



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

**FACULTAD DE PSICOLOGÍA**

**División de Estudios Profesionales**

**Propuesta para mejorar los reportes de resultados  
de pruebas estandarizadas de aprovechamiento**

Edward Ortega Serrano

Tesina para obtener el título de Licenciado en Psicología

Director

Lic. Fernando García Cortés

Sinodales

Lic. Irma Graciela Castañeda Rodríguez

Dra. Gilda Teresa Rojas Fernández

Mtro. Miguel Herrera Ortiz

Mtra. Patricia Bermúdez Lozano



México, D.F

Mayo de 2014



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## Agradecimientos

“La memoria no es lo que recordamos, sino lo que nos recuerda.  
La memoria es un presente que nunca deja de pasar”.

Octavio Paz

Cuando se inicia un proyecto, las esperanzas y expectativas se comparten con la gente que lo acompaña. Y en algunas ocasiones el compartir también se vuelve un proyecto por sí mismo. Este proyecto fue muy largo, abandonado muchas veces y desplazado; pero siempre presente de forma sutil, cotidiana y alentado a finalizarlo por mucha gente que me ha rodeado.

Finalizarlo es cerrar un ciclo enorme, distante.

Y tanta, tanta gente compartió esto, que en este trabajo están ellos. Los vivos y no vivos, los que dejaron de ser y los que ahora son más.

Y yo no soy sin ellos.

Al Licenciado Fernando García, por su confianza y gran calidad profesional, que se convirtió en ejemplo personal.

A los que siempre motivan, acompañan y hacen que muchos milagros sucedan en la vida: Lorena y Harvey.

A quien con paciencia espera que las cosas pasen; confiando callada y sabiamente: mi madre.

A quien no está físicamente, pero es un eterno presente en la memoria: mi padre.

Y a tantos que ayudaron de mil formas.

Gracias

Edward Ortega Serrano

<b>INTRODUCCIÓN</b> .....	1
Planteamiento del problema .....	2
Justificación .....	3
Objetivos generales .....	6
Objetivos particulares .....	6
<b>CAPÍTULO I. EL ANDAMIAJE TEÓRICO</b> .....	8
¿Qué es una prueba psicológica? .....	9
¿Aplicar una prueba es evaluar? .....	11
Tipos de pruebas psicológicas .....	13
Características de las pruebas de aprovechamiento .....	14
Usos de las pruebas de aprovechamiento .....	16
Indicadores de calidad de una prueba: la confiabilidad y la validez .....	17
La validez .....	18
La confiabilidad .....	22
¿Qué es más importante para determinar la calidad de una prueba, la validez o la confiabilidad? .....	24
Documentos que guían el diseño y construcción de las pruebas estandarizadas.....	25
Pasos para elaborar una prueba .....	27
Los reportes de resultados de la prueba .....	29
Lineamientos para elaborar reportes .....	29

Literatura sobre reportes de resultados en pruebas .....	30
<b>CAPITULO II. LAS PRUEBAS ESTANDARIZADAS DE APROVECHAMIENTO EN MÉXICO .....</b>	<b>32</b>
Las instituciones de evaluación de la educación en México .....	33
Participación de México en evaluaciones internacionales .....	36
Pruebas estandarizadas de gran escala en México .....	38
La difusión de resultados de las pruebas estandarizadas en México .....	39
<b>CAPITULO III. LOS REPORTES DE RESULTADOS DE LAS PRUEBAS ESTANDARIZADAS .....</b>	<b>41</b>
Estándares, normas y recomendaciones sobre reportes de resultados .....	42
Diseño de los reportes de calificación .....	46
Componentes de un reporte de resultados .....	48
Aspectos que hacen poco efectivo a un reporte de resultados .....	49
Elementos de un buen reporte de resultados acorde con investigaciones .....	50
Cómo generar un reporte de resultados de una prueba .....	55
Estructurando información en un reporte .....	56
Metodología .....	57
El tipo de investigación .....	57
El enfoque de la investigación .....	58
La población o universo de estudio .....	58
Unidad de análisis .....	58

Técnicas de recolección de información .....	58
Fuentes de información .....	59
<b>CAPÍTULO IV. LA PROPUESTA DE TRABAJO .....</b>	<b>60</b>
Primer paso. Elaborar una ficha de componentes de un reporte de resultados de una prueba estandarizada .....	62
Segundo paso. Determinar los estándares sobre la entrega de resultados y el uso de una prueba estandarizada .....	64
Tercer paso. Reconocer las consideraciones generales y particulares de cada elemento del reporte de una prueba estandarizada acorde con los expertos .....	66
Cuarto paso. Recomendaciones para verificar un reporte de resultados de una prueba estandarizada y determinar las mejoras .....	71
<b>CAPÍTULO V. APLICACIÓN DE LA PROPUESTA A UNA PRUEBA ESTANDARIZADA .....</b>	<b>74</b>
La prueba Enlace EMS .....	75
El antecedente: porqué se decide crear una prueba de gran escala a nivel nacional	75
Estatus actual de la prueba .....	80
Reconocimiento de los componentes de un reporte de resultados de la prueba Enlace EMS .....	80
Estándares sobre la entrega de resultados y el uso de la prueba Enlace EMS .....	82
Reconocimiento de consideraciones particulares de cada elemento del reporte de una prueba estandarizada .....	85
Recomendaciones para elaborar el reporte de resultados de la prueba Enlace EMS .....	86

