre.

#### 2.3. Respuestas construidas

Ver los formatos de respuestas cortas (página 44).

#### 2.4. Formato de ensayo

Los ensayos son respuestas largas que los estudiantes brindan a un tema o a un "punto de vista".

El formato de ensayo es utilizado para evaluar la habilidad de los estudiantes de construir una respuesta considerada, por ejemplo, para desarrollar un argumento construido lógicamente o a una narrativa estructurada con contexto, personaje, problema y solución.

Los ensayos pueden ser estructurados (por ejemplo, "explicar brevemente", "describir tres métodos para...") o abiertos en su formulación. A continuación, se muestran ejemplos de ensayos estructurados y abiertos.<sup>3</sup>

#### Ejemplo: Evaluar un argumento construido lógicamente

#### Tema de ensayo estructurado:

Seleccione uno de los siguientes y argumente sobre los problemas medioambientales creados:

- a. Drenar aceite usado en la tierra
- b. Chorrear aceite usado en el drenaje

Mantenga su respuesta a menos de media página.

#### Tema de ensayo abierto:

Evalúe críticamente el impacto de la legislación antidiscriminación en los puestos de trabajo de Australia.

<sup>&</sup>lt;sup>3</sup> Clayton, B. (1995). Focussing on Assessment, p. 1139. Adelaide: National Centre for Vocational Education Research.

# 2.5. Investigaciones

Las investigaciones son tareas abiertas que requieren que los estudiantes apliquen conceptos familiares y habilidades en situaciones no familiares. Los estudiantes que completan investigaciones escritas brindan reportes escritos de sus hallazgos.

Las investigaciones son normalmente utilizadas para evaluar la comprensión matemática de un estudiante, su habilidad para resolver problemas y el uso del idioma matemático. A continuación, se muestra un ejemplo de investigación matemática.<sup>4</sup>

familiares y habilidades en sit	as abiertas que requieren que tuaciones no familiares. Este e	e los estudiantes apliquen conceptos ejemplo puede utilizarse para evaluar ara la solución de problemas y el uso
Utilice los patrones de bloq	jues para investigar	
Utilice los patrones de bloque más grande.	es para investigar el cuadro, e	el cuadro más grande, y el que es aun
Cuadro ¿Qué patrones se pueden en ¿Qué sucede con las otras for		Cuadro más grande

<sup>&</sup>lt;sup>4</sup> Holcroft, L. y Coates, R. Monitoring Learning in Mathematics, p. 15. Queensland: Mount Coot-tha School Support Centre.

#### 2.6. Escalas Likert

Las escalas Likert son utilizadas mayormente como formatos de respuestas en cuestionarios diseñados para evaluar opiniones, actitudes, valores e intereses. Una escala Likert consiste en un conjunto de categorías ordenadas: Fuertemente en desacuerdo, En desacuerdo, De acuerdo y Fuertemente de Acuerdo. Los evaluados registran las respuestas a cada pregunta utilizando las alternativas brindadas.

Los estudiantes que responden de manera anónima al cuestionario diseñado por la Escuela John XXIII en Perth, Australia Occidental, para evaluar los niveles de "conciencia" del estudiante registran sus respuestas para cada afirmación utilizando una escala de puntaje de 4 puntos: "Fuertemente en desacuerdo", "En desacuerdo", "De acuerdo", "Fuertemente de acuerdo". El extracto que mostramos a continuación muestra algunas de las afirmaciones de la escala de "Conciencia".				
	FED	ED	DA	FDA
Me sentiría mal si robara algo.				
Si encontrara una billetera con identificación, trataría de encontrar al propietario.				
Si mis amigos planean robar, trataría de convencerlos de no hacerlo.				
Preferiría hacer mi propio trabajo, aunque no sea tan bueno, a copiar y hacerlo bien.				

# 2.7. Ítems correspondientes

Los ítems correspondientes consisten en dos listas: una lista de premisas y una lista de respuestas. Las indicaciones explican las bases en las que se realiza la correspondencia entre ambas.

Cuestionario desarrollado por el Consejo Australiano para la Investigación Educativa Camberwell Australia, para el John XXIII College, Western Australia.

Los ítems correspondientes son normalmente utilizados para evaluar la habilidad de recordar el conocimiento y comprender las relaciones, por ejemplo, para recordar términos y sus definiciones, o para comprender principios e ilustraciones.

Instrucciones: La columna A es una lista de características, mientras que la columna B es una lista de planetas. Relacione las características con los planetas ubicando la letra correcta en el espacio a la izquierda de donde se encuentran los números de la columna A.<sup>6</sup>

C	olumna A	Co	lumna B
1	Más cercano al Sol	a	Tierra
2	Más lejano del Sol	b	Júpiter
3	Planeta más grande	с	Marte
4	Muchas lunas	d	Mercurio
5	Planeta más pequeño	е	Neptuno
		f	Plutón
		g	Saturno
		h	Urano
		i	Venus

# 2.8. Formato de opciones múltiples

Los formatos de opciones múltiples contienen una pregunta o una afirmación incompleta seguida por una serie de alternativas entre las cuales los estudiantes seleccionan la respuesta.

Algunas veces, las alternativas brindan al estudiante varias respuestas posibles y la respuesta correcta ("clave") es la que más se acerca de estas. Algunas veces, las alternativas a las claves son equivocadas; a pesar de ello, algunos estudiantes pensarán que son las correctas.

<sup>&</sup>lt;sup>6</sup> Clayton, B. (1995: 124).

# 

# Ejemplo 2: evaluar la comprensión literal La respuesta correcta (clave) se encuentra marcada con un asterisco. Mi nombre es Anita. Tengo ocho años. Nací en Canadá. Vine a Nueva Zelanda cuando tenía 6 años. Tengo un primo que vive en Wellington y otro que vive en Auckland. Me gustan las vacaciones y la playa. No me gusta ir a la escuela. ¿Dónde nació Anita? Auckland Nueva Zelanda Welllington Canadá\*

Las preguntas de opciones múltiples son utilizadas mayormente para evaluar el conocimiento de información fáctica o comprensión literal. Los ejemplos 1 y 2 de esta página son muestras de preguntas con opciones múltiples de este tipo. Sin embargo, las preguntas de opciones múltiples también pueden ser construidas para evaluar las habilidades de razonamiento complejo y comprensión sofisticada. En el ejemplo 3 de la siguiente página, los estudiantes necesitan razonar lógicamente para responder a las preguntas. En el ejemplo 4, en la página 43, los estudiantes necesitan hacer una inferencia del texto.

Una crítica realizada a las preguntas de opción múltiple es que no requieren que los estudiantes generen una respuesta y, por lo tanto, serían "inauténticas" o inválidas. Esta característica puede, sin embargo, permitir la evaluación de un sutil entendimiento antes de que los estudiantes puedan expresar su comprensión en la escritura.

Otra crítica realizada a este tipo de preguntas de opción múltiple es que no requieren que los evaluadores realicen un dictamen cuando se evalúan las respuestas del estudiante. Esto es cierto. Se realiza un dictamen sobre la respuesta correcta durante la construcción de la pregunta.

Ejemplo 3: evaluar el razonamiento lógico	
La respuesta correcta (clave) se encuentra marcada con un asterisco.	
Pregunta 1	
Juan quiere hacer un pastel y trata de recordar la receta. Sin embargo, su primer intento hacer el pastel no tuvo éxito.	de
Él trata nuevamente, esta vez agrega dos nuevos ingredientes: una cucharadita de bicarbor de sodio y un huevo. No hay otros cambios en la receta o en el método de hacer el pastel utilizó en el primer intento. Esta vez el pastel es un éxito.	
¿Cuál de las siguientes conclusiones es la mejor sobre el éxito del segundo intento?7	
El éxito se debe a que agregó:	
Ya sea el bicarbonato de sodio o el huevo.	
Ambos, el bicarbonato de sodio y el huevo juntos.	
Ya sea el bicarbonato de sodio o el huevo, o ambos juntos.*	
Ni el bicarbonato de sodio ni el huevo.	
Pregunta 2  Los siguientes tres guantes mostrados se encuentran en la mesa. Cada uno de ellos mue	stra
un patrón solo en la superficie de su parte trasera (encima). Solo se puede ver la parte tras del guante.	
¿Cuál de los siguientes pares de guantes serían consistentes con la mano derecha e izquie de los guantes? <sup>8</sup>	erda
Solo I y II*	
Solo I y III	

<sup>&</sup>lt;sup>7</sup> Victorian Board of Studies (1998). *Victorian Certificate of Education General Achievement Test*.

<sup>&</sup>lt;sup>8</sup> Victorian Board of Studies (1997). Victorian Certificate of Education General Achievement Test.

Ejemplo 4: Evaluación de la comprensión inferencial
Las respuestas correctas (clave) se encuentran marcadas con un asterisco.
Pregunta 1
Las líneas 9 y 10 se refieren al desempeño de una mujer frente a Rubinstein:9
"Una vez ella tocó frente a Rubinstein, quien bostezó."
Esta referencia sugiere que:
El talento musical de la mujer fue limitado. *
La mujer pudo haber sido un gran músico.
Rubinstein subestimó el talento de la mujer.
La mujer conserva ese momento como el más enriquecedor de su vida.
Pregunta 2
"Cuando escuches ruido de cascos, piensa en caballos antes que en cebras". 10
Dicho médico anónimo
Este dicho alarma a los doctores sobre:
La ocurrencia más común es la causa más probable. *
Los síntomas de las enfermedades no son las que parecen.
El mejor diagnóstico requiere amplitud y profundidad de pensamiento.
Ellos deben conocer todos los síntomas antes de hacer un diagnóstico.

#### 2.9. Parafraseo

El parafraseo requiere que los estudiantes relaten, en sus propias palabras, algo que han visto, leído o escuchado. Los estudiantes que completan el parafraseo escrito brindan sus respuestas por escrito.

El parafraseo es utilizado inicialmente para evaluar la comprensión de los estudiantes, la calidad y la organización de sus ideas. El parafraseo escrito

<sup>&</sup>lt;sup>9</sup> Queensland Core Skills Test (1997). Core Skills Test Multiple Choice Questions II. Queensland: Board of Senior Secondary Studies.

<sup>&</sup>lt;sup>10</sup> Victorian Board of Studies (1996). Victorian Certificate of Education General Achievement Test.

también brinda una oportunidad para evaluar el control de los estudiantes sobre las características del lenguaje (gramática, ortografía, y puntuación).

# 2.10. Formatos de respuestas cortas

Las evaluaciones de respuestas cortas, o preguntas de "respuestas construidas", requieren que los estudiantes generen una respuesta breve. La respuesta puede ser a una pregunta o puede requerir el completar una oración.

Las preguntas de respuestas cortas pueden ser "cerradas" o "abiertas". Las preguntas cerradas solo tienen una respuesta aceptable. Las preguntas abiertas pueden tener más de una respuesta correcta.

Las preguntas de respuestas cortas son utilizadas para evaluar la comprensión conceptual de los estudiantes al igual que el conocimiento fáctico. El ejemplo 1 evalúa la comprensión conceptual del estudiante en ciencia. Las respuestas de los estudiantes demuestran los diferentes niveles de comprensión de las reacciones químicas y el conocimiento sobre la conservación de la materia. El ejemplo 2 es una pregunta abierta que permite varias alternativas según el texto.



# Ejemplo 1: evaluar los niveles de comprensión conceptual<sup>11</sup>

Las respuestas de los estudiantes demuestran los diferentes niveles de comprensión sobre las reacciones químicas y el conocimiento acerca de la conservación de la materia.



¿Adónde va la madera cuando se quema el fósforo?

¿El fósforo tiene el mismo peso antes y después de quemarse? Explique su respuesta.

<sup>&</sup>lt;sup>11</sup> Doig, B. y Adams, R. (1993). *Tapping Students' Science Beliefs*. Camberwell: The Australian Council for Educational Research.

Ejemplo 2: evaluar múltiples lecturas del texto<sup>12</sup>



En aquel tiempo (Olimpiadas de Moscú, 1980), este dibujo apareció en el periódico *New York Daily News*. ¿Qué significan las diferentes características del dibujo sobre la mezcla entre el deporte y la política?

Asegúrese de describir el punto central al que hace referencia el dibujante.

#### 2.11. Autorreflexiones

Las autorreflexiones son las reflexiones que hacen los estudiantes sobre su propio trabajo y progreso.

Las reflexiones son normalmente utilizadas como un dispositivo curricular para ayudar a los estudiantes a negociar metas específicas y hacerlos conscientes de su propio conocimiento, habilidad y proceso de pensamiento. También

<sup>12</sup> Queensland Core Skills Test (1997). Core Skills Test Paper 3 Short Response Items. Queensland: Board of Senior Secondary Studies.

son útiles para la propia evaluación. Los profesores, algunas veces, evalúan las autorreflexiones de los alumnos (ver página 82 para ver el ejemplo de una guía de evaluación que brinda un puntaje a la autoevaluación del alumno). Un ejemplo de autorreflexión se muestra a continuación.<sup>13</sup>

Nombre:
Fecha:/
Descripción o título del trabajo:
¿Cuánto tiempo le tomó realizar este trabajo?
¿Dónde lo realizó?
Mayormente en la escuela  Mayormente fuera de la escuela  Mitad y mitad
¿Alguien lo ayudó con este trabajo?
¿Qué es lo correcto que hizo en este trabajo?
¿Qué es lo que hace buena a esa parte del trabajo?
Si pudiera cambiar algo sobre este trabajo, ¿qué cambiaría?
¿Por qué lo cambiarías?
¿Te gustó hacer este trabajo?
Sí No
Un poco

<sup>&</sup>lt;sup>13</sup> Adaptado de Hibbard, M. et al. (1996). Performance-based Learning and Assessment, p. 204. Virginia: Association for Supervision and Curriculum Development.

# 3. EVALUACIÓN ESCRITA DISEÑADA POR EL PROFESOR

Los profesores algunas veces piden a los estudiantes completar evaluaciones escritas con la intención de utilizar sus respuestas para evaluar niveles individuales de comprensión, habilidad, conocimiento o actitudes. Los ejemplos de tareas escritas diseñadas por los profesores incluyen evaluaciones de opción múltiple, respuestas cortas y preguntas ensayo, y cuestionarios diseñados para evaluar valores y actitudes.

Esta sección observa las evaluaciones diseñadas por los profesores. Las evaluaciones de este tipo son diferentes a las observaciones informales que realizan los profesores sobre los estudiantes involucrados en actividades de escritura en el transcurso de la enseñanza y el aprendizaje diario. También son diferentes a las evaluaciones escritas desarrolladas por desarrolladores profesionales de evaluaciones para su uso en contextos de alto rendimiento, en los que se requieren esfuerzos especiales para asegurar la comparación de evaluación de estudiante a estudiante, evaluador a evaluador y escuela a escuela (ver páginas 102-105).