Reportes individuales del nivel de dominio que el alumno obtuvo en la prueba Enlace EMS .....	91
Reportes individuales con los reactivos por cada campo disciplinario .....	97
Reportes para los profesores .....	105
Reportes para autoridades educativas .....	112
<b>CAPITULO VI. CONCLUSIONES Y RECOMENDACIONES</b> .....	120
Conclusiones .....	121
Limitaciones de la propuesta y sugerencias para trabajos futuros .....	124
<b>REFERENCIAS</b> .....	127
<b>ANEXOS</b> .....	140

## INTRODUCCIÓN

En el ámbito educativo, en particular en el proceso de enseñanza y de aprendizaje la evaluación es constante y hasta cotidiana, si entendemos por evaluación como la formulación de un juicio que se obtiene de manera organizada, se produce cuando un profesor obtiene y proporciona información sobre el desempeño de los alumnos en una o varias asignaturas o cuando se proporcionan datos sobre el desempeño de los alumnos a nivel local, regional o nacional. En cualquier caso hay una evaluación abordada de manera diferenciada.

Cuando hay que obtener información a gran escala lo usual es el uso de las pruebas estandarizadas, que son instrumentos que miden una muestra representativa de uno o varios dominios de manera rápida, homogénea y objetiva. Pero elaborar una prueba estandarizada es un proceso costoso y complejo en términos técnicos, involucra a mucha gente y meses de elaboración para finalizar como un instrumento aplicable a los usuarios y que sirva para brindar información.

Brindar información es el propósito de toda prueba estandarizada, puede o no ser de interés público pero la brinda, y generalmente a través de los reportes de resultados para diferentes audiencias, pero ¿cómo asegurar que se proporciona la información adecuada? ¿Cómo saber si la manera en la que se proporciona la información es la mejor? ¿Hay criterios que nos permitan conocer si cualquier prueba cumple con sus objetivos? ¿Cómo saber si una prueba elabora buenos reportes? y tan importante ¿una prueba que reporte resultados puede mejorar la manera en la que los ofrece?

Estas son las cuestiones que se abordan en este trabajo que tiene como propósito hacer una propuesta para verificar si un reporte de una prueba estandarizada tiene los contenidos correctos, esperados acorde con estándares internacionales, acorde con criterios de expertos y sobre todo que permitan ser útiles para las poblaciones para las que fueron diseñadas.

## **Planteamiento del problema**

En México la evolución de las pruebas estandarizadas en ámbitos educativos ha sido discontinua porque no han existido instrumentos con una sólida temporalidad, si bien el país tiene instancias creadas profesionalmente para realizar evaluaciones no sujetas a presupuestos de orden federal o instancias gubernamentales que hacen instrumentos de evaluación; no hay agencias, marcos legales o estándares nacionales o estatales que determinen el qué o cómo se deben diseñar instrumentos de evaluación como las pruebas estandarizadas. Si bien estos referentes legales no son la garantía para una buena prueba el no tenerlos puede provocar que quien las elabore lo haga atendiendo a contextos inmediatos o metodológicamente poco rigurosos o discontinuos, al no estar sujetos a reglas claras o entrega de resultados técnicos.

Abordar la elaboración de estándares o reglas para la creación de una prueba es un tema complejo que involucra a diversos actores de distintos ámbitos de la sociedad y que debe ser abordado por expertos en disciplinas como sociología, psicología, psicometría, matemáticas, actuaría, pedagogía o ciencias de la comunicación. Estos componentes no deben limitar el hecho de que las pruebas estandarizadas que ya existen deben ser analizadas, revisadas o descritas en función de referentes técnicos que pueden estar al alcance de un público poco especializado en el tema; las pruebas proporcionan información a los usuarios y ésta debe ser elaborada cuidadosa y profesionalmente. Cómo presentar o diseñar dicha información de forma organizada no está presente en la literatura habitual de diseño de pruebas; en apariencia se da por hecho que el reporte de resultados obedece naturalmente a diferentes fases previas y que no ha requerido tanta atención como otros componentes documentados abundantemente en la literatura.

Ante estas razones, el problema de la investigación gira en torno a tres elementos:

-La ausencia de criterios o estándares de distintas fuentes agrupados en un conjunto que dicten la manera de hacer reportes de resultados en pruebas estandarizadas.

-La ausencia de criterios o recomendaciones a los reportes de resultados en una prueba estandarizada nacional.

-La necesidad de crear reportes de resultados útiles y que generen conocimiento para diferentes usuarios.

## **Justificación**

Uno de los elementos primordiales para el adecuado desarrollo de un país es su sistema educativo. Este transmite conocimientos, saberes y valores para formar personas capaces de desempeñarse de manera eficiente en su ámbito cotidiano y formar individuos que reconozcan la forma en la que se estructura el mundo; permite que los ciudadanos desarrollen fortalezas y habilidades útiles en los diferentes ámbitos de la vida.

El sistema educativo entendido como un sistema es un concepto abstracto; es un conjunto de elementos que se pueden abordar desde diferentes enfoques o perspectivas de interés o funcionalidad como la infraestructura, los agentes involucrados de manera directa o indirecta, los modelos en los que se basa el proceso de enseñanza o aprendizajes o la diversidad o la profundidad de los currículos. La manera en la que se abordan es relevante en función de la información que se quiera obtener, es decir, dónde se indaga determina lo que se encuentra.

Saber indagar en un ámbito tan complejo es importante porque los datos responden a lo inquirido. La forma en la que se obtiene información deber ser

coherente con los propósitos y los instrumentos que ayudan a recabar la información deben ser considerados cuidadosamente en función de diversos componentes como su utilidad, aplicabilidad, costo, insumos adicionales, tiempo de obtención de resultados, etcétera. Los instrumentos para coleccionar información son usuales en el ámbito educativo; son cotidianos, constantes y permean actividades que van desde circunstancias personalísimas como la evaluación de un profesor para saber el grado de aprendizaje de un alumno al finalizar un curso hasta aquellas que implican la evaluación del aprendizaje de las competencias en los alumnos con el fin de compararlos con sus pares de distintos países. En cualquier circunstancia es relevante poder obtener información por medio de un instrumento de evaluación.

Los instrumentos de evaluación que se aplican en el ámbito educativo de forma masiva son herramientas costosas; el diseño, la estandarización y sobre todo la aplicación o calificación requieren insumos materiales especializados con personal capacitado en tareas muy específicas. Su elaboración no puede ser superflua ni considerada sencilla o simple porque independientemente de lo que ofrezca (si es información con fines diagnósticos para tomar decisiones de alto o de bajo impacto) debe ser concebida con gran calidad y profesionalismo en todas sus fases. En países habituados a la elaboración de estos instrumentos es común encontrar documentos, decretos, normas o leyes que dictaminan el diseño o sus componentes; donde muchos aspectos sobre las pruebas son detallados.

En países con poca cultura respecto al uso o diseño de pruebas estandarizadas la reglamentación legal respecto a estas es poca, ambigua o con desconocimiento técnico por parte de sus elaboradores; y si bien no es garantía de un buen instrumento, ayuda al proponer condiciones mínimas y estandarizadas de elaboración. Pero algo se tiene por cierto: las pruebas estandarizadas se elaboran con o sin buenos marcos legales, son elaboradas para proporcionar información aun sin criterios claros sobre su diseño o elaboración. Se puede creer que ninguna prueba estandarizada debe elaborarse sin reglas claras, pero ¿es así? ¿Cómo

podemos saber si una prueba se elaboró de forma adecuada, siguiendo reglas o procedimientos técnicos pertinentes?

Hay diferentes procedimientos que permiten verificar si una prueba estandarizada se elaboró acorde con adecuados criterios técnicos en distintas categorías y en diversos niveles de profundidad o detalle técnico; lo más frecuente es hablar del grado de confiabilidad o de evidencias sobre si los puntajes que se obtienen en la prueba apoyan a las interpretaciones que se hacen sobre ésta (validez) con muchas formas de obtener datos, bien documentadas y con un buen nivel de descripción por parte de los expertos en psicometría. Pero hay un componente de las pruebas que recibe poca atención: los reportes de resultados.

Los reportes de resultados son el componente más evidente, tangible y probablemente el más socializado de todo el proceso de construcción de una prueba: el constructo que se aborda, los ítems elaborados, la escala de calificación, los niveles o criterios en los que se reporta el desempeño, las consideraciones sobre lo que la prueba ofrece o no, los documentos complementarios y recomendaciones a los usuarios son elementos que se esperan de un reporte de resultados; pero en comparación con otros temas técnicos hay poca información sobre lo que debe contener un buen reporte de resultados, más aun, no hay muchos procedimientos que permitan verificar si los reportes cumplen con los objetivos de calidad y hay muy pocos para un público no especializado. ¿Con qué criterios se puede juzgar si una prueba cumple con los estándares adecuados? ¿Qué se espera de un buen reporte de resultados? ¿Se asegura que la información es útil para el usuario?

No hay agencias nacionales o institutos que normen de manera detallada lo que se espera de un reporte de resultados, incluso el Instituto Nacional para la Evaluación de la Educación no lo ha definido y esto provoca que no haya información sobre qué esperar o exigir, sobre referentes que ayudan a determinar

si la información es útil o suficiente o sobre si la información ya presentada es la adecuada.

Nuestra propuesta es una respuesta para mejorar la construcción de pruebas estandarizadas al crear un modelo que permita verificar que los reportes a los usuarios cumplen con estándares de calidad, recomendaciones de expertos e información que ayude a los usuarios a generar conocimiento.

### **Objetivos generales**

Realizar un estudio de los estándares internacionales, investigaciones y recomendaciones de expertos sobre el diseño de reportes de resultados en una prueba estandarizada para elaborar una propuesta de verificación de reportes.

Elaborar una propuesta para verificar si los reportes de resultados cumplen con lo que se espera de ellos y que permita, basada en los diagnósticos o resultados, elaborar mejores reportes.

### **Objetivos particulares**

Revisar y clasificar los estándares internacionales sobre la elaboración de reportes de resultados en pruebas estandarizadas.

Revisar las investigaciones referidas al análisis de reportes de resultados y agrupar las conclusiones en categorías.

Revisar las recomendaciones de expertos en el tema de elaboración de reportes y agrupar las conclusiones en categorías.

Conjuntar todo lo referido a reportes de resultados y elaborar una propuesta de verificación de éstos.

Elaborar una propuesta de verificación que permita generar diagnósticos y recomendaciones.

Aplicar la propuesta a una prueba estandarizada que proporcione datos de sus reportes: la prueba Enlace de nivel Medio Superior.

Con base en el diagnóstico derivado de la propuesta, elaborar reportes de resultados de la prueba Enlace Media Superior para los estudiantes, docentes y autoridades educativas.

# **CAPÍTULO I. EL ANDAMIAJE TEÓRICO**

En este capítulo se describen conceptos que consideramos necesarios para entender el ámbito en el que se inserta el reporte de resultados de una prueba; están presentes directa o indirectamente al mencionar un reporte de resultados y es preciso hacer hincapié en ellos ya que proporcionan al lector no especializado un bagaje específico sobre una prueba estandarizada de aprovechamiento.