Las evaluaciones escritas diseñadas por el profesor son utilizadas para recolectar información que los profesores no puedan ver diariamente y para hacerlo eficientemente. Las evaluaciones se planean con anticipación para brindar evidencia de resultados particulares en un área de aprendizaje, y los profesores deciden anticipadamente el formato de respuesta y el método para evaluar el trabajo del estudiante.

Los profesores utilizan las evaluaciones escritas para recolectar información con diferentes propósitos, incluyendo el diagnosticar fortalezas y debilidades de los estudiantes, para informar el nivel de enseñanza en un nivel global y para inferir los niveles de logro de los estudiantes para reportarlos a los padres.

Cuando se evalúa para un diagnóstico o con propósitos formativos, los profesores algunas veces enfocan las evaluaciones en un aspecto particular de

conocimiento o habilidad. Por ejemplo: "¿El estudiante entiende sustracciones de dos dígitos cuando se requiere reagrupación?". El profesor necesita saber cuándo el concepto particular ha sido entendido y cuándo es necesaria una enseñanza adicional. El propósito es tener una conclusión sobre la comprensión del estudiante en un aspecto particular de sustracción para informar el proceso de enseñanza.

Por el contrario, cuando el propósito de evaluación es reportar el nivel de logro en un área de aprendizaje, los profesores utilizan registros del desempeño del estudiante en tareas dirigidas a un rango de habilidades relacionadas de forma parecida, para poder inferir un nivel de logro. Por ejemplo, una conclusión sobre el nivel de logro del estudiante en "Números" podría requerir la evaluación de un rango mayor de comprensión que la sustracción de números de dos dígitos.

#### 3.1. Planeamiento de la evaluación escrita

La utilidad de tareas escritas planeadas como fuente de evidencia sobre el logro o actitud de un estudiante depende de:

- qué tan bien dirigidas están las tareas a metas instruccionales, o los resultados de un área de aprendizaje (relevancia, un tema validado);
- la cantidad de evidencia brindada (un tema confiable);
- qué tan bien las tareas permiten a los estudiantes mostrar sus fortalezas y debilidades; y
- qué tan fiables son las tareas para los estudiantes con diferentes antecedentes de lenguaje, culturales y socioeconómicos.

#### **Orientar resultados**

En las evaluaciones de desarrollo, los profesores recolectan evidencia del logro de los estudiantes usando evaluaciones que permiten ubicar resultados en un mapa de progreso. Las evaluaciones escritas son útiles para recolectar evidencia sobre un amplio rango de resultados de aprendizajes en muchas áreas curriculares. En la página 18 se muestran algunos ejemplos de resultados

El propósito de una evaluación escrita determina el rango de resultados relacionados. Las evaluaciones de diagnóstico pueden estar más enfocadas en un rango delimitado de conocimiento y comprensión. Las evaluaciones para reportar el nivel de logro en un dominio más amplio dirigirán un rango más amplio de resultados de aprendizaje.

logrados que pueden obtenerse utilizando evaluaciones escritas.

Cuando se planean pruebas escritas, es importante tener en claro el propósito de la evaluación. Cuando la evaluación es para un propósito de diagnóstico o formativo, se puede orientar un rango limitado de resultados y comprensión.

Cuando la evaluación es para propósitos sumativos, para inferir el nivel de logro de los estudiantes en un área de aprendizaje, es importante que las actividades permitan un amplio rango de resultados valiosos. Esto significa que las tareas necesitan brindar una cobertura curricular. Si las actividades brindan una cobertura limitada de resultados, luego las inferencias sobre el logro del estudiante serán limitadas. Evaluar un rango limitado de resultados también puede enviar un mensaje no intencional a los estudiantes sobre lo que es valioso.

Cualquiera que sea el propósito de la evaluación, es esencial que exista un número suficiente de tareas desde las cuales se pueda realizar una inferencia sobre el conocimiento y comprensión de los estudiantes. Mientras más oportunidad tenga el profesor de observar lo que el estudiante sabe y puede hacer, mayor será la fiabilidad en su calificación sobre las habilidades subyacentes del estudiante.

Este principio aplica ya sea si la habilidad subyacente es un dominio definido estrictamente, como en la "sustracción", o un dominio definido de manera más amplia, como la habilidad en "numeración". Las observaciones del logro del estudiante necesitan ocurrir un gran número de veces.

Cuando se planea una tarea escrita es importante tener claros los resultados que se quieren obtener. Las tareas pueden enfocarse para brindar información y se puede seleccionar apropiados formatos de repuestas. A continuación, mostramos algunos ejemplos de resultados y ejemplos de formato de respuestas que pueden ser utilizados en su evaluación.

Resultados	Formato de respuesta
Escribe varios párrafos de información desarrollando un problema y una solución cuando es apropiado para el tema.¹ (Pensilvania, Lectura, 5.to Grado)	ensayo
Refleja de forma crítica en significados y valores asociados con un trabajo de arte visual particular.² (Australia, Artes Visuales, Nivel 8)	respuesta larga, ensayo
Identifica partes decimales de un total. <sup>3</sup> (Ontario, Matemáticas, 3. <sup>er</sup> Grado)	opciones múltiples
Enumera la forma en la que los materiales son utilizados para diferentes propósitos. <sup>4</sup> (Australia, Ciencia, Nivel 2)	respuestas cortas
Participa en el proceso político y contribuye al servicio comunitario de manera consistente con la de un buen ciudadano. <sup>5</sup> (Australia Occidental, Valores)	cuestionario

Pennsylvania Department of Education. Academic Standards for Reading, Writing, Speaking and Listening, p. 9.

<sup>&</sup>lt;sup>2</sup> Curriculum Corporation (1994). The Arts – A Curriculum Profile for Australian Schools, p. 155. Carlton: Curriculum Corporation.

<sup>&</sup>lt;sup>3</sup> Ministry of Education and Training. *Ontario Principal Standards: Mathematics*.

<sup>&</sup>lt;sup>4</sup> Curriculum Corporation (1994). Science – A Curriculum Profile for Australian Schools, p. 35. Carlton: Curriculum Corporation.

<sup>&</sup>lt;sup>5</sup> Curriculum Council Western Australia (1998). Curriculum Framework, p. 324.

También es esencial considerar la validación de la inferencia. Los profesores realizan una inferencia sobre habilidades más generales de comprensión desde las respuestas de los estudiantes hacia tareas específicas. Cuando se realiza una inferencia sobre el nivel de logro sobre un conjunto de resultados en un campo, los profesores necesitan considerar la validación de este conjunto. Por ejemplo, puede ser válido realizar una inferencia sobre el nivel de logro del estudiante en escritura a partir de una variedad de tareas orientadas a un rango de resultados de escritura. Sin embargo, puede que no sea válido realizar una inferencia sobre el nivel de logro de un estudiante en Lenguaje haciendo un conglomerado sobre resultados en lectura, visualización, expresión oral, comprensión auditiva y escritura. Los logros en estas áreas de aprendizaje pueden necesitar ser reportados por separado.

# Seleccionar el formato de respuesta

Cuando se elige de una cantidad de formatos de respuesta posibles, algunas veces ayuda enfocarse en la parte principal del resultado que se quiere obtener. A continuación, mostramos algunos ejemplos de verbos comúnmente utilizados en los resultados y ejemplos de formatos de respuesta que pueden ser utilizados en su evaluación.

Resultados que requieren de los estudiantes:		Posibles formatos de respuesta	
calcular distinguir identificar determinar	recordar seleccionar definir	opciones múltiples	
enunciar definir determinar clasificar	identificar calcular describir nombrar	respuesta corta rellenar espacios en blanco	
comparar evaluar contrastar traducir desarrollar	analizar interpretar argumentar planear	ensayo investigación	

Una vez que se aclara el propósito de la evaluación y se identifica el resultado al que se quiere orientar, el siguiente paso es considerar el formato de respuesta. Algunos formatos de respuesta se adaptan mejor a la evaluación de tipos particulares de resultados que otros. Por ejemplo, si un resultado de aprendizaje requiere que el estudiante demuestre su comprensión sobre la relación entre eventos históricos, o que estructure un argumento lógico, entonces un formato de ensayo puede ser el más apropiado. Si el resultado requiere que los estudiantes demuestren su conocimiento en fechas, entonces se puede utilizar un formato de opciones múltiples.

Cuando se seleccionan posibles formatos de respuesta, algunas veces es de ayuda enfocarse en el verbo de la oración del resultado que se quiere conseguir. Por ejemplo, un resultado que requiere que los estudiantes "identifiquen" o "recuerden" puede conseguirse utilizando un formato de respuestas de opciones múltiples. Un resultado que requiere que los estudiantes "comparen" o "evalúen" puede conseguirse de una mejor forma utilizando un formato de respuesta que requiere a los estudiantes generar un formato de respuesta-ensayo o formato de respuesta corta.

# Planeamiento de tareas imparciales

Es importante que las tareas escritas sean diseñadas de manera imparcial. Las tareas o preguntas individuales pueden permitir a los estudiantes demostrar

Consideraciones a tomar al momento de planear preguntas imparciales:

- · Enfoque claro
- Accesibilidad del lenguaje
- Tareas inclusivas
- Sin trucos
- Revisar las tareas

lo que saben y pueden hacer. En el diseño de tareas escritas, son importantes diferentes consideraciones:

# Enfoque claro

La evaluación de conocimiento y habilidades en un área no debería ser excesivamente influido por el conocimiento y habilidad del estudiante en otras áreas. Por ejemplo, los problemas matemáticos que requieren una gran cantidad de lectura pueden dificultar a los estudiantes con bajas habilidades de lectura el demostrar su comprensión matemática. Si las pruebas son administradas en computadora, los estudiantes con bajas habilidades al utilizar el teclado pueden encontrarse en desventaja.

# Accesibilidad del lenguaje

El lenguaje utilizado en las instrucciones y la presentación de la tarea debe ser lo más simple y claro como sea posible. El lenguaje utilizado en las preguntas no debe exceder el lenguaje requerido por la tarea. Por ejemplo, la comprensión de una pregunta de opción múltiple no debe ser más difícil de leer que un pasaje de lectura.

A menos que el conocimiento de un término particular se encuentre en evaluación, las preguntas deben ser escritas en un lenguaje simple. Los estudiantes pueden fallar en contestar una pregunta correctamente porque ellos no llegan a comprender el lenguaje utilizado en la pregunta. Cambiar una sola palabra en una pregunta de opción múltiple (por ejemplo, "solvente no acuoso" a "líquido diferente al agua") puede reducir la dificultad de la pregunta.

#### Tareas inclusivas

Las tareas necesitan fijarse en contextos accesibles a los estudiantes con diferente historial de género, cultura y lenguaje. Por ejemplo, una serie de problemas matemáticos fijados en el contexto de competencia de motos y carros puede no incluir de manera conjunta a las niñas para que puedan demostrar su comprensión matemática.

#### Sin trucos

Es importante que los estudiantes no sean engañados injustamente por las preguntas. Por ejemplo, deben evitarse las palabras ambiguas a menos que la ambigüedad sea intencional. Un estudiante no debe ser penalizado por una lectura divergente de una palabra ambigua. Se deben evitar las preguntas de opción múltiple escritas de forma negativa: "¿Cuál de las siguientes no es una representación acertada del problema?". Los estudiantes normalmente no toman en cuenta el "no" a menos que esté subrayado o resaltado.

#### Revisión

El desarrollo de las preguntas normalmente toma tiempo, y requiere retroalimentación y refinamiento. Es ideal que las preguntas sean desarrolladas de forma adecuada antes del momento de ser utilizadas. Esto permite tener tiempo para revisarlas antes de administrarlas a los estudiantes. Incluso los desarrolladores de pruebas más experimentados se benefician al hacer que sus preguntas sean revisadas por sus colegas. La tarea del panel es revisar los objetivos de las preguntas que

Consideraciones a tomar al momento de planear evaluaciones válidas, fiables e imparciales:

- Seleccionar un rango de preguntas
- Seleccionar tareas de diferentes niveles de dificultad
- Ordenar las preguntas cuidadosamente
- Asegurar independencia entre preguntas
- Mensaje claro
- Tiempo adecuado para finalizar

se van a realizar para revisar los errores o ambigüedades en las tareas y para sugerir mejoras. A continuación, mostramos un ejemplo de una puntuación para una lista de verificación de diseño de evaluación.

Cuando se utilizan las pruebas sobre un período de tiempo con diferentes grupos de estudiantes, ellas también pueden ser mejoradas con base en las respuestas de los estudiantes.

Diseñar listas de verificación para la revisión de un colega			
Este extracto viene de una lista de verificación desarrollada para una revisión de un colega sobre el diseño de una evaluación, por el Centro de Aprendizaje, Evaluación y Estructura Escolar. Los colegas evalúan el diseño sobre tres criterios: credibilidad del diseño, valor instruccional del diseño, y uso amigable del diseño. <sup>6</sup> Aquí se muestran los indicadores para el criterio de "valor instruccional". Primero se realiza una pregunta Sí/No para cada indicador. En la base de estas calificaciones se toma una decisión final sobre el logro que realiza este criterio en una escala de 1 al 4, donde 4 es el nivel más alto de logro.			
Sí No	¿La pregunta requiere aprendizajes del núcleo curricular?		
Sí No	¿La pregunta vale el tiempo requerido para completarla de manera correcta?		
Sí No	¿La pregunta es exigente para los estudiantes?		
Sí No	¿La retroalimentación a los estudiantes les permitirá autoevaluarse y autorregularse?		
Sí No	¿Los estudiantes podrán verificar los resultados y utilizar la retroalimentación para mejorar su posterior desempeño?		
PUNTAJE GENERAL DEL	VALOR INSTRUCCIONAL 1 2 3 4		

<sup>&</sup>lt;sup>6</sup> Wiggins, G. (1998). *Educative Assessment*, p. 151. San Francisco: Jossey-Bass Publishers.

# Pruebas válidas, fiables e imparciales

En el planeamiento de cómo agrupar una serie de preguntas para hacer una prueba válida, fiable e imparcial, se debe tomar en cuenta un número de consideraciones importantes.

# Seleccionar el rango de preguntas

Para ser válida, una prueba debe incluir un rango de preguntas orientadas a un rango relevante de resultados. Una forma de revisar la cobertura de una prueba es realizar un cuadro del tipo que mostramos en la siguiente página.

Este cuadro muestra las preguntas desarrolladas para una prueba de estadística. Para revisar la cobertura del resultado, las preguntas son categorizadas de acuerdo al tipo de resultado que se desea: "recordar los hechos", "habilidades de cálculo" y "comprensión".