### **¿Qué es una prueba psicológica?**

Diversos autores definen de manera muy similar qué es una prueba psicológica, pero cada uno enfatiza ciertas características dependiendo de su interés o enfoque.

- Medida objetiva y estandarizada de una muestra de conducta (Anastasi, 1998).
- Un proceso o recurso estandarizado que proporciona información de una muestra de conducta o de un proceso cognitivo de una forma cuantificable (Hogan, 2003).
- Procedimiento sistemático para observar el comportamiento y describirlo con la ayuda de escalas numéricas o categorías fijas (Cronbach, 1998).

Estas definiciones breves y sencillas son el punto de partida; a continuación se describen algunos elementos y consideraciones relevantes.

- No se pueden obtener componentes exhaustivos de una conducta, esto es técnicamente imposible por la amplísima variedad de expresiones que puede

tener; así que obtener una muestra representativa es una manera que ha demostrado ser eficiente (Anastasi, 1998; Fernández-Ballesteros, 2004; Hogan, 2003; Martínez, 1995).

- La **subjetividad** (entendida como el criterio individual o personal para determinar las respuestas correctas en una prueba) está controlada o disminuida porque la calificación de la prueba no depende del juicio o criterio de una sola persona o calificador, sino de expertos que predeterminan los criterios de calificación (González, 2007; Fernández Ballesteros, 2004; Cronbach, 1998; Anastasi, 1998).
- Una prueba se aplica inicialmente a un grupo de personas con características similares a quienes la presentarán de manera habitual y con este grupo inicial se construye una referencia o norma contra la que se contrastarán los siguientes resultados; esto da como resultado una **tipificación** de la interpretación de los puntajes (Mc Intire, 2000; Fernández Ballesteros, 2004).
- Las pruebas tienen reglas o normas sobre la forma en la que se aplican, condiciones del lugar de aplicación, horas destinadas a la aplicación, instrucciones para los que responden la prueba, métodos de recolección del instrumento y métodos para la calificación o asignación de puntajes y todas estas condiciones deben ser siempre las mismas para cada ocasión que la prueba se presenta; a esto se le conoce como **estandarización** (Cronbach, 1998; González, 2007; Martínez, 1995;). Esto garantiza que las variables externas a las pruebas sean similares y controladas.
- Existe una cuantificación o categorización como resultado de la aplicación de las pruebas. La información que proporcionan puede estar en términos numéricos, que pueden ser tan sencillos como la cantidad de respuestas contestadas correctamente, hasta puntajes ponderados que consideran diferentes aspectos para proporcionar la calificación; su expresión puede ser

categoría en niveles como suficiente o insuficiente hasta más desagregados o detallados (Hogan, 2003).

Con los elementos anteriores podemos ampliar la definición de una prueba y concebirla como un instrumento de medición o cuantificación de una muestra de conducta, que se administra de forma controlada y sistemática, cuyos criterios de calificación son siempre los mismos y las calificaciones están basadas en una referencia o norma de sujetos con las mismas características.

Al inicio mencionamos que el concepto de prueba no cambia sustancialmente y que cada autor enfatiza elementos importantes; en una descripción más amplia, una prueba es “un método para adquirir una muestra de conducta preestablecida bajo condiciones controladas, donde la aplicación y calificación se realizan de manera siempre similar y cuya interpretación de resultados es con los mismos parámetros” (Walsh y Betz, 2001).

Más que elegir una definición, preferimos tener presentes los elementos que ayudan a comprender cualquier concepto sobre las pruebas estandarizadas y a lo largo del trabajo no utilizaremos una concepción autoral particular.

### **¿Aplicar una prueba es evaluar?**

Una distinción que consideramos importante es la diferencia entre **evaluar** y **aplicar** una prueba estandarizada; esta distinción es relevante porque es frecuente encontrar la expresión (especialmente en medios noticiosos) “evaluación” cuando se hace referencia a la aplicación de una prueba estandarizada y hacer esto lleva en muchas ocasiones a interpretaciones o afirmaciones inadecuadas.

Para el Grupo de Evaluación de las Naciones Unidas, una red profesional que agrupa a unidades responsables de evaluación en el sistema de las Naciones Unidas e incluye agencias especializadas, fundaciones, programas y organizaciones afiliadas; la evaluación es:

... una valoración, lo más sistemática e imparcial posible, de una actividad, proyecto, programa, estrategia, política, tópico, tema, sector, área operativa, desempeño institucional, etcétera. Incide principalmente sobre los logros esperados y alcanzados, examinando la cadena de resultados, los procesos, los factores contextuales y la causalidad, a fin de entender los logros o la ausencia de éstos. Su objetivo es determinar la relevancia, el impacto, la efectividad, la eficiencia y la sostenibilidad de las intervenciones y contribuciones de las organizaciones del sistema de las Naciones Unidas. Una evaluación debe suministrar información basada en evidencia que sea creíble, fiable y útil, facilitando la incorporación oportuna de los hallazgos, recomendaciones y lecciones en los procesos de toma de decisiones de las organizaciones del sistema de las Naciones Unidas y de sus miembros (UNEG, 2005).