#### Incluir preguntas de diferentes niveles de dificultad

La mayoría de pruebas necesitan incluir preguntas que sean apropiadas para los estudiantes con diferentes niveles de habilidades. Esto significa que se necesita incluir preguntas de diferentes niveles de dificultad. Donde se incluyen preguntas más largas, las pruebas necesitan ser lo suficientemente abiertas para ser accesibles a los estudiantes de diferentes habilidades, a fin de permitirles demostrar sus conocimientos, habilidades y comprensión.

# Ordenar las preguntas

Mientras sea posible, la pregunta más fácil debe aparecer al inicio de la prueba para que el estudiante más débil no se desanime por la dificultad de las preguntas desde el inicio y tenga la oportunidad de demostrar lo que sabe y puede hacer.

#### Asegurar la independencia de las preguntas

Todas las preguntas en la prueba necesitan ser independientes. La posibilidad de responder una pregunta correctamente no debería depender de la comprensión que se haya formado en una pregunta anterior. El fallar en una pregunta no debe influir en el éxito o fracaso en otra.

Rango e imparcialidad de las preguntas				
Una forma de revisar la cobertura del resultado de una prueba es realizar un cuadro. Este cuadro muestra las preguntas desarrolladas para una pregunta de estadística. Las preguntas son categorizadas de acuerdo al tipo de resultado dirigido: "recordar los hechos", "habilidades de cálculo" y "comprensión". La mayoría de las preguntas en esta prueba se orientan a la "comprensión".				
Contenido	Recordar los hechos	bjetivos Habilidades de cálculo	Comprensión	Totales
Distribución de frecuencia	1 pregunta	za.curo	2 preguntas	3
Significados	1 pregunta	2 preguntas	1 pregunta	4
Variaciones	1 pregunta	2 preguntas	1 pregunta	4
Correlaciones	2 pregunta	2 preguntas	6 preguntas	10
Posiciones relativas	2 pregunta		4 preguntas	6
totales	7	6	14	27

# Entrega

Las evaluaciones escritas necesitan entregarse en un formato que no coloque en una desventaja injusta a los alumnos. Es mejor entregar pruebas tipeadas en vez de las escritas a mano, que pueden ser difíciles de leer para los alumnos.

<sup>&</sup>lt;sup>7</sup> Izard, J. F. (1977). Construction and Analysis of Classroom Tests, p. 8. Hawthorn: Australian Council for Educational Research.

#### Tiempo

El tiempo brindado debe permitir que entre el 90% y el 95% de estudiantes complete el conjunto de preguntas.

# 3.2. Construyendo evaluaciones: algunos ejemplos

Los diferentes formatos de respuestas escritas brindan diferentes acercamientos para evaluar lo que los estudiantes conocen, pueden hacer, y sus valores. De la misma manera que las consideraciones generales para desarrollar las evaluaciones discutidas anteriormente, hay consideraciones particulares para cada formato de respuesta.

A continuación, se describe una guía para construir evaluaciones utilizando ocho formatos de respuestas escritas comunes.

Los formatos son los siguientes:

- Procedimiento para rellenar espacios en blanco
- Mapas de conceptos
- Formatos de ensayos
- Investigaciones

- Escalas Likert
- Opciones múltiples
- Formato de respuestas cortas

# Procedimiento para rellenar espacios en blanco

Decidir qué resultados se desea obtener. Por ejemplo, el formato de rellenar los espacios en blanco puede ser utilizado como un vehículo para evaluar la comprensión de lectura, la comprensión de conceptos científicos o conocimientos de lazos cohesivos en la construcción del texto. Seleccionar un texto apropiado y dejar suficiente texto introductorio intacto para ser configurado en el contexto del pasaje.

Omitir las palabras o frases que se orientan a los resultados a ser evaluados. Considerar cada omisión cuidadosamente para ver cómo la cohesión del texto es interrumpida y ver qué tanta referencia hacia adelante o hacia atrás es requerida para mantener el significado. Utilice solo uno o dos espacios por oración. Evite pistas específicas como "una" antes de un espacio.

Brindar espacio suficiente para la respuesta y asegurarse de que todos los espacios sean iguales en tamaño.

Asegurar que las instrucciones a los estudiantes incluyan leer todo el pasaje antes de empezar a rellenar los espacios, leer el texto luego de los espacios para tener pistas del contexto y volver a leer el pasaje luego de completar los espacios para revisar que el pasaje tenga sentido.

Recordar a los estudiantes que más de una palabra puede haber sido omitida, si ese es el caso.

A continuación, mostramos un ejemplo de una prueba de espacios en blanco. Esta prueba ha sido diseñada para evaluar la comprensión de los estudiantes en cohesión a través del uso de pronombres para referirse a participantes establecidos en un texto.

# **Construyendo evaluaciones**

Este extracto de una prueba de espacios en blanco está diseñado para evaluar la comprensión de los estudiantes en cohesión a través del uso de pronombres para referirse a participantes establecidos en un texto.<sup>8</sup>

Había una vez un sapo que soña	oa con ser alguien especial, alguie	en valiente y noble
quería ser un rey, pero	sabía que eso no podría pasarle a	mientras
viviera en un estanque de lirios. F	or eso se aventuró haci	ia el mundo para buscar una
princesa. Como todos los sapos,	escuchó la historia d	de la princesa cuyo beso de
amor convertía a un sapo en un p	oríncipe encantador.	

<sup>8</sup> Clarke, J. *Monitoring Learning in English*, p. 59. Mount Coot-tha Support Centre.

# Mapas de conceptos

Decidir el área de conocimiento a ser mapeado.

Asegurar que los estudiantes comprendan el propósito del mapa de conceptos: producir una representación diagramática en la que los conceptos relacionados se encuentren conectados por tiempos, por fechas y por organización estructural.

Los estudiantes completan el mapa en cuatro pasos:

- Lluvia de ideas para identificar el vocabulario relacionado al concepto brindado;
- Clasificar las palabras jerárquicamente desde lo general hasta lo especifico;
- Dibujar líneas o flechas entre conceptos relacionados o interrelacionados; y
- Agregar etiquetas para aclarar y enlazar los conceptos.

# Formato de ensayo

Decidir los resultados que necesitan ser encontrados. Por ejemplo, si el área de estudio es Lenguaje, ¿necesitará utilizar el formato de ensayo para brindar la evidencia del control de los estudiantes sobre las características del texto (la calidad de las ideas y el sentido de la audiencia y el propósito) y/o las características del lenguaje (control de la ortografía, puntuación y gramática)?

Hacer que la introducción sea lo más clara posible, asegurando que los estudiantes sean orientados hacia la demostración de los resultados que se quiere obtener. Por ejemplo, si evalúa el control de los estudiantes sobre las características del texto del formato narrativo, asegúrese de que la introducción dirija a los estudiantes a desarrollar una narrativa y no un relato o una exposición.

# **Construyendo evaluaciones**

La siguiente pregunta está diseñada para obtener resultados de ciencia para las facetas de "trabajando científicamente" y "vida y vivienda" del perfil australiano:

"Explicar cómo los seres vivientes obtienen, almacenan y transportan nutrientes, transportan energía y administran los desperdicios" ("vida y vivienda", Nivel 6).

"Seleccionar las formas de presentar información que aclare los patrones y ayude a hacer generalizaciones".9

#### Nutrición de las plantas

Las personas normalmente dicen que las plantas obtienen su alimento de la tierra. ¿Está de acuerdo con esta afirmación o no? Explique cuidadosamente lo que piensa, poniendo atención especial a lo que se refiere por alimento.

Asegurar que la introducción no demande conocimientos o habilidades que no son importantes para su propósito. Por ejemplo, si va a evaluar las habilidades del estudiante para construir un argumento lógico, asegúrese de que la introducción no requiera otro conocimiento fuera de su experiencia. Los estudiantes de 6.<sup>to</sup> año puede ser capaces de construir un argumento orientado por la pregunta "¿los gatos como mascota deberían mantenerse dentro de la casa por las noches?", pero no podrán construir un argumento sobre los méritos de la cobertura de salud privada. Si el resultado que se desea obtener requiere que los estudiantes demuestren un conocimiento base, asegúrese de que la pregunta requiera a los estudiantes mostrar dominio de conocimiento esencial.

Asegurar también que la introducción sea lo suficientemente abierta para permitir a los estudiantes de diferentes antecedentes culturales y de lenguaje, y de diferentes niveles de habilidades, involucrarse con la pregunta.

<sup>&</sup>lt;sup>9</sup> Curriculum Corporation (1994). Science – A curriculum Profile for Australian Schools, p. 93. Carlton: Curriculum Corporation.

Asegurar que los estudiantes comprendan lo que se quiere obtener. Por ejemplo, si le dice a sus estudiantes que solo quiere conocer su comprensión de conceptos científicos, entonces no es ético calificar en contra por faltas de ortografía, puntuación o caligrafía.

Pruebe la pregunta escribiendo una respuesta ideal para ella.

En la página anterior, se muestra un intento de una introducción para un ensayo. Esta pregunta está diseñada para obtener resultados de ciencia en las facetas de "trabajando científicamente" y "vida y vivienda" del perfil australiano.

# Investigaciones

Decidir sobre los resultados que se quiere obtener.

Asegurar que el tema esté estructurado de manera suficiente para dirigir a los estudiantes a obtener los resultados que desea evaluar y que la pregunta sea apropiada para el tiempo brindado. Asegurar que el contexto sea familiar, interesante y cautivador.

Los buenos temas, al mismo tiempo que brindan estructura, también están abiertos a permitir a los estudiantes llevar a cabo la investigación en una variedad de formas y permiten a los estudiantes brindar múltiples soluciones.

#### **Escalas Likert**

Decidir los resultados, opiniones, actitudes, valores o intereses que se quiere obtener.

Decidir el conjunto de categorías ordenadas para ser utilizadas, por ejemplo: Fuertemente en desacuerdo, En desacuerdo, De acuerdo, Fuertemente de acuerdo. Tratar de tener un número equitativo de categorías para prevenir "observadores pasivos" que seleccionan la categoría central.

Las oraciones deben ser lo más claras y específicas posibles. Las oraciones que son muy generales pueden sugerir respuestas como "depende".

Si el propósito del cuestionario es reunir información del nivel de la clase, los estudiantes deben responder de forma anónima. De esa manera pueden responder honestamente.

En la siguiente página, se muestra un ejemplo de una encuesta estudiantil diseñada para brindar al profesor información sobre aspectos importantes de la experiencia del salón de clase desde el punto de vista de los estudiantes. El profesor que utiliza este cuestionario puede considerar respuestas a cada pregunta por separado.



# **Cuestionarios** Esta encuesta estudiantil está diseñada para brindar a los profesores información sobre aspectos de la experiencia del salón desde el punto de vista del estudiante. 1º Los estudiantes pueden responder anónimamente a una encuesta de este tipo. Las respuestas a cada pregunta son consideradas Inseguro de dónde estoy Falto de organización Mal sobre el trabajo Inquieto, aburrido No preparado No servicial Incómodo Ignorado Injusto Solo por separado y no se combinan para obtener una medición de las actitudes de los estudiantes. **Construyendo evaluaciones** Involucrado en las lecciones En esta clase me siento Seguro de dónde estoy El profesor ha estado Bien sobre el trabajo Parte de un grupo Bien organizado Importante Preparado Cómodo Servicial Justo

<sup>10</sup> Wiggins, G. (1998). Educative Assessment, p. 294. San Francisco: Jossey-Bass Publishers.

Confuso sobre lo que espera Insensible a mis necesidades Aparentemente aburrido Sin saber el tema	No al tanto del tema  Aburrido Inefectivo para ayudarme a aprender Muy lento Muy difícil Muy impredecible Muy simple Demasiado
Transparente sobre lo que espera Susceptible a mis necesidades Completamente involucrado y emocionado	Dispuesto a hacer las ideas difíciles accesibles e interesantes  Nuestro trabajo ha sido en general Incitante a pensar Efectivo para ayudarme a aprender Muy rápido Muy fácil Muy abstracto Muy poco

# **Opciones múltiples**

Decidir los resultados que se desea conseguir y diseñar cada pregunta para obtener un resultado de aprendizaje.

# Subcategorías

Las subcategorías (preguntas o afirmaciones incompletas) deberían enfocar a los estudiantes en lo que se les ha preguntado. Los estudiantes deben tener alguna idea sobre lo que se les ha preguntado antes de leer las opciones.

Luego	o Sally
	cenó.
	fue a casa.
	se acostó.
	leyó un libro.
¿Qué	hizo luego Sally?
	Cenó.
	Fue a casa.
	Se acostó.
	Leyó un libro.

Para minimizar la cantidad de lectura requerida, se debe colocar la mayor parte de palabras posibles en las subcategorías.

Jim fue		
	al río con sus amigos.	
	al parque con sus amigos.	
	a la playa con sus amigos.	
	al cine con sus amigos.	

Jim y sus amigos fueron
al río.
al parque.
a la playa.
al cine.
Evite utilizar lenguaje superfluo en las subcategorías. "Uno de los colores en la pintura es" es preferible a "la pintura utiliza tres colores. Uno de ellos es".
Fije las subcategorías de forma positiva cuando sea posible.
Las alternativas
Si las alternativas completan las subcategorías para formar una oración, entonces cada alternativa debería ser finalizada con un punto final. Si las alternativas forman oraciones completas por sí solas, ellas deberían ser puntuadas según corresponda. Si las alternativas son palabras solas o frases, no se necesita puntuación.
Por ejemplo:
El símbolo en la figura muestra la forma del café.
tienda.
baños.
guardería.
¿Qué le dijo el padre a su hijo cuando se molestó?
"¿Cómo pudiste hacer eso?"
"¡Ve a tu cuarto de una vez!"

	"Cuando te comportas así me siento muy molesto."
	"Esta es la última vez que te llevo al fútbol."
¿Cómo se le	notificará al ganador?
	Por carta
	Por teléfono
	En televisión
	En el periódico

La respuesta debe ser correcta y no solo la mejor de las alternativas presentadas. La respuesta no debe sobresalir entre las alternativas porque es más larga, mejor escrita, o si tiene alguna otra cualidad superficial. Cuando se coloca una serie de preguntas de opción múltiple en una prueba, revise que la respuesta correcta no se encuentre siempre en la misma posición en el conjunto de alternativas.

Construyendo evaluaciones		
Esta pregunta de opción múltiple consigue el resultado: "Identificar los procesos de energía transferida y condiciones que las afectan". Note que los distractores son diseñados para diagnosticar la mala interpretación conceptual de los alumnos.  La cantidad de energía disipada cada segundo en un circuito depende de		
solo el voltaje. solo la corriente. la corriente y el voltaje. la resistencia y el voltaje.		

<sup>&</sup>lt;sup>11</sup> Curriculum Corporation (1994). Science – A curriculum Profile for Australian Schools, p. 7. Carlton: Curriculum Corporation.

Los distractores (alternativas incorrectas) deben ser completamente incorrectos, pero razonables y plausibles. Cualquier distractor que sea absurdamente incorrecto reduce el número de opciones reales disponibles al estudiante, y no contribuye en nada a la pregunta. Los distractores deben tener diferentes significados entre ellos. El significado de una alternativa nunca debe ser incluido por otro como en "ella tomó: una manzana, un suéter, algo de protector solar, algo de comida". Las alternativas deben ser fijadas desde la más corta a la más larga, de la más larga a la más corta, o en orden lógico.

Utilizar cuatro o cinco alternativas para reducir la oportunidad del estudiante a seleccionar la respuesta correcta adivinando.

Evitar utilizar "todas las anteriores" o "ninguna de las anteriores".

Algunas veces, los distractores pueden ser diseñados para recopilar información sobre la mala interpretación conceptual de los estudiantes, como en el ejemplo de la página 69. Esta pregunta ha sido diseñada para conseguir el resultado de "Identificar los procesos de energía transferida y condiciones que las afectan".

# Formato de respuestas cortas

Decidir qué resultados se quiere obtener. Asegurar que cada pregunta esté dirigida hacia ese resultado. Decidir, mientras las preguntas son desarrolladas, si hay solo una respuesta correcta o múltiples respuestas aceptables. Si hay múltiples respuestas, decidir si se utilizará una guía de puntaje parcial.

# 3.3. Calificar y registrar

Los profesores utilizan una variedad de métodos para calificar y registrar la calidad de las respuestas de los estudiantes a preguntas escritas. Algunos formatos de respuestas escritas (por ejemplo, preguntas de opción múltiple o de verdadero/falso) no requieren que los profesores utilicen un dictamen

cuando evalúan las respuestas de los estudiantes. Se ha realizado un dictamen sobre la respuesta correcta durante la construcción de la pregunta. Otros formatos, sin embargo, requieren que los profesores califiquen la calidad de la respuesta.

Para desarrollar un criterio de calificación de la calidad de las respuestas de los estudiantes, son importantes algunas consideraciones. El criterio necesita orientarse a los resultados que serán medidos. Ellos necesitan ser claros para los estudiantes, padres y otros profesores. Deben ser imparciales y no reflejar variables sobre las cuales los estudiantes o los profesores no tienen control. Y necesitan ser útiles. Lo que significa, para cualquier pregunta, que el número de criterios en los que el trabajo será calificado necesita ser delimitado, o el calificar consumirá mucho tiempo y será muy pesado.

Los métodos para calificar y registrar la calidad de las respuestas escritas de los estudiantes incluyen:

- guía de puntaje;
- categorías de diagnóstico;
- puntuación analítica; y
- puntuación holística.

Las consideraciones para el desarrollo del criterio para calificar la calidad de las respuestas de los estudiantes son:

- orientar resultados
- claridad
- imparcialidad
- utilidad

Los métodos para calificar y registrar la calidad de las respuestas escritas de los estudiantes incluyen:

- guía de puntaje;
- categorías de diagnóstico;
- puntuación analítica; y
- puntuación holística.

# 3.4. Guía de puntaje

Las guías de puntaje son utilizadas para calificar la calidad de las respuestas de los estudiantes a preguntas cortas. Algunas veces las respuestas son calificadas de forma dicotómica (correcto/incorrecto). Algunas veces, las respuestas son calificadas de una forma que permita a los profesores reconocer y registrar los diferentes niveles del estudiante sobre un éxito parcial, su comprensión y estrategia parcialmente correcta. La puntuación de este tipo es llamada puntuación de "crédito parcial".

El primer paso para desarrollar una guía de puntaje es tener claro el resultado que se quiere obtener con la pregunta. Luego, puede desarrollarse el criterio de evaluación para calificar las respuestas del estudiante. Por ejemplo, si se evalúa la comprensión conceptual del estudiante en ciencia o en comprensión de lectura, entonces el criterio de evaluación probablemente no incluirá una correcta ortografía.

Las guías de puntuación son las más utilizadas cuando el criterio para evaluar es respaldado por ejemplos de respuestas correctas o incorrectas del estudiante. A continuación se muestra un ejemplo de una guía de puntaje dicotómica para una pregunta de

matemáticas y un ejemplo de una guía de puntaje de crédito parcial para una pregunta de comprensión visual.

# Calificar y registrar

# **Puntuación dicotómica**

Mostramos un ejemplo de guía de puntaje dicotómico para una pregunta de matemáticas. La pregunta viene de una unidad diseñada sobre un mapa de un zoológico. En esta pregunta, se califica el uso del lenguaje de posición y dirección del estudiante. Tome nota de los ejemplos de las respuestas de los estudiantes utilizados para respaldar la quía de puntaje. 12

#### Pregunta:

Explique cómo ir desde la entrada del tren hasta la zona de las focas.

#### Guía de puntaje:

- Brindar detalles adecuados y correctos: debe ser desde la entrada del tren, y las direcciones para girar deben ser las correctas.
- Ej.: Ir de frente y girar hacia la derecha en las cacatúas.
  - Ir de frente hasta pasar el teléfono en la cacatúa y girar hacia a la derecha.
  - Pasar el baño y la zona de picnic. Continuar hasta pasar los leones, los quioscos, los felinos pequeños y, luego, girar en la primera esquina izquierda.

Marcar como correcta una respuesta que utiliza la convención de norte como la parte superior de la página, es decir "ir hacia el sur y luego voltear hacia el oeste".

#### 0 Direcciones incorrectas

Ej.: • Pasar la escuela del zoológico y girar hacia la izquierda.



<sup>&</sup>lt;sup>12</sup> Recht, E.; Forster, M. y Masters, G. (1998). *DART Mathematics Upper Primary*. Camberwell: Australian Council for Educational Research.

# **Calificar y registrar**

# Puntaje de crédito parcial

El siguiente ejemplo de guía de puntaje de crédito parcial es para una pregunta de comprensión visual. Los estudiantes ven un video corto y luego responden una serie de preguntas. Estas preguntas evalúan las habilidades del estudiante para reconocer el significado central de un evento. Los estudiantes que reconocen la posición narrativa de la escena, pero no la importancia central, obtienen solo un crédito parcial por comprensión. Tome nota que los ejemplos utilizados sobre las respuestas de los estudiantes respaldan la guía de puntaje.<sup>13</sup>

#### Pregunta:

¿Por qué la historia comienza con el sol creciente?

#### Guía de puntaje:

- 2 Muestra explícitamente la importancia del trabajo de una lechuza o el rol central del sol en la historia
- Ej.: Porque se trata de cómo el sol no asciende luego.
  - Para que sepas lo que hace la lechuza.
  - Es el tema principal.
- 1 Se refiere al uso "inicio" de una manera general o respuesta narrativa (ulular de la lechuza)
- Ej.: Porque es el inicio de la historia.
  - Es el inicio del día.
  - Porque la lechuza ulula.

#### 0 Otra razón

- Ej.: Porque es de día.
  - · Para darle a la historia diferentes imágenes.
  - Para que los animales despierten.

# 3.5. Categorías de diagnóstico

Otro método para calificar la calidad de las respuestas de los estudiantes es distribuirlas en categorías predefinidas de respuestas a una pregunta. Estas categorías pueden ser construidas para representar el incremento de comprensión sofisticada de una pregunta o, como en el siguiente ejemplo, brindar información sobre los diferentes tipos de comprensión y estrategias.

<sup>&</sup>lt;sup>13</sup> Bodey, W.; Darkin, L.; Forster, M. y Masters, G. (1997). DART English Middle Primary. Camberwell: Australian Council for Educational Research.

		•	e respuestas para para completar es	pacios en blanco.		
Palabras borradas en los espacios en blanco	Palabras utilizadas	¿Palabra correcta?	Semánticamente aceptable	Sintácticamente correcta	Análisis de ortografía	Comentarios
correr	resbalar	No	Sí	Sí	Doble consonante	Repasar dobles consonantes
nadó	nadó	Sí				
camisa	camesa	Sí	Sí		Confusión -e, -i	Agregar a la lista de ortografía

#### 3.6. Escalas de puntuación

Las escalas de puntuación son normalmente utilizadas para ayudar al evaluador a hacer calificaciones sobre respuestas extensas. Al utilizar escalas de puntuación, los evaluadores califican la calidad del trabajo del estudiante sobre un criterio específico. Los evaluadores realizan calificaciones analíticas del trabajo cuando califican diferentes aspectos del trabajo. Ellos realizan calificaciones holísticas cuando hacen una sola calificación basada en la impresión general del trabajo.

Al desarrollar escalas de puntuación analítica y holística, el primer paso es considerar el resultado que se quiere obtener. El criterio de evaluación puede ser luego desarrollado para reflejar aquellos resultados. Por ejemplo, un formato de respuesta de ensayo puede ser utilizado para reunir evidencia sobre las habilidades del estudiante para estructurar un tipo particular de texto y controlar las características más evidentes de la escritura. Luego, el

<sup>&</sup>lt;sup>14</sup> Clarke, J. *Monitoring Learning in English*, p. 55. Mount Coot-tha Support Centre.

criterio en una escala de calificación analítica incluirá el control de una forma de texto particular (por ejemplo, narrativo o expositivo) de la misma forma que el control sobre las características más evidentes de la escritura (ortografía, puntuación y gramática).

A continuación, mostramos un ejemplo de escala de calificación analítica para escritura de no ficción, y para una investigación matemática y un mapa de conceptos en las páginas 77 y 78. En las páginas 79-80, mostramos un ejemplo de escala de calificación holística para evaluar una investigación matemática de respuesta abierta.

Calificar y registrar Escala de calificación analítica para escritura de no ficción							
Los profesores que utilizan la siguiente guía de evaluación, califican diez criterios utilizando una escala de 3 puntos. El criterio orienta los r no ficción así como a resultados relacionados con habilidades generale mecánica de la escritura. <sup>15</sup>	resultados espe	ecíficos a es	critura de				
Criterio de evaluación	Alto	Medio	Bajo				
<ol> <li>La escritura demuestra una habilidad de interpretar ideas significativamente en el contexto.</li> </ol>							
2 La "gran idea" del texto es interesante y clara.							
3 Todas las ideas principales son claras en relación a la "gran idea".							
4 Las ideas principales son organizadas en una secuencia lógica.							
5 Las transiciones desde una idea principal a la siguiente son fluidas.							
6 Hay suficientes detalles apropiados y acertados para respaldar cada idea principal.							
7 La elección de palabras es apropiada, variada y crea una voz natural.							

<sup>&</sup>lt;sup>15</sup> Adaptado de Hibbard, M. et al. (1996). A Teacher's guide to Performance-Based Learning and Assessment, p. 186. Virginia: Association for Supervision and Curriculum Development.

8 La mecánica y gramática se integran al significado y el efecto de la escritura.	
<b>9</b> La variedad estilística y la elección son evidentes en elementos retóricos como dicción, sintaxis, estructura y lenguaje figurativo.	
10 El texto es impecable y presentable.	

# Calificar y registrar Escala de calificación analítica para una investigación matemática

Los profesores que utilizan esta guía de puntaje en la Escuela Estatal Gap evalúan investigaciones matemáticas sobre seis criterios utilizando una escala de puntaje de 4 puntos. Dos de los criterios ("habilidades grupales" y "actitud y esfuerzo") requieren la observación del profesor a los estudiantes mientras ellos realizan el trabajo.¹6

	Estándar muy alto	Estándar alto	Satisfactorio	Experimenta dificultades
	Bien establecido	Establecido	En desarrollo	Inicial
Lenguaje matemático	Extenso y uso relevante.	Uso competente.	Uso apropiado de términos básicos.	Uso incorrecto o insuficiente de términos.
Comunicación	Ideas claras, concisas, informativas, bien organizadas.	Ideas explicadas con comprensión.	Comprensión básica.	Habilidad limitada para explicar las ideas de forma lógica.
Habilidades grupales	Participación activa.	Participación frecuente.	Participación ocasional.	Poca o ninguna participación.
Estrategias de investigación	El acercamiento es lógicamente consistente y sistemático.	El acercamiento es estructurado y sistemático.	Intenta utilizar un acercamiento estructurado y sistemático.	Poca evidencia de algún plan sistemático.

<sup>&</sup>lt;sup>16</sup> Holcroft, L. y Coates, R. *Monitoring Learning in Mathematics*, pp. 20-21. Brisbane: Mount Coot-tha School Support Centre.

Interpretación de información y representación	Demuestra presentación de información acertada y adecuada y un amplio rango de habilidades manipulativas.	Vínculos estructurados en la presentación de la información. Demuestra una mayoría de habilidades manipulativas.	Demuestra habilidades manipulativas básicas. La presentación de la información no siempre es completa.	La presentación de la información es incompleta e inadecuada con pocas habilidades manipulativas.
Actitud y esfuerzo	Consistentemente entusiasta, confiado, persistente. Creativo e innovador.	Entusiasta, y generalmente confiado y persistente.	Muestra interés, pero falta de confianza y persistencia.	Muestra poco esfuerzo, interés o entusiasmo.

Calificar y registrar Escala de calificación analítica para mapas de conceptos							
Los profesores que utilizan esta escala de c evalúan la calidad de los mapas de conceptos una escala de puntuación de 4 puntos. <sup>17</sup>							
Criterio de evaluación	Con ayuda	Satisfactorio	Capaz	Eficiente			
1 Identifica ideas principales.							
2 Clasifica y categoriza información.							
3 Hacer inferencias al enlazar la información							

<sup>&</sup>lt;sup>17</sup> Basado en Clarke, J. *Monitoring Learning in English,* pp. 85. Brisbane: Mount Coot-tha School Support Centre.

## Calificar y registrar Escala de calificación holística para problemas matemáticos de respuesta abierta

Esta guía de puntaje genérico para problemas matemáticos con concluyentes fue desarrollada en el Programa de Calificación de California.<sup>18</sup> Las calificaciones holísticas pueden ser utilizadas para evaluar las respuestas a un rango de problemas matemáticos de diferentes niveles de diferentes.

#### Rúbrica para problemas matemáticos de respuesta abierta Competencia demostrada

#### Respuesta ejemplar

6 puntos

Brinda una respuesta completa con una explicación clara coherente, sin ambigüedades y elegante. Incluye un diagrama claro y simplificado. Comunica de manera efectiva a la audiencia identificada. Muestra comprensión de las ideas y los procesos del problema matemático. Identifica todos los elementos importantes del problema. Puede incluir ejemplos y contra ejemplos. Presenta fuertes argumentos de respaldo.