La descripción representa los alcances de una actividad evaluativa; en términos aplicables al ámbito psicológico, una evaluación psicológica es la “recolección e integración de datos relacionados con la psique individual con el propósito de hacer una evaluación psicológica, concretada mediante el uso de herramientas como pruebas, entrevistas, estudios de caso, observación conductual, así como equipos y procedimientos diseñados específicamente para la medición de ésta” (Cohen y Swerdlick, 2006).

Al asumir que en un ejercicio de evaluación se proporciona una valoración por medios sistemáticos es comprensible que una prueba sea el instrumento adecuado por sus características de estandarización, objetividad y representatividad de cierto rasgo o atributo expresable en un puntaje o categoría; pero es importante considerar que una evaluación puede ser más que la aplicación de una prueba porque se pueden utilizar diversos instrumentos o acciones para obtener información.

La adecuada comprensión nos ayuda a entender que para evaluar podemos usar una prueba con el conocimiento del alcance y naturaleza de ésta, pero que no es

el único instrumento disponible para realizar una evaluación. Se aplica una prueba como un ejercicio de evaluación, pero la evaluación no siempre es la aplicación de una prueba.

## Tipos de pruebas psicológicas

¿Por qué abordar los tipos de pruebas si el objetivo son las de aprovechamiento? Porque es importante conocer las clasificaciones de las pruebas ya que nos permite reconocer que, acorde con diversas intenciones, existen diferencias.

Las clasificaciones son diversas y dependen básicamente de la perspectiva con que se analicen; en la tabla 1 se muestra una clasificación que contiene las categorías habituales sobre pruebas en la literatura especializada.

**Tabla 1. Tipos de pruebas acorde con su intención**

Intención	Categoría
Sobre el desempeño o actuación en la prueba	Máximo desempeño
	Observación de la conducta
	Autorreportes
Sobre la población de referencia	Estandarizada
	No estandarizada
Por la intención al responder	Objetiva
	Subjetiva o proyectiva
Por las dimensiones de interés	De desempeño
	De aptitud
	De inteligencia
	Inventarios de interés
	De personalidad

*Fuente:* Elaboración propia a partir de *Foundations of Psychological Testing* de Mc Intire S. y Miller L. (2000). EE. UU.: McGraw-Hill.

Las categorías no siempre son absolutas, incluso pueden presentarse en varios escenarios con diferentes intenciones; por ejemplo, una prueba que evalúe el aprendizaje de un idioma extranjero puede servir como instrumento de máximo

desempeño para tomar decisiones de alto impacto, como el ingreso a una institución educativa o como instrumento de diagnóstico en un instituto que imparte idiomas.

No es nuestro objetivo hacer una descripción de cada categoría; la literatura sobre pruebas psicológicas lo hace de manera adecuada y extensa (Anastasi, 1998; Aiken, 1996; Cohen y Swerdlick, 2006; Cronbach, 1998; Mc Intire, 2000). Lo que interesa a nuestro trabajo son aquellas categorizadas como pruebas de desempeño, descritas a continuación.

### **Características de las pruebas de aprovechamiento**

Dependiendo del autor en idioma español o del traductor de la obra al español, las pruebas que llamaremos de aprovechamiento tienen diferentes denominaciones; en ocasiones son mencionadas como test de rendimiento (Aiken, 1996, Martínez, 1995), pruebas de logro (Cohen & Swerdlick, 2006), test de nivel o dominio (Jornet, 1996), test de desempeño (Mc Intire, 2000) o test de aprovechamiento (Anastasi, 1996; UNESCO; 2000); todas hacen referencia a la prueba que tiene como intención medir un atributo aprendido. Acorde con diferentes autores, estas pruebas miden:

- Qué tanto ha aprendido un alumno para pasar de un curso a otro o si tiene los niveles mínimos que exige un programa de formación (Jornet, 1996).
- La cantidad de conocimientos que fueron aprendidos u olvidados (UNESCO; 2000).
- Los efectos de los programas de instrucción o entrenamiento (Anastasi, 1998).

- Las adquisiciones de aprendizaje o conocimientos específicos, con finalidades de diagnóstico y /o pronóstico concretas (Martínez, 1995).
- Los conocimientos adquiridos o el grado de aprendizaje que ha tomado lugar (Cohen y Swerdlik, 2006).

Podemos afirmar a partir de las descripciones anteriores que el elemento común es qué o cuánto aprendió un individuo de un universo de contenidos homologados, como suelen ser los aprendidos en el aula. Esta afirmación es más relevante al considerar que hay pruebas que miden lo que un individuo ha aprendido en contextos más diversos que los escolares y suelen ser usadas para hacer predicciones o para tener elementos que puedan ayudar a interpretar conductas futuras; éstas son categorizadas como pruebas de aptitud (Anastasi, 1998), y es necesario hacer distinciones entre éstas y las de aprovechamiento.

Las características principales de las pruebas de desempeño con relación a las de aptitud son, según Anastasi, que las pruebas de aprovechamiento:

- Miden los efectos de conjuntos relativamente estandarizados de experiencias, como los cursos escolares.
- Son enfáticas en conocer lo que el individuo puede hacer en ese momento.
- Definen de forma empírica y teórica el universo de referencia mejor que las pruebas de aptitud porque dicho universo suelen ser programas, cursos o materias cuyos contenidos son finitos y más o menos concretos.
- Pueden aspirar a verificar el proceso y no sólo el producto por estar referidas a dominios concretos donde se pueden identificar unidades en el dominio.

- El estándar de referencia para dar un puntaje suele ser un criterio, más que una referencia con la población. Es decir, importa satisfacer un criterio antes que saber el desempeño del individuo respecto a los demás sujetos que hayan presentado la misma prueba.