#### Respuesta competente

5 puntos

Brinda una respuesta completamente válida con explicaciones razonablemente claras. Puede incluir un diagrama apropiado. Comunica efectivamente a la audiencia identificada. Muestra comprensión de las ideas y los procesos del problema. Identifica los elementos más importantes del problema. Presenta sólidos argumentos de respaldo.

#### Respuesta satisfactoria

#### Fallas menores pero satisfactorio

4 puntos

Completa el problema de forma satisfactoria, pero la explicación puede ser confusa. La argumentación puede estar incompleta. El diagrama puede ser inapropiado y confuso. Comprende las ideas matemáticas subyacentes. Utiliza las ideas de forma efectiva.

#### Fallas mayores, pero casi satisfactorio

3 puntos

Empieza el problema apropiadamente, pero puede fallar en completar u omitir partes significativas del problema. Puede fallar en demostrar una comprensión completa de las ideas y los procesos matemáticos, puede cometer errores de cálculo. Puede hacer mal uso o fallar en el uso de términos matemáticos. La respuesta puede reflejar una estrategia inapropiada para resolver problemas.

<sup>&</sup>lt;sup>18</sup> Wiggins, G. (1998: 160).

#### Respuesta inadecuada

#### Empieza, pero falla al completar el problema

2 puntos

La explicación no es comprensible. El diagrama es confuso. No muestra comprensión de la situación del problema. Puede cometer errores de cálculo mayores.

#### No empieza efectivamente

1 punto

Las palabras que utiliza no reflejan el problema. Los dibujos no representan la situación del problema. Falla al indicar qué información es apropiada.

No hace el intento

0 puntos

#### 3.7. Diferentes evaluadores

A pesar de que las respuestas de las evaluaciones escritas son normalmente calificadas por los profesores, ellas también pueden ser evaluadas por otros compañeros de clase o por el mismo estudiante.

La evaluación de un compañero o la autoevaluación puede ayudar al estudiante a reflejar lo que saben y cómo han aprendido, y así fomentar la responsabilidad de aprender.

En las página siguiente y en la 84 mostramos dos ejemplos de guías de puntaje desarrolladas para el uso del estudiante. El primer ejemplo ayuda a los estudiantes jóvenes a evaluar sus descripciones sobre el ciclo de vida de una mariposa. El segundo ejemplo, desarrollado por la Facultad de Leyes de la Universidad de Wollongong, es utilizado para recolectar evidencia de las habilidades del estudiante para evaluar su propio desempeño, un objetivo de aprendizaje establecido de la facultad.

En la página 82, también se muestra un ejemplo de una escala de calificación analítica utilizada por los profesores para evaluar la autorreflexión de los estudiantes.

# 3.8. La necesidad de comparar las calificaciones

En el contexto del salón de clases, normalmente no es de gran preocupación si otro profesor realiza la misma calificación al trabajo de un estudiante utilizando la misma guía de puntaje. Sin embargo, en contextos de alto rendimiento, la fiabilidad entre evaluadores es crucial.

Usualmente, mientras mayor sea el requerimiento de comparación, mayor será el requerimiento de especificar el criterio de evaluación. Los evaluadores están capacitados y el proceso de calificación es cuidadosamente monitoreado para asegurar un alto nivel de acierto entre evaluadores. En algunos casos de alto rendimiento, el trabajo escrito es calificado por varios evaluadores.

Autoevaluación utilizando la escala d	е санпса	cion a	inalitica
Esta guía de autoevaluación fue desarrollada por estudio califican su descripción del ciclo de vida de la mariposa escala de puntaje de 3 puntos. 19			
	Necesita trabajar	ОК	Magnífico
1 ¿Dibujé las cuatro partes del ciclo de la mariposa?	<b>:</b>	<u></u>	☺
2 ¿Mostré detalles?	<b>(1)</b>	<b>:</b>	☺
3 ¿Etiqueté las partes de cada figura?		<b>(</b>	☺
4 ¿Escribí una oración para cada figura?		<u></u>	☺
5 ¿Mis oraciones dicen lo que muestran las figuras?		<u></u>	☺
6 ¿Empecé cada oración con una letra mayúscula?		<u></u>	☺
7 ¿Terminé cada oración con un punto?		<u></u>	☺
8 ¿Mi trabajo es impecable?	<b>:</b>	٥	©

Calificar y registrar

<sup>&</sup>lt;sup>19</sup> Hibbard, M. et al. (1996: 132-133).

# El siguiente es un extracto de una escala de calificación analítica utilizado por profesores para evaluar la autoevaluación de los alumnos.<sup>20</sup> Los profesores califican las autoevaluaciones sobre criterios utilizando una escala de puntaie de 5 puntos. Aquí mostramos cuatro de los once criterios. **Evidencia** extensa Evaluar la autoevaluación del estudiante utilizando la escala de calificación analítica **Evidencia** significativa Alguna evidencia Poca evidencia Sin evidencia ¿El estudiante ha descrito su fortaleza de escritura de ¿El estudiante ha expresado el sentido general sobre ¿El estudiante ha descrito su debilidad de escritura ¿El estudiante ha descrito el uso de su proceso de quién es como escritor/escritora? escritura de forma acertada? Calificar y registrar de forma acertada? forma acertada?

<sup>20</sup> Hibbard, M. et al. (1996: 207).

# Calificar y registrar

#### **Autoevaluación**

El siguiente ejemplo de la guía de autoevaluación de un estudiante fue desarrollado por la Facultad de Leyes de la Universidad de Wollongong.<sup>21</sup> La escala de calificación analítica aparece en la siguiente hoja. A continuación, reproducimos la explicación brindada por los estudiantes. La hoja de autoevaluación es utilizada como evidencia de la habilidad de los estudiantes para evaluar su propio desempeño, un estado de aprendizaje objetivo de la facultad.

#### Una nota sobre la autoevaluación

Es necesario que evalúe su propia participación en clase. Esto no es una evasión de responsabilidad del personal académico, puesto que realizaremos una evaluación de su participación. Pero uno de los objetivos de los temas legales de Wollongong es permitir que ustedes desarrollen sus propias habilidades de aprendizaje. Usted necesita evaluar su propio desempeño de aprendizaje en varias etapas de su carrera, por eso, en esta clase, y en otras clases, recibirá ayuda para desarrollar su habilidad para evaluar su propio desempeño.

La evaluación académica pretende ser un proceso en el cual se mida el logro de un objetivo específico. Los objetivos de la participación en clase son:

- 1 Brindar una oportunidad para obtener el desarrollo de sus habilidades de aprendizaje:
  - · reforzar su propio aprendizaje
  - brindar un incentivo para su plan, distribuir y manejar su propio tiempo incluyendo el desarrollo de técnicas de lectura selectiva
  - realizar calificaciones sobre la prioridad que se le da a diferentes tareas
- 2 Ayudar a otros miembros de la clase a dar y recibir retroalimentación sobre su aprendizaje.
- 3 Indicar al personal académico si ha comprendido el material o si necesita ayuda.
- 4 Ayudarlo a desarrollar las habilidades de:
  - expresar argumentos de forma oral
  - · expresar ideas abstractas y complejas de forma oral
  - escuchar
  - comprender

En este contexto, hemos estipulado lo que creemos son los objetivos del proceso. En la vida real, necesitará trabajar en sus propios objetivos y en la forma en la que son medidos. Para este propósito, usted tendrá que trabajar el criterio en el que quiere evaluar si logra o no los objetivos.

Cada cuatro semanas se le pedirá que complete un formulario que le pedirá calificar el logro en cada uno de los objetivos enlistados anteriormente en una escala de seis puntos y que brinde un puntaje general. Este formulario también le dará la oportunidad de indicar los factores anormales que afectan su trabajo. Compararemos su autoevaluación con nuestra evaluación de su participación, y discutiremos con usted si hay una percepción diferente de lo que encontramos en su participación. Debe estar consciente de los factores que no conocemos y que afectan su participación. La última figura registrada debe ser un acuerdo; sin embargo, nuestro personal académico se reservará el poder final de decisión. Cuando hemos utilizado este método anteriormente, hemos encontrado que, inicialmente, los estudiantes tienden a subestimar su propia participación, pero al final de la sesión existe una buena medición de acierto entre el personal y los estudiantes.

<sup>&</sup>lt;sup>21</sup> Nightingale, P.; Te Wiata, I.; Toohey, S.; Ryan, G.; Hughes, C. y Magin, D. (1996). *Assessing Learning in Universities*, pp. 243-244. University of New South Wales: Professional Development Centre.

# UNIVERSIDAD DE WOLLONGONG **FACULTAD DE LEYES** Hoja de autoevaluación N.º:\_\_\_\_\_ Nombre del estudiante:\_\_\_ Profesor/nombre del ponente:\_\_\_\_\_ \_\_\_\_ Desde: \_\_\_/\_\_\_/\_\_\_ Hasta: \_\_\_/\_\_\_/\_\_\_ Durante este período evaluaré mi participación en cada una de las siguientes áreas: Nota: MS (86-100%) sobresaliente en todos los aspectos S (76-85%) muy sobre el promedio. Muy buen logro general SP (65-75%) sobre el promedio. Muy buen logro parcial P (50-64%) aceptable / promedio B (45-50%) bajo MB (0-44%) extremadamente bajo MARQUE LA CASILLA EN LA COLUMNA **APROPIADA** Área de actividad MB Aprendizaje y comprensión del material Planeamiento de estudio; cobertura del material asignado Organización de estudio y preparación Ayuda a otros miembros de la clase en grupos completos de la clase en pequeños grupos y sindicatos informalmente fuera de la clase Expresar y mostrar argumentos de forma oral Comprender y expresar ideas complejas y abstractas de forma oral Escuchar Comprensión de la discusión en clase Los factores especiales que me afectaron durante esta clase fueron:

En general, evalúo mi participación en clase durante este período con \_\_\_\_\_\_%

#### 3.9. Estimar y reportar el logro

Las evaluaciones escritas pueden brindar a los profesores evidencia que puede ser utilizada para tener una inferencia y reportar sobre el nivel de logro de un estudiante en un aspecto de su aprendizaje. Para poder estimar y reportar este logro, los profesores realizan una calificación equilibrada de sus registros del trabajo del estudiante. Las calificaciones de este tipo se realizan con base en evaluaciones que solo se enfocan en un área de aprendizaje. Por ejemplo, es usual evaluar el logro de un estudiante en razonamiento numérico de forma separada a alguna otra dimensión junto con el posible progreso del estudiante (como lectura, escritura, y lenguaje hablado) a pesar de que estos aspectos de desarrollo puedan estar relacionados.

Los profesores utilizan las evaluaciones escritas para recolectar información para muchos propósitos, incluyendo reunir información sobre las actitudes y reflexiones del estudiante sobre su propio progreso. Algunas veces, la intención de reunir este tipo de información es reportar las respuestas por separado a cada pregunta, y no combinar respuestas a lo largo de las preguntas. Por ejemplo, un profesor puede construir un cuestionario hecho con preguntas como: "¿te gustan las clases de matemáticas?" y "¿qué área de las matemáticas te parece más difícil?"

Para poder estimar y reportar el logro del estudiante, los profesores necesitan hacer una calificación equilibrada de los registros de los trabajos del estudiante. La validación, fiabilidad y objetividad de los estimados del logro del estudiante dependerán de la calidad de evidencia en la que se han basado los estimados.

con la intención de analizar la respuesta de los estudiantes a cada pregunta por separado.

Pero en una evaluación de desarrollo, las respuestas de los estudiantes a las preguntas o tareas se combinan para estimar la ubicación de cada estudiante en un mapa de desarrollo de conocimiento, habilidades comprensión, actitudes o valores.

Estas preguntas pueden ser afirmaciones en un cuestionario de actitud (por ejemplo, diseñadas para medir la actitud hacia las matemáticas) o preguntas en una prueba (por ejemplo, diseñadas para medir el logro en un aspecto de las matemáticas).

Un mapa de progreso describe el camino de un desarrollo normal, por eso no es común una coincidencia perfecta entre lo que se describe en el mapa y la evidencia reunida para un estudiante en particular. Se debe hacer el mejor estimado de la evidencia disponible. La validación, fiabilidad y objetividad del estimado dependerá de la calidad de la evidencia en la que se basa el estimado.

La validación del estimado depende de la relevancia de la evidencia. Cuando se planea una evaluación escrita, los profesores necesitan asegurar que las preguntas brindan la evidencia sobre los resultados de aprendizaje relevantes.

La *fiabilidad* del estimado depende de la cantidad de evidencia en la que se ha basado. Generalmente, mientras más evidencia sea utilizada, más fiable será el estimado.

La *objetividad* del estimado depende de la medida en que este no haya sido afectado por la elección de la pregunta o la elección del evaluador.

El proceso de estimar un nivel de logro del estudiante en un mapa de progreso es relativamente directo cuando las calificaciones coinciden directamente con resultados y niveles en un mapa de progreso. Por ejemplo, las escalas de puntuación pueden ser construidas de un modo en que el criterio obtenga resultados en un mapa de progreso para que los puntos en la escala correspondan exactamente con los niveles de un mapa de progreso.

Cuando los profesores basan sus procedimientos de reporte en los principios de la evaluación de desarrollo, los reportes se construyen en torno al concepto de un mapa de progreso. Estos reportes brindan un estimado del nivel de logro de los estudiantes y descripciones de los tipos de conocimiento, habilidades, actitudes o valores en cada nivel.

Las descripciones usualmente se encuentran en el lenguaje de los resultados del mapa de progreso, pero sin el argot. El reporte de los padres algunas veces incluye la información sobre el logro de otros estudiantes del mismo grado.

En las páginas 88 y 89 mostramos dos ejemplos de reportes de Lenguaje. El primero muestra el logro de un estudiante en relación al 80% de estudiantes en la clase. El segundo, muestra el logro del estudiante en relación al grado percentil (en este caso la media del 80% y el 10% más alto y más bajo de estudiantes en el grado) y el promedio del grado.

Ambos reportes están acompañados por descripciones de los tipos de conocimiento, habilidades y comprensión normalmente demostrados en cada nivel de logro.

# **Ejemplo 1:** Este reporte brinda información sobre el nivel de logro del estudiante en un mapa de progreso e información normativa sobre el logro en relación a la clase.<sup>22</sup> Profesor:\_\_\_\_\_ Clase:\_\_\_\_ Fecha:\_\_\_ 2 Hablar y escuchar Leer Ecribir 80% de estudiantes de la clase en este rango. Nivel de logro de los estudiantes. Comentarios generales: Este tipo de reporte brinda información normativa sobre el logro de los estudiantes en relación a la clase.



<sup>&</sup>lt;sup>22</sup> NSW Department of School Education (1994). *Preliminary Ideas on Assessment and Reporting,* p. 50.

# **Ejemplo 2:** Este reporte brinda información sobre el nivel de logro de un estudiante en un mapa de progreso e información normativa sobre lo que ha logrado en relación al grado percentil y el promedio.23 Estudiante: Profesor:\_\_\_\_\_ Fecha: **NIVEL** Faceta Transición 3 Base 1 2 Hablar y escuchar Leer Escribir Nivel de logro de los estudiantes Promedio del grado Representa el logro del 10% inferior y superior de estudiantes en el grado Comentarios: Firma del profesor:\_\_\_\_\_ Firma del director:\_\_\_\_ Comentario de los padres:\_\_\_ Este tipo de reporte brinda información normativa sobre el estudiante en relación al grado percentil y el promedio.

<sup>&</sup>lt;sup>23</sup> NSW Department of School Education (1994: 51).

# 3.10. Calificar estrategias de evaluaciones y reporte

Algunos profesores califican sistemáticamente sus evaluaciones y estrategias de reporte para poder mejorar la calidad de sus prácticas. En esta página mostramos un ejemplo de cuestionario para la autoevaluación de profesores. Este cuestionario se enfoca en la efectividad de una evaluación en particular.

En la siguiente página, mostramos un ejemplo de cuestionario para el alumno. Este cuestionario está diseñado para brindar retroalimentación en la efectividad de la evaluación y reporte de un curso completo.

Calificar estrategias de evaluación		Autoevaluación
Calificar regularmente las estrategias de evaluación puede evaluación. Este cuestionario de autoevaluación puede se evaluaciones escritas. <sup>24</sup>	•	•
Evento de evaluación:		
Grupo:		Fecha://
Aspecto de evaluación	✓	Comentarios
¿Su evaluación se realizó de acuerdo al plan?		
¿La evaluación fue válida? (Por ejemplo, nivel apropiado, cobertura de habilidades)		
¿La evaluación fue confiable?		
¿La evaluación fue imparcial para todos los estudiantes?		
¿Qué cambios le haría a la evaluación antes de utilizarla nuevamente?		
¿Obtuvo algún comentario de los estudiantes sobre la		

<sup>&</sup>lt;sup>24</sup> Adaptado de Clayton, B. (1995). Focussing on Assessment South Australia, p. 81. National Centre for Vocational Education Research Ltd.

	ra los estudiantes s escritas. <sup>25</sup>	Fuertemente de acuerdo						
	iste cuestionario pai ngo de evaluacione	De acuerdo						
	oceso de evaluación. E ado para evaluar un raı	En desacuerdo						
	r útil para redefinir el pr acional puede ser utiliz	Fuertemente en desacuerdo						
Calificar estrategias de evaluación Cuestionario a los estudiantes	Calificar regularmente las estrategias de evaluación puede ser útil para redefinir el proceso de evaluación. Este cuestionario para los estudiantes desarrollado por el Centro de Investigación de Educación Vocacional puede ser utilizado para evaluar un rango de evaluaciones escritas. <sup>25</sup>		Me gustaría haber tenido más información sobre esta evaluación al inicio del módulo.	Tuvimos muchas evaluaciones en este módulo. La evaluación en el módulo fue lo que esperaba.	No pude demostrar por completo mis habilidades durante la evaluación.	Mi asesor brindó comentarios favorables durante/ luego de la evaluación.	La evaluación fue una pérdida de tiempo. El resultado de mi evaluación fue justo.	

<sup>25</sup> Clayton, B. (1995: 82).

En las evaluaciones escritas desarrolladas profesionalmente es posible, y usual, implementar una variedad de controles de calidad que son más difíciles de implementar en las pruebas desarrolladas por los profesores.

# 4. PRUEBAS DESARROLLADAS PROFESIONALMENTE

Las evaluaciones escritas se basan normalmente en preguntas y cuestionarios construidos por agencias profesionales para desarrollar las pruebas. Estos instrumentos son diseñados algunas veces para el uso en el salón complementando las evaluaciones desarrolladas por el profesor, ya sea para brindar medidas más amplias del logro del estudiante o medidas más enfocadas para propósitos de diagnóstico. Las evaluaciones escritas construidas profesionalmente también son utilizadas comúnmente en postulantes seleccionados para ingresar en cursos educacionales competitivos y como base para el otorgamiento de becas. Otros usos de las pruebas escritas incluyen el monitorear los estándares educacionales (a través de encuestas internacionales. nacionales, estatales/territoriales sobre el logro del estudiante), pruebas de competencia con el propósito de graduarse y certificarse, y medición del resultado del estudiante para el uso en investigación educacional.

En las evaluaciones escritas desarrolladas profesionalmente es posible, y usual, implementar una variedad de controles de calidad que son más difíciles, y posiblemente menos necesarios, de implementar en las pruebas desarrolladas por los profesores. Este

artículo resalta un rango de consideraciones comunes en el desarrollo y uso de pruebas escritas y cuestionarios diseñados profesionalmente.

#### 4.1. Planeamiento de medición

El punto de partida en la construcción de cualquier prueba es la intención de estimar el nivel de comprensión de los estudiantes en algunos aspectos de su desarrollo, por ejemplo, estimar sus niveles de logro en algún área curricular de la escuela como álgebra, escritura narrativa o conocimientos de los trabajos gubernamentales; para estimar los niveles de comprensión en habilidades multidisciplinarias como razonamiento verbal, razonamiento numérico, comprensión mecánica, y para estimar los niveles de desarrollo de actitudes y valores como son la empatía, orientación a las metas, y actitudes hacia las matemáticas.

#### Diseñar una prueba

El trasfondo de cada prueba es la intención de estimar el nivel de comprensión de los estudiantes en solo un aspecto o dimensión de su desarrollo. Esta intención se ve reflejada en la intención de resumir el desempeño de la prueba en un solo puntaje para que todos los estudiantes puedan posicionarse con respecto a los otros en un solo orden de puntaje junto con el mismo mapa de progreso. Algunas pruebas están diseñadas para brindar estimados del nivel de comprensión del estudiante en varias áreas de logro (por ejemplo, una prueba de matemáticas puede estar diseñada para brindar puntaje en numeración, medición y espacio). Las pruebas compuestas de este tipo son en realidad tres pruebas administradas en la misma sesión. En este artículo nos referimos a una "prueba" como un conjunto de preguntas diseñadas para brindar un solo puntaje para cada estudiante en un área definida de desarrollo del estudiante.

El primer y más importante criterio al diseñar una prueba es el unir las preguntas para brindar evidencia sobre un aspecto de desarrollo que la prueba está diseñada para medir. Aquí hay dos consideraciones separadas: ¿las preguntas

de la prueba trabajan en conjunto para definir una dimensión? (el criterio de consistencia interna), y, ¿es esta la dimensión que la prueba tiene la intención de medir? (validación de construcción).

Al diseñar una prueba, los desarrolladores deben considerar si las preguntas trabajan en conjunto para brindar evidencia sobre un aspecto del desarrollo que la prueba está prevista para medir.

- ¿Las preguntas de la prueba trabajan en conjunto para definir una dimensión? (consistencia interna)
- ¿Es esta la dimensión que la prueba estaba prevista para medir? (validación de construcción).

Al seleccionar una prueba diseñada profesionalmente, los profesores necesitan considerar la relevancia de la prueba de acuerdo con su currículo.

Por ejemplo, la validación de una prueba diseñada para medir la comprensión en algún aspecto del currículo escolar (escribir y balancear ecuaciones químicas) dependerá de la medida en que los resultados frente a las tareas asignadas se pueda resumir significativamente en un solo puntaje y mapearse en una dimensión de logro, y de la medida en que la tarea brinde cobertura en esta área de aprendizaje de química.

Cuando el currículo se expresa en términos de resultados de aprendizaje previstos, la prueba debe brindar la cobertura más amplia posible de los resultados relevantes.

Para asegurar la cobertura del rango de resultados importantes en un área de aprendizaje, es común para los desarrolladores de pruebas utilizar una "cuadrícula" de especificación de la prueba. El propósito del cuadro es planear

y luego monitorear en qué medida las preguntas de la prueba brindarán cobertura de los resultados dentro del aspecto de aprendizaje.

#### Seleccionar pruebas desarrolladas profesionalmente

Al seleccionar pruebas construidas profesionalmente, los profesores necesitan considerar la relevancia de la prueba respecto del currículo prevista. ¿La prueba obtiene el rango de resultados de aprendizaje sobre lo que se ha estado enseñando? ¿Tiene la dificultad apropiada para este grupo de estudiantes? Si la prueba solo obtiene algunos de los resultados previstos, ¿hay un riesgo de que la prueba distorsione la enseñanza y la comprensión de lo que los estudiantes consideran importante? ¿Hay otras pruebas que se alinean de mejor forma con las intenciones curriculares?

# Desarrollar/seleccionar las pruebas apropiadas<sup>1</sup>

- Definir lo que medirá la prueba y para qué será utilizada. Describir la población para la que está dirigida.
- Representar de forma acertada las características, usos y limitaciones de las pruebas para su audiencia prevista.
- Explicar los conceptos de medición relevantes mientras sea necesario para aclarar el nivel de detalle que es apropiado para la audiencia prevista.

- Definir primero el propósito de la prueba y la población a ser evaluada. Luego, seleccionar una prueba para el propósito y, en base a la población, revisar la información disponible.
- Investigar las potenciales fuentes útiles de información, además de puntuación de las pruebas para corroborar la información brindada en las pruebas.
- Leer los materiales brindados por los desarrolladores de pruebas y evitar utilizar pruebas para las cuales se ha brindado información incompleta o confusa.

<sup>&</sup>lt;sup>1</sup> Extracto obtenido del *Code of Fair Testing Practices in Education* (1998), preparado por un comité conjunto de la American Educational Research Association, American Psychological Association y el National Council on Measurement in Education.

- Describir el proceso del desarrollo de la prueba. Explicar cómo el contenido y las habilidades a ser evaluadas fueron seleccionadas.
- Brindar evidencia probando que la prueba cumple con el propósito previsto.
- 6. Brindar ejemplos representativos o copias completas de preguntas de las pruebas, instrucciones, hojas de respuestas, manuales, y reporte de puntaje para calificar a los usuarios.
- 7. Indicar la naturaleza de la evidencia obtenida con respecto a la propiedad de cada prueba para los grupos de diferentes antecedentes raciales, étnicos o lingüísticos que puedan ser evaluados.
- 8. Identificar y publicar cualquier habilidad especializada que se necesite para administrar cada prueba y para interpretar el puntaje correctamente.

- **4.** Familiarizarse con el cómo y cuándo fue desarrollada la prueba y su uso.
- Leer evaluaciones independientes de la prueba y su posible medición alternativa. Buscar evidencia requerida para respaldar el pedido de los desarrolladores.
- **6.** Examinar el conjunto de pruebas por separad o ejemplo de preguntas, instrucciones, hojas de respuestas, manuales y reportes de puntajes antes de seleccionar una prueba.
- Determinar si el contenido de la prueba y la norma grupal o comparación grupal son apropiados para las personas previstas a ser evaluadas.
- 8. Seleccionar y utilizar solo las pruebas para las cuales se han establecido las habilidades necesarias y los puntajes han sido correctamente interpretados.



## 4.2. Escribir preguntas

Una vez que el aspecto de desarrollo o logro a ser medido se ha identificado y aclarado, el segundo paso en el proceso de desarrollar una prueba es decidir qué tipo de preguntas o de apuntes se utilizarán.

El formato de respuesta más común en pruebas desarrolladas profesionalmente ha sido la respuesta de opción múltiple. La razón por la que se ha expandido su uso es porque maximiza la fiabilidad de la prueba. De hecho, muchas pruebas de opción múltiple de gran escala son evaluadas por una máquina. Otros formatos de respuesta utilizados comúnmente en pruebas desarrolladas profesionalmente son ensayos y respuestas cortas: ambos son usualmente evaluadas a mano e involucran un grado de juicio por parte de los evaluadores. Con el advenimiento de tecnología que puede leer la caligrafía de los estudiantes se vuelve más común tener una máquina evaluadora de respuestas cortas e incluso de ensayos.

Cuando se diseña una pregunta, los desarrolladores enfrentan el cuestionamiento de brindar una pregunta que puede ser puntuada como correcta o incorrecta o una pregunta más abierta que permita un amplio rango de respuestas por parte de los estudiantes. Las preguntas abiertas no

El formato de respuesta más común en pruebas desarrolladas profesionalmente ha sido tradicionalmente el de respuestas de opción múltiple. La razón por la que se ha expandido el uso de este formato es porque maximiza la fiabilidad de la evaluación.

Para asegurar que las pruebas sean lo más imparciales posible, los desarrolladores deben:

- Evitar temas particulares.
- Minimizar el uso de términos y conceptos poco familiares.
- Verificar sesgos culturales o de género.
- Invitar a grupos de interés específicos para garantizar la inclusión.
- Revisar las estadísticas sobre sesgos en las preguntas.

siempre son evaluadas como correctas o incorrectas, pero se evalúan en diferentes categorías para reconocer distintos niveles de calidad en las respuestas de los estudiantes. La ventaja de las preguntas abiertas de este tipo es que pueden ser utilizadas con estudiantes en diferentes niveles de habilidad o de logro. Idealmente, estas preguntas permiten a los estudiantes con un bajo nivel de logro responder, quizá en formas poco sofisticadas, pero también retan a los estudiantes con un mayor nivel de logro. Las puntuaciones y esquemas del puntaje de "crédito parcial" se desarrollan para capturar esos diferentes niveles de respuestas.

## Minimizar sesgos

Es importante que las pruebas desarrolladas profesionalmente sean lo más imparciales posible para todos los estudiantes. Los desarrolladores intentan eliminar los sesgos en las pruebas evitando temas particulares y minimizando el uso de términos y conceptos que podrían ser poco familiares para grupos particulares de estudiantes. Se pone especial atención en las posibilidades de sesgos culturales y de género.

En programas de pruebas de gran escala, es común invitar a grupos de interés específicos para revisar las preguntas de las pruebas para buscar inclusión. Por ejemplo, los borradores de las pruebas pueden referirse a personas familiarizadas con dificultades especiales experimentadas por estudiantes que no hablan inglés como primera lengua, o a representantes de minorías para asegurar que las preguntas no tengan un sesgo cultural.

Es común en el proceso de desarrollo llevar a cabo una revisión estadística para revisar los sesgos. Esta revisión, algunas veces llamada "funcionalidad diferencial" (dif, por sus siglas en inglés) está diseñada para identificar preguntas que son inusualmente difíciles para cierto grupo de estudiantes. La identificación estadística de la funcionalidad diferencial puede ser útil para llamar la atención de preguntas que necesitan ser examinadas con más minuciosidad por contar con posibles sesgos.