La ventaja de utilizar pruebas de aprovechamiento es que, si tienen un buen modelo de construcción, se asegura la uniformidad y la adecuada cobertura de los contenidos, así como la disminución de los factores aleatorios o irrelevantes en la calificación (Anastasi, 1998). Como afirmábamos en las secciones iniciales, esto no significa que las pruebas de aprovechamiento se puedan utilizar como únicos instrumentos para tomar decisiones o que se descuiden otros aspectos de mediciones en el ámbito educativo; hay autoridades educativas que advierten sobre los límites de uso de una prueba, como aquellas en donde se utilizan como el único elemento para tomar decisiones respecto a la calidad educativa en general (Popham, 1999). Reconocer las características adecuadas de una prueba de aprovechamiento es un requisito fundamental para quienes analicen el tema de las pruebas en ámbitos educativos.

### **Usos de las pruebas de aprovechamiento**

Las pruebas de aprovechamiento se utilizan en diferentes escenarios, siendo común el educativo, aunque también se usa en el organizacional (Anastasi, 1998; Hogan, 2003). Son las pruebas más usuales de todas, y también las más evidentes para el público, probablemente por su difusión en los medios y el impacto que pueden tener en los usuarios. Los usos más comunes de estas pruebas son:

- Evaluar la competencia mínima en habilidades básicas.
- Examinar en el nivel universitario, donde se utilizan para la admisión, colocación, diagnóstico o egreso de alumnos.

- Admisión a las escuelas de posgrado, donde la mayoría de las pruebas son una combinación de pruebas de inteligencia o de habilidades y de aprovechamiento.
- Diagnosticar y pronosticar donde se utilizan para señalar las debilidades o fortalezas en un tema particular e incluso sugerir causas o líneas de acción para resolver algunos problemas; en su mayoría son pruebas relativas al lenguaje, matemáticas o lectura.
- Certificar al proporcionar un documento que avale la capacidad del ejercicio profesional en un determinado campo.

Al reconocer que cada prueba de aprovechamiento tiene su propia intención contamos con mejores elementos para abordar el análisis de cualquiera de ellas. Al cuestionar cuál es la intención de una prueba debemos saber que existen diferentes opciones y buscar información en función de lo que significa medir, pero para esto es necesario saber cómo evaluar una prueba, o cuáles son los elementos que nos brindan información sobre su calidad. En general, los indicadores son dos: la confiabilidad y la validez.

### **Indicadores de calidad de una prueba: la confiabilidad y la validez**

Toda vez que determinamos anteriormente las características de las pruebas de aprovechamiento y sus usos, es necesario conocer las cualidades que conforman una prueba de buena calidad o los elementos que ayudan a decidir si su construcción es la adecuada. Diversos elementos nos ayudan a entender estas cuestiones: algunos son indicadores obtenidos a través de procedimientos estadísticos o matemáticos, otros a través de procedimientos que involucran juicios de diferentes expertos sobre la pertinencia de algunos componentes de la prueba; todos estos procedimientos nos brindan evidencias de calidad, en algunos casos algunas evidencias son más relevantes que otras, e incluso no siempre se

cumplen todos los criterios de calidad, sin que esto signifique un mal desempeño de una prueba. Explicaremos esto más adelante.

Las pruebas no tienen un único indicador de su calidad; al ser instrumentos complejos que obtienen evidencia de un atributo y entender que el atributo es una construcción teórica de la realidad, la certeza de lo que pretenden medir está sujeta de forma ineludible a la manera en la que se concibe ese atributo independientemente de la prueba. Por ejemplo, una prueba que evalúe el desempeño en un área de conocimientos particular tiene un universo más concreto de medición que otra que evalúe el desempeño de un profesional al egreso de su carrera, cuyo universo de medición es amplio y puede ir desde conocimientos conceptuales hasta demostraciones del desempeño como recién egresado. Tomando esta consideración, las evidencias de que una prueba es adecuada son de naturaleza diversa y se dirigen a diferentes componentes que, en su conjunto, nos indican si una prueba es pertinente y cumple con adecuados criterios de calidad.

### **La validez**

Acorde con los *Standards for Educational and Psychological Testing* (1999) nombrados a partir de aquí simplemente como *estándares*, la validez es el grado en el que la evidencia y la teoría apoyan la interpretación de los puntajes con relación a los usos propuestos en una prueba. Debemos considerar varios aspectos para comprender mejor el concepto (Cronbach, 1998):

- No son las pruebas las que demuestran validez, sino las puntuaciones o interpretaciones que se derivan de éstas; no se puede hablar, en términos técnicos de una prueba válida o inválida, sino del grado en el que la evidencia obtenida nos dice si los puntajes se relacionan con el uso propuesto de cada prueba.

- Para apoyar las interpretaciones se necesitan diferentes evidencias en forma de datos o información; no es viable una única fuente de información, por muy sólida que sea. Dado que se pueden realizar diferentes inferencias con los puntajes, se debe proporcionar diferente información para cada una de ellas. Así también podemos afirmar que un proceso de validación puede estar conformado por diferentes e independientes subprocesos.
- No se puede evidenciar a través de un solo índice numérico, porque no hay uno único; lo que se puede inferir de los índices es el grado o fuerza en el que cada evidencia apoya la interpretación.