#### 4.3. Estandarizar condiciones de pruebas

Es usual en pruebas desarrolladas profesionalmente especificar las condiciones bajo las cuales se toman las pruebas. Se estandarizan las condiciones de las pruebas para asegurar que los resultados puedan compararse sobre diferentes ubicaciones y tiempo. El tiempo disponible para las pruebas, los materiales que los estudiantes deben llevar (notas, libros, calculadoras), las instrucciones introductorias y las condiciones del centro de pruebas tienen que ser especificadas para intentar asegurar que algunos estudiantes no se encuentren en una injusta ventaja, o en una injusta desventaja por las condiciones bajo las cuales se administra una prueba.

# Minimizar las ventajas injustas

Es común en pruebas desarrolladas profesionalmente que los estudiantes tengan algunas preguntas de práctica antes de comenzar la evaluación. El propósito de estas prácticas es darle la oportunidad al estudiante de familiarizarse con el tipo de pregunta y con el método de registro de la respuesta. El propósito es, nuevamente, minimizar las ventajas injustas que

Para asegurar que los resultados son comparables entre diferentes lugares de evaluación y a través del tiempo, los test desarrollados profesionalmente usualmente especifican las condiciones bajo las cuales las pruebas deben ser tomadas.

puedan tener algunos estudiantes debido a su familiaridad con los procedimientos de pruebas.

Muchos programas de pruebas elaboran colecciones de preguntas de muestra, disponibles para los candidatos antes de la fecha de la prueba. Estas preguntas pueden tiene la forma de una evaluación completa. La disponibilidad de practicar con pruebas brinda al estudiante una oportunidad de practicar las respuestas bajo similares condiciones a aquellos que ya han tomado una prueba similar. De este modo, los estudiantes tienen la oportunidad de practicar.

Las respuestas a las pruebas de práctica y las soluciones (incluyendo ejemplos de ensayos de los estudiantes) brindan más oportunidad para asegurar que los estudiantes no se sientan en desventaja y puedan mostrar sus habilidades.

A menudo, se toman disposiciones especiales para asegurar que los estudiantes no se vean en desventaja por otros factores irrelevantes respecto de la prueba, como la falta de visión o la habilidad manual limitada. Muchos programas de pruebas grandes proporcionan pruebas en Braille y con letras grandes.

Se pueden proporcionar escribas para los estudiantes que no pueden registrar sus

propias respuestas; y —donde las habilidades de lectura no están siendo evaluadas— las pruebas pueden leerse a los estudiantes con habilidades de lectura limitadas.

Los test desarrollados profesionalmente pueden tener disposiciones también para candidatos que, por razones de índole religiosa, no sean capaces de realizar la prueba en la fecha indicada o deban ausentarse en parte del período de aplicación de esta. Por ejemplo, en la Prueba de Admisión de la Escuela de Medicina de Australia, un grupo de candidatos que no podía asistir a la evaluación en la fecha señalada por motivos religiosos fue supervisado desde el momento de la prueba hasta el día siguiente, cuando estuvieron en posibilidad de desarrollarla.

Para asegurar que todos los estudiantes son tratados equitativa y justamente, las pruebas profesionalmente desarrolladas usualmente

- proveen de preguntas "de práctica"
- brindan disposiciones para estudiantes con discapacidad
- cuentan con disposiciones para estudiantes con necesidades religiosas
- proveen instrucciones para minimizar la posibilidad de hacer trampa
- poseen lineamientos en caso de eventos impredecibles
- especifican mecanismos de apelación

En un intento de estandarizar las condiciones de las pruebas, de manera que se garantice que todos los estudiantes son tratados de forma equitativa y justa, los test elaborados profesionalmente pueden estar acompañados por instrucciones diseñadas para minimizar la posibilidad de hacer trampa (por ejemplo, reglas sobre los tipos de dispositivos electrónicos aceptados en el aula, organización de los asientos, arreglos para acompañar a los estudiantes cuando estos deban ausentarse del salón de examen). Los controladores

también pueden proveer lineamientos para el manejo de errores de impresión en los cuadernillos de las pruebas, reclamos acerca de errores en algún ítem, alertas debido a desastres u otros eventos impredecibles (¡un centro de evaluación australiano se enfrentó al problema de un ave volando en círculos en el interior del aula durante toda la evaluación!).

Como un intento final de asegurar que todos los estudiantes son tratados justamente gracias a las condiciones de estandarización tomadas, muchos programas de evaluación proveen mecanismos de apelación que permiten a los estudiantes presentar casos para especial consideración surgidos debido a enfermedades o estrés en el día de la prueba.

#### Administración de pruebas desarrolladas profesionalmente

Los usuarios de las pruebas desarrolladas profesionalmente tienen la responsabilidad ante los examinadores de asegurarse de que las pruebas se administran bajo las condiciones especificadas por los desarrolladores. Es posible comparar las actuaciones de los estudiantes de un centro de pruebas a otro y con las normas de prueba publicadas solo si se siguen instrucciones estandarizadas. Los usuarios también tienen la responsabilidad de garantizar que las personas no se vean en desventaja por factores irrelevantes para el contenido de la prueba.

#### 4.4. Marcado estandarizado

En las pruebas desarrolladas profesionalmente, se toman medidas para asegurar que todas las respuestas ante los estudiantes son juzgadas y registradas de la misma manera. La consistencia o calificación es más fácil o segura en pruebas de opción múltiple en las que la regla de puntuación implica la identificación de una sola alternativa correcta (la "clave"). En pruebas de este tipo, la puntuación puede automatizarse y las posibilidades de calificación injusta pueden minimizarse o eliminarse.

Sin embargo, incluso en las pruebas de elección múltiple, los controles de rutina pueden incorporarse al proceso de calificación. Los lectores de marcas ópticas pueden ser programados para identificar preguntas para las cuales las respuestas de los estudiantes son ambiguas (por ejemplo, donde dos alternativas parecen haber sido marcadas). Una vez identificadas, estas respuestas poco claras a menudo se controlan a mano y se toma una decisión sobre cómo se van a calificar. La identificación automática de los patrones atípicos de respuesta también es utilizada en algunos programas de pruebas. Un patrón atípico de respuestas puede surgir porque una hoja de respuestas fue desalineada al ser leída o porque un estudiante volteó dos páginas en vez de una y, por lo tanto, no vio algunas preguntas.

# Asegurar la coherencia de la calificación

La calificación de las pruebas de respuestas abiertas casi siempre implica un grado de juicio por parte de los evaluadores. Para ayudar en la evaluación, se proporcionan a los evaluadores criterios y ejemplos de respuestas aceptables e inaceptables a cada pregunta. Cuando se dispone de calificación parcial para respuestas incompletas o respuestas parcialmente correctas a una pregunta, se proporcionan criterios y ejemplos para ilustrar cada una de las puntuaciones posibles

Para asegurar que las respuestas de todos los estudiantes son evaluadas de la misma manera, las pruebas profesionalmente desarrolladas proveen

- criterios de evaluación
- ejemplos de respuestas aceptables e inaceptables para preguntas de respuesta abierta, y
- ejemplos de respuestas de los estudiantes equivalentes a cada punto en una escala de valor

en la pregunta. La identificación de niveles variables de respuesta a un ítem puede ayudar a ilustrar niveles a lo largo de un mapa de logros crecientes (ver ejemplo en la página 106).

Las pruebas de escritura desarrolladas profesionalmente proporcionan ejemplos de los criterios que deben usarse para evaluar los escritos de los estudiantes (vea la página 107), así como ejemplos anotados de escritura de los estudiantes para ilustrar la aplicación de esos criterios. Las evaluaciones de la escritura de los estudiantes implican siempre un juicio de equilibrio de la calidad tomando en cuenta la pieza o trabajo como un todo. Por esta razón, los exámenes de escritura desarrollados profesionalmente suelen proporcionar una serie de ejemplos de escritura de los estudiantes en cada momento en una escala de rastreo para ilustrar la variedad de maneras en que los estudiantes pueden lograr la misma puntuación.

En los programas de pruebas a gran escala, es usual establecer procedimientos para monitorear la consistencia de los evaluadores. Se les pide a los diferentes evaluadores que evalúen el mismo trabajo del estudiante (por ejemplo, un ensayo) y se comparen las evaluaciones realizadas con ese trabajo. A través de comparaciones de este tipo es posible identificar evaluadores "discrepantes". Los evaluadores discrepantes a veces son más indulgentes (tienden a dar puntuaciones más altas) que otros evaluadores, a veces son más rigurosos (tienden a dar puntuaciones más bajas) y a veces tienden a hacer menos uso del rango de puntuación disponible (puntúa todo el trabajo de forma muy similar). Cuando se observan grandes discrepancias, el trabajo de los estudiantes es comúnmente observado por un tercer evaluador.

Incluso con el uso de procedimientos para identificar y manejar las discrepancias entre los evaluadores, las diferencias pequeñas, pero consistentes, en la rigurosidad/indulgencia del evaluador pueden permanecer sin ser detectadas. En un intento de asegurar que todos los estudiantes son tratados de manera justa, a veces se hacen ajustes estadísticos a los resultados de los estudiantes para eliminar los efectos del evaluador (por ejemplo, para aumentar

ligeramente el resultado final de los estudiantes que han sido calificados por dos evaluadores rigurosos).

## Uso de guías de evaluación desarrolladas profesionalmente

Los usuarios de pruebas desarrolladas profesionalmente tienen la responsabilidad de probar las evaluaciones para asegurar que las guías y criterios de evaluación proporcionados se usen de acuerdo con las directrices de los desarrolladores de pruebas. Solo si se siguen estas pautas es posible comparar los resultados de los estudiantes de un centro de pruebas a otro y con las normas de prueba publicadas.



#### Estandarizar la evaluación

#### Asegurar una calificación consistente

La siguiente pregunta abierta de una prueba de lectura DART para estudiantes de primaria superior brinda una puntuación de 1 (crédito parcial) para respuestas que brindan solo ejemplos, pero que no generalicen. Una puntuación de 2 (crédito completo) es brindado para generalizaciones mostrando una comprensión más profunda del término "suelos adecuados". La figura muestra que un puntaje de 1 en la pregunta 7 ilustra el nivel 3 de logro de lectura; un puntaje de 2 ilustra el nivel 5 de logro de lectura.

#### 7. En algunas regiones arenosas se dejarán rastros en el suelo adecuado. ¿Qué es un suelo adecuado?

7.2 Generalización

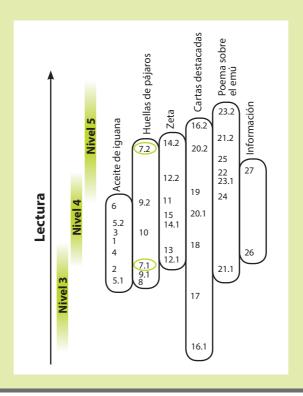
Ej.: "Los suelos son tan suaves que ellos dejan un rastro claro"."Los suelos que permiten que se realice una clara impresión".

7.1 Solo ejemplo

Ej.: "Arena"

7.0 Respuesta incorrecta

Ej.: "Marcas, suelos"



<sup>&</sup>lt;sup>2</sup> Forster, M.; Mendelovits, J. y Masters, G. N. (1995). *Developmental Assessment Resource for Teachers DART Upper Primary*. Melbourne: Australian Council for Educational Research.

# Guía de evaluación para evaluar las características del lenguaje en escritura del estudiante

#### Fuente: Encuesta de Alfabetización Nacional de Lenguaje<sup>3</sup>

	ente: Encuesta de Alfabetización Nacional de Lenguaje
5	Revisar que la escritura sea consistente en contenido y estilo.  Experimenta con oraciones reorganizadas  Controla las estructuras gramaticales y puntuación en oraciones complejas.  Organiza la escritura de forma coherente y apropiada al contexto (por ejemplo, párrafo para la narración títulos y subtítulos para un texto de información).  Utiliza vocabulario preciso y efectivo.  Aproxima la ortografía de palabras particularmente difíciles utilizando patrones.
4	<ul> <li>Empieza a adoptar convenciones organizacionales de formatos estructurados (por ejemplo, oración introductoria general hacia un reporte).</li> <li>Contiene una variedad de formas de oraciones (por ejemplo, oraciones simples y complejas).</li> <li>Utiliza puntuación apropiada la mayoría del tiempo.</li> <li>La escritura es clara al inicio y final y tiene posibles divisiones de párrafos.</li> <li>Utiliza vocabulario apropiado la mayor parte del tiempo.</li> <li>Deletrea la mayoría de las palabras correctamente.</li> </ul>
3	Muestra alguna evidencia de planeamiento, revisión y corrección en su propia escritura.  Controla estructura de oraciones simples e intenta estructuras más completas. Intenta una variedad de inicio de oraciones. Intenta estructurar piezas (por ejemplo, noción de inicio y final).  Deletrea muchas palabras comunes de forma correcta.  Escribe legiblemente.
2	Utiliza oraciones simples. Utiliza estructura de oraciones repetitivas. Utiliza conjunciones simples (por ejemplo, "y" y "pero"). Controla la misma puntuación común algunas veces o todo el tiempo (por ejemplo, letras mayúsculas, puntos finales). Deletrea palabras de uso frecuente de forma correcta la mayor parte del tiempo. Escribe de manera que pueda ser generalmente interpretado por los demás.
1	Utiliza algunas convenciones básicas.  (por ejemplo, escribe de izquierda a derecha, coloca espacios entre las palabras).  Contiene algunas palabras conocidas, o palabras representadas por su letra inicial.  Utiliza la inicial correcta y otros sonidos.  Puede leer mientras escribe.

<sup>&</sup>lt;sup>3</sup> Masters, G. N., Forster, M. (1997). *Mapping Literacy Achievement,* Canberra: Department of Employment, Education, Training and Youth Affairs.

#### 4.5. Interpretar el resultado de las pruebas

En evaluaciones desarrolladas profesionalmente, se toman análisis estadísticos para establecer en qué medida las preguntas ayudan a que la evaluación funcione en conjunto para respaldar la intención de resumir cada respuesta del estudiante en un solo puntaje. Estos métodos estadísticos se conocen como modelo de "respuesta a pregunta".

La calibración de elementos de prueba desarrollados profesionalmente a lo largo de un mapa de progreso permite a los usuarios de la prueba interpretar las puntuaciones de los exámenes de los estudiantes en términos de los tipos de conocimientos, habilidades y entendimientos típicamente asociados con cada puntaje.

Un resultado de aplicar este modelo de respuesta a pregunta a un grupo de preguntas es que las preguntas que conforman una prueba son "calibradas" de acuerdo con su dificultad. Un ejemplo del resultado de calibrar un juego de preguntas se muestra en la página 106. La imagen en la página 106 muestra los resultados de utilizar las respuestas en un gran grupo de estudiantes de la escuela primaria para calibrar las preguntas en una prueba de lectura. Las preguntas que los estudiantes responden correctamente con más frecuencia se muestran hacia la parte baja de la figura. La pregunta más fácil para los estudiantes obtuvo 16.1 (en otras palabras, una puntuación de 1 en la pregunta 16). Las preguntas que fueron menos respondidas correctamente se muestran en la parte superior de la figura. La pregunta más difícil fue 23.2 (en otras palabras, puntaje 2 en la pregunta 23).

La calibración estadística de preguntas de esta manera proporciona un "mapa" de preguntas cada vez más difíciles. Cuando los elementos de dos, tres o más pruebas se calibran juntos, un mayor número de estas se encuentra en cada

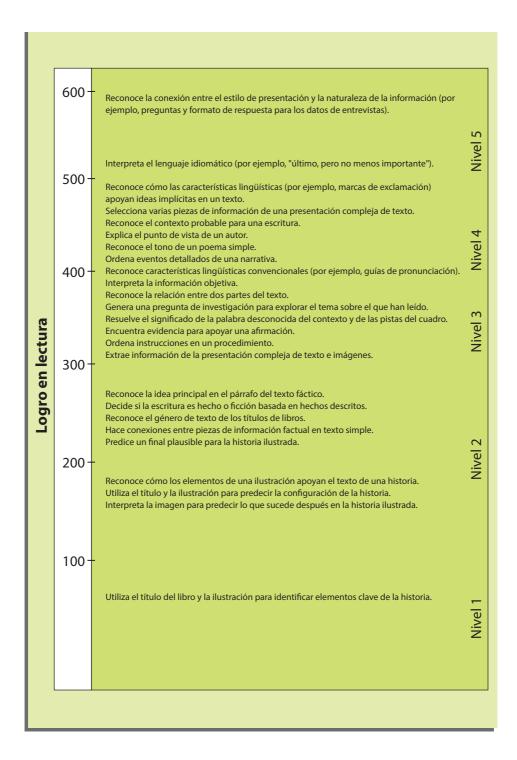
<sup>&</sup>lt;sup>4</sup> Masters, G. N., Forster, M. (1997: 108).

posición a lo largo del mapa. Al examinar y describir lo que los estudiantes tienen que hacer para responder a estos puntos, se puede desarrollar una descripción cada vez más rica de las posiciones a lo largo del mapa.

#### Usando mapas de progreso

La imagen de la página 110 muestra descripciones de preguntas calibradas a lo largo de un mapa de desarrollo en lectura. En evaluación del desarrollo, a este tipo de mapa se le llama "mapa de progreso". La pregunta más fácil en esta evaluación de lectura está calibrada cerca de la base de la figura. Esta pregunta requiere de niños pequeños predecir el contenido de un libro a partir del título y la ilustración de la carátula. La pregunta más difícil del test está calibrada cerca de la parte superior de la figura; y requiere de los niños evaluar la propiedad de un texto de acuerdo con su objetivo (a saber, la propiedad de un formato de pregunta y respuesta para la grabación de datos de una entrevista). Leyendo hacia arriba la página, se puede "ver" el nivel creciente en habilidad de lectura que se requiere para responder las preguntas cada vez más difíciles sobre el texto. El desempeño de los estudiantes en esta evaluación se expresa en una escala que va desde cerca de 100 hasta cerca de 600. El texto a lo largo de la escala muestra el tipo de tareas de lectura que los estudiantes de cada nivel deberían dominar.

La calibración de preguntas a lo largo de un mapa de progreso de pruebas desarrolladas profesionalmente permite a los usuarios de estas pruebas interpretar los puntajes obtenidos por los estudiantes en términos de tipos de conocimiento, habilidades y comprensiones asociados típicamente con cada puntaje. Considere, por ejemplo, a un estudiante que obtiene un puntaje de 350 en esta escala. Debajo del puntaje 350 se encuentra la descripción "Encuentra evidencia para apoyar una afirmación". Este es un ejemplo del tipo de habilidades que un estudiante en este nivel de competencia de lectura tiene una oportunidad razonable de demostrar. Habilidades por debajo de este nivel son más probables de ser demostradas; mientras que las que se encuentran sobre dicho nivel tienen menos probabilidades de mostrarse.



#### Uso de las normas de pruebas

En las pruebas desarrolladas profesionalmente, es común proporcionar a los usuarios información sobre los resultados de alguna población relevante de estudiantes en una prueba. Estas "normas" permiten que los resultados de las pruebas de los estudiantes sean interpretados en términos de las actuaciones de otros estudiantes de la misma edad o en el mismo grado. Por ejemplo, una prueba de matemáticas de 6.<sup>to</sup> grado puede mostrar cómo una muestra nacional de estudiantes de sexto grado se desempeñó en esa prueba.

Usualmente, las normas de prueba se informan como "percentiles". Un estudiante que obtiene puntajes en el percentil 80, puntúa tan bien o mejor que el 80% de los estudiantes del grupo comprendido en la norma. Los manuales para las pruebas estandarizadas normalmente contienen tablas de normas que muestran el porcentaje de estudiantes que califican en o por debajo de cada puntaje posible de la prueba.

Estos percentiles pueden ser mostrados para toda la población estudiantil, así como para subgrupos relevantes (por ejemplo, quinto grado, sexto grado, niños, niñas).

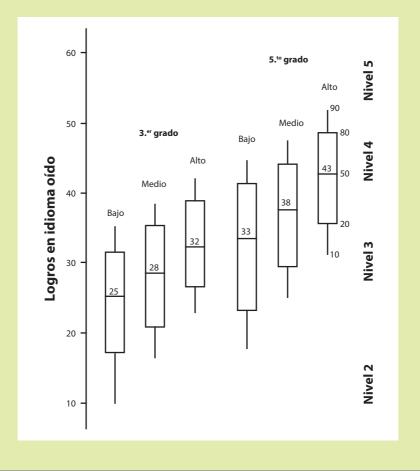
Además de proporcionar tablas de normas que permitan a los profesores comparar los resultados de las pruebas con los resultados típicos de la edad o el grado, los manuales de prueba a veces proporcionan visualizaciones gráficas de información normativa. Un formato para mostrar las normas de prueba es un diagrama de "caja y bigote" (vea la página 112) en el que se marcan los puntos percentiles (por ejemplo, percentiles 10, 25, 50, 75 y 95).

Al comparar los resultados de las pruebas de los estudiantes con las normas de prueba publicadas, es importante que los usuarios de la prueba consideren la relevancia de la población normativa y aseguren que la prueba se administre bajo las condiciones y procedimientos recomendados por el desarrollador de la prueba.

# Interpretación de los resultados de las pruebas

# Uso de las normas de pruebas

Los maestros que utilizan las pruebas de visualización estandarizadas construidas para la Encuesta de Nacional de Alfabetización en Inglés a Nivel Escolar son capaces de comparar los resultados de las pruebas de comprensión de un idioma oído con los resultados de una muestra nacional de estudiantes en estas pruebas. El diagrama de cajas y bigotes muestra el desempeño de los estudiantes de los 3.er y 5.to años en grupos socioeconómicos bajos, medios y altos. Por ejemplo, un estudiante del 3.er año con un rendimiento de prueba de 34 en la escala de logro de audición se ha desempeñado por encima del percentil 50 para los estudiantes del 3.er año en el grupo socioeconómico más alto y también por encima del percentil 50 para estudiantes de 5.to en el nivel socioeconómico más bajo.



<sup>&</sup>lt;sup>5</sup> Masters, G. N., Forster, M. (1997: 182).

## 5. DISEÑANDO LA EVALUACIÓN ESCRITA

En la evaluación del desarrollo, los maestros monitorean el progreso del estudiante comparando con un mapa de desarrollo de habilidades, conocimientos, comprensión, actitudes o valores. La evaluación escrita es un método que se puede usar para recopilar evidencia del logro del estudiante. Diferentes tipos de formatos de respuesta pueden proporcionar diferentes tipos de evidencia que los profesores pueden usar para estimar las ubicaciones de los estudiantes en un mapa de progreso.

En este acápite se enumeran las cuestiones que los profesores deben tener en cuenta al diseñar evaluaciones escritas. Estos temas incluyen preguntas sobre el propósito de la evaluación, métodos para juzgar y registrar el desempeño de los estudiantes, y maneras de estimar y reportar los niveles de logro de los estudiantes. Al final del artículo, se incluye un resumen a modo de "lista de verificación" del proceso de diseño de la evaluación.



#### 5.1. Planificación de evaluaciones escritas

¿Cuál es el propósito de la evaluación?

Por ejemplo, ¿desea diagnosticar las fortalezas y debilidades de los individuos en un área particular de aprendizaje, informar el proceso de enseñanza a nivel de toda la clase o inferir los niveles de logro de los estudiantes para reportar a los padres?

La forma en que responda a esta pregunta determinará el rango de metas o resultados del plan de estudios que usted dirige (consulte las páginas 12-18 y 49-53).

¿Para qué objetivos o resultados del plan de estudios reunirá evidencia?

¿Es pertinente la evidencia que está reuniendo? Es decir, ¿se centra en resultados u objetivos de instrucción explícitos?



Dependiendo del propósito de su evaluación, ¿la evidencia que está reuniendo aborda la gama completa de resultados (vea las páginas 12-18 y 49-53)?

¿Cuál formato de respuesta es el más adecuado para los resultados que se están abordando?

Por ejemplo, si el resultado requiere que los estudiantes demuestren conocimiento factual, entonces se puede usar un formato de opción múltiple. Si los resultados requieren que los estudiantes estructuren un argumento lógico, entonces se puede usar un formato de ensayo (Vea las páginas 52-53 y 59-70).

## ¿Son justas las pruebas planificadas?

¿Las pruebas tienen un enfoque claro? ¿Las tareas incluyen a estudiantes de diferentes géneros y culturas? ¿El lenguaje utilizado para presentar las tareas es accesible? ¿Ha tenido tiempo de revisar las tareas antes de administrarlas (vea las páginas 52-59)?



## 5.2. Calificación y registro

¿Cómo calificará y registrará el trabajo de los estudiantes?

¿Cuál método para calificar y registrar el trabajo del estudiante es el más apropiado para su propósito de evaluación y formato de respuesta? ¿Utilizará guías de puntuación, categorías de diagnóstico, clasificaciones analíticas o calificaciones holísticas (consulte las páginas 72-80)?

¿Qué criterios utilizará?

¿Sus criterios de evaluación proporcionan evidencia que es relevante? Es decir, ¿se refieren a los resultados que se están evaluando? ¿Son claras, justas y utilizables? ¿Tiene ejemplos de trabajo del alumno (ver páginas 72-87)?

¿Cuán grande es la necesidad de comparabilidad del puntaje?



## ¿Quién hará la evaluación?

¿Las tareas serán evaluadas por el profesor, por los compañeros o por el mismo estudiante (vea la página 80)?

## 5.3. Resumir el trabajo del estudiante

¿Cómo calculará los niveles de logro de los estudiantes en un mapa de progreso?

¿De qué evidencia estimará el nivel de logro de los estudiantes? ¿Utilizará calificaciones holísticas o evaluaciones analíticas del trabajo de los estudiantes o criterios de calificación?

¿Ha comprobado que sus estimaciones son válidas? (es decir, ¿las estimaciones se basan en pruebas pertinentes?)



¿Ha comprobado que las estimaciones son fiables? (es decir, ¿se basan en pruebas suficientes?)

¿Ha comprobado que las estimaciones son objetivas? Es decir, no se ve afectada por la elección de la tarea o la elección del evaluador (véanse páginas 85-87).

# 5.4. Informando el trabajo del estudiante

¿Cómo reportará el logro estudiantil?

¿Interpretará los niveles de logro descriptivamente? ¿Mostrará los logros de forma gráfica (consulte las páginas 88-89)?



# 5.5. Un resumen del proceso de diseño de una evaluación escrita

Etapa de diseño	Estrategias de diseño
Decidir el propósito de la evaluación.	<ul> <li>Describa el propósito de la evaluación.</li> <li>Revise estas descripciones en relación con objetivos y resultados importantes del plan de estudios.</li> </ul>
Decidir las metas del plan de estudios o los resultados a obtenerse.	Enumere las metas o los resultados.
Decidir sobre el formato de respuesta.	Compruebe que el formato es adecuado para los resultados que se tratan.
Revisión antes de la administración.	<ul> <li>Compruebe la imparcialidad (incluyendo claridad, inclusión, lenguaje accesible).</li> </ul>
Decidir sobre un procedimiento para calificar y registrar pruebas.	<ul> <li>Decida quién evaluará (autoevaluación, compañero, maestro).</li> <li>Desarrolle guías de puntaje (criterios de puntuación o escalas de calificación).</li> <li>Revise estos resultados en relación con los que se están evaluando.</li> <li>Revise la claridad y la utilidad.</li> </ul>
Decidir sobre un procedimiento para estimar los niveles de logro en un mapa de progreso.	<ul> <li>Describa el procedimiento para estimar los niveles de logro.</li> <li>Revise estas descripciones en relación con la tarea, el propósito y la audiencia.</li> </ul>
Decidir sobre un procedimiento para reportar los niveles de logro.	<ul> <li>Describa el procedimiento para reportar los niveles de logro.</li> <li>Revise estas descripciones en relación con la tarea, el propósito y la audiencia.</li> </ul>

# Evaluación escrita Recursos de evaluación del Consejo Australiano para la Investigación Educativa 8

Durante la elaboración de estándares de aprendizaie en aue estuvo abocado el SINEACE entre los años 2009-2015, se conoció la experiencia australiana de evaluación de los aprendizajes realizada por el Consejo Australiano para la Investigación Educativa (ACER). El SINEACE suscribió un convenio con el ACER para traducir los folletos que comprenden el Kit de Recursos de Evaluación, que fueron un valioso material durante el proceso; esto permite poner al alcance de los lectores de habla castellana el octavo v último número de la serie. el cual rescata la importancia de la evaluación escrita como herramienta para recolectar evidencia que permita evaluar el desarrollo de los estudiantes.

La primera parte de la publicación describe la variedad de evaluaciones escritas y para qué propósitos se utilizan; la segunda parte explora el diseño de estas; la tercera ahonda en las consideraciones a tomar en cuenta en el desarrollo y uso de pruebas diseñadas profesionalmente. Finalmente, se brinda una revisión completa del proceso de diseño de las pruebas escritas.

# SERIE DOCUMENTOS TÉCNICOS