En palabras de Lee J. Cronbach “la meta final de la validación es la explicación y la comprensión”; si hacemos algunos planteamientos al analizar una prueba desde la perspectiva de la validez, algunos serían: ¿cuál es el alcance que tiene esta prueba respecto con lo que dice que mide? ¿Qué evidencia existe para asegurar que el criterio que mide la prueba es el adecuado? ¿Los ítems que conforman la prueba representan lo que dice ésta que evalúa? Estos planteamientos abordan diferentes formas o procedimientos de obtener información sobre lo válido de una prueba. Las clasificaciones habituales que se encuentran en la literatura (American Educational Research Association & American Psychological Association, & National Council on Measurement in Education, 1999; Cronbach, 1998; Downing & Haladyna, 2012; Gonzalez, 2007; Mc Intire & Miller, 2000) sobre la validez son las mostradas a continuación.

#### Validez relativa al contenido de la prueba

Proporciona información sobre si los ítems son representativos del universo o dominio que pretende cubrir la prueba. Sirve para sustentar que cada muestra individual de la conducta corresponde con una porción del rasgo que mide la prueba; en un total se puede afirmar que el conjunto de reactivos miden el rasgo propuesto.

### Validez relativa al constructo

Determina si la información que proporciona la prueba está relacionada con un sustento o modelo teórico. Proporciona evidencia para fundamentar la hipótesis de una relación entre un nivel teórico no observable directamente y uno observable a través de los datos obtenidos en una prueba. Hay dos formas de demostrar esta validez.

1. Convergente. Demuestra que el constructo que teóricamente debe estar relacionado con los puntajes realmente lo está.
2. Divergente. Muestra que los constructos que no deben estar relacionados con el puntaje de la prueba realmente no lo están. En ocasiones son atributos con los cuales uno puede confundirse o ser parecidos.

### Validez referida al criterio

Es el grado en que la medida de un atributo demuestra una asociación con algún indicador independiente o externo de ese mismo atributo. Hay dos tipos:

**Predictiva.** Cuando un criterio es medido posteriormente a la obtención de los puntajes del rasgo que se considera predictor; lo que se busca es saber cómo el estatus presente de la prueba predice el estatus futuro de un criterio.

**Concurrente.** Es cuando el predictor y el puntaje son obtenidos al mismo tiempo; el interés es conocer el estatus presente en la prueba y en el criterio de interés.

Los estándares mencionan que si bien la validez debe considerarse como un concepto unitario, se obtienen evidencias de ella a través de diferentes aproximaciones, que categoriza de la forma siguiente.

### Evidencia basada en el contenido de la prueba

Es la que se obtiene a través de analizar la relación entre el contenido de la prueba y el constructo que se pretende medir; el contenido pueden ser los temas,

redacción o formatos de los ítems, las tareas, preguntas o incluso las guías de cómo administrarlos o manuales técnicos. Esta información puede usarse para mostrar diferencias en el significado o las interpretaciones de los puntajes entre subgrupos de examinados, el grado en el que el constructo tiene poca representatividad o componentes irrelevantes que pueden favorecer a algunos subgrupos de examinados.

#### Evidencia basada en los procesos de respuesta

Es información que proviene del análisis de respuestas individuales, del análisis teórico o empírico de los procesos de quienes respondieron la prueba para obtener evidencia de la relación entre el constructo y el desempeño o respuesta de los examinados.

#### Evidencia basada en la estructura interna

Es información del grado de relación entre los ítems o componentes de la prueba y el constructo en el cual se basan las interpretaciones de la prueba. Puede estar restringida a una dimensión o varias que se espera sean homogéneas. El tipo de análisis y su interpretación dependerán del uso de la prueba.

#### Evidencia basada en su relación con otras variables

Es el análisis de la relación de los puntajes de una prueba con variables externas, que pueden ser mediciones del mismo criterio que el medido por la prueba, relaciones con otras pruebas que midan lo mismo o pruebas que midan constructos similares o diferentes.

#### Evidencia basada en las consecuencias de la prueba

En primer lugar es importante distinguir que hay evidencias de las consecuencias de las pruebas que sirven a la validez y evidencias que dan información para tomar decisiones sobre políticas sociales pero que están fuera del ámbito de la validez.

La evidencia derivada de las consecuencias da información de motivos de invalidez como una baja representación del constructo o componentes irrelevantes a éste. Si las pruebas se administran con la expectativa de proporcionar beneficios a partir de su uso, un propósito fundamental de la validez es dar información de tal beneficio. En ocasiones las afirmaciones del beneficio van más allá del uso directo de los puntajes (como en pruebas educativas donde se afirma que ayudan a motivar a los estudiantes a mejorar su desempeño o donde afirman cambios en las prácticas escolares a través de proporcionar valiosos resultados de aprendizaje a los educadores); el examen directo de las consecuencias de la prueba es de gran importancia. El proceso de validación en estos casos debe dar evidencia de la existencia de dichos beneficios.

En general, la literatura sobre pruebas aborda el tema de validez desde enfoques puramente métricos (Fernández-Ballesteros, 2004; Martínez, 1995; Patton, 2002; Stake, 2010) hasta consideraciones reflexivas u orientadoras (Cronbach, 1998; Walsh & Betz, 2001; Weiss, 2008) de manera muy extensa y detallada en la mayoría de los casos. Debemos considerar, a partir de lo expuesto anteriormente, que la evidencia que proporcione información sobre el uso de los puntajes con relación al uso propuesto de la prueba determina la fuerza respecto a lo que se puede afirmar de una prueba; si una prueba afirma medir algo es necesario mostrar evidencia acerca de ello y nosotros debemos tener la capacidad de interpretarla y cuestionarla de manera adecuada.

### **La confiabilidad**

Otra cualidad esperada en las pruebas es la confiabilidad de la medida y hace referencia a que, cada ocasión que se mide un rasgo bajo las mismas condiciones se espera obtener resultados similares en los puntajes (Walsh, 2001). O en forma de planteamiento, según Cronbach: si la gente fuera estimada dos veces, ¿coincidirían las puntuaciones? ¿hasta qué punto? Para comprender mejor el concepto es necesario entender tres afirmaciones.

- Todos los que responden la prueba tienen una cantidad del rasgo que se mide, nombrado *puntaje verdadero*. Si aplicáramos la misma prueba al mismo sujeto de manera constante, se obtendría su calificación real que no depende de factores aleatorios.
- Cada vez que se aplica una prueba existe un grado de *error de medición* en las puntuaciones, que puede ser originado por las características del sujeto, de la prueba o de situaciones que no tienen relación con el rasgo a medir, pero que afectan a las puntuaciones.
- La unión de los dos elementos anteriores implica que lo que se obtiene entre el puntaje verdadero y un grado de error es el *puntaje observado*, es decir, el puntaje de un individuo al responder la prueba.

En una prueba considerada confiable las fuentes de error son controladas y mantenidas hasta ciertos límites. La confiabilidad se expresa a través de un índice y cada ocasión que se obtiene un puntaje es necesario saber el rango de error, es decir, ¿qué puntaje habría obtenido cualquiera si se le hubiera aplicado la prueba en otra ocasión? Este rango se expresa como Error Estándar de Medida (SEM, por sus siglas en inglés); cuanto mayor es éste, menos precisa es la medida; es un rango que nos indica la variación de la medida en una persona, por ejemplo, si alguien obtiene un puntaje de 80 en un campo dado y un SEM de 3.5 significa que su puntuación verdadera se encuentra entre 76.5 y 83.5.

La confiabilidad, a diferencia de la validez, sí puede expresarse de manera numérica con diferentes procedimientos, los más comunes son el test-retest, formas alternas, división por mitades y consistencia interna.

## **¿Qué es más importante para determinar la calidad de una prueba, la validez o la confiabilidad?**

Entre las evidencias de validez y los indicadores de confiabilidad, ¿qué es más importante para determinar su calidad? Dependerá de las intenciones propuestas de la prueba; aunque es importante distinguir entre afirmar la importancia de una evidencia sobre otras y no dar información de éstas. Es muy importante la información sobre la validez de constructo en una prueba, pero no significa que no es necesaria la información sobre la confiabilidad. Para poder determinar la relevancia de cada proceso de evidencia debemos tener claro el propósito de la prueba, alcance, limitaciones, y quién la diseña y elabora (Cronbach, 1998).

Para algunos expertos la confiabilidad es la principal característica porque una prueba que no es confiable puede contener muchas respuestas azarosas o un rango de error que la convierte en una prueba poco útil. En palabras de Walsch (2000):

La dependencia de la validez en la confiabilidad se basa en el hecho de que una prueba no confiable no se relaciona a algo en concreto y por lo tanto será inútil. En otras palabras, si no tenemos confianza en la precisión, consistencia o repetibilidad de nuestra medida, no vale la pena intentar relacionar el instrumento con otras conductas, rasgos o medidas. Así, la confiabilidad es un prerrequisito necesario para la validez y es la esencia del concepto de la medición. (p 56).

Sin embargo, para Cronbach la validez es el criterio más relevante; este autor menciona que:

La base fundamental para elegir un test es su validez. ¿Podemos interpretar el test claramente? ¿Sirve la información obtenida a nuestros propósitos? La fiabilidad es una consideración suplementaria. Una excelente fiabilidad no puede compensar una validez inaceptable. Si la validez es aceptable, la información sobre la fiabilidad ayuda a interpretar las puntuaciones y puede sugerir cómo incrementar la validez (por ejemplo, repitiendo las medidas). (p 244).

Si bien podemos tomar alguna de las perspectivas anteriores, debemos ser capaces de reconocer la evidencia de ambos procesos y reconocerlos como información valiosa sobre la adecuada construcción de la prueba. Las pruebas que no dan información de la validez o de la confiabilidad deben también dar cuenta de por qué no lo hacen.

Si bien ya hemos mencionado los elementos que nos dan información para saber si una prueba se diseñó de acuerdo con su intención y es fiable, ¿dónde podemos obtener información para saber si una prueba está bien diseñada? ¿Dónde la obtienen los propios diseñadores y usuarios de las pruebas? Existen diversos documentos que sirven a estos propósitos.

### **Documentos que guían el diseño y construcción de las pruebas estandarizadas**

Existen diferentes instituciones y documentos que proporcionan normas sobre el diseño, aplicación, derechos y responsabilidades de los constructores y usuarios de las pruebas (American Educational Research Association & American Psychological Association, & National Council on Measurement in Education, 1999; Downing & Haladyna, 2012; Comité conjunto de estándares para la evaluación educativa, 1998; National Council on Measurement in Education, 1995) Son referencias comunes en prácticamente todos los textos básicos sobre pruebas psicológicas y están redactadas en forma de sentencias breves, algunas con más o menos detalle sobre su contenido. A continuación se mencionan las más citadas (Anastasi, 1998; Cohen y Swerdlick, 2006; Cronbach, 1998; Hogan, 2003) y en secciones posteriores haremos mención a contenidos particulares de algunos de ellos.

- *Standards for Educational and Psychological Testing*. Elaborado por el *Joint Committee on Standards for Educational and Psychological Testing of the*

*American Educational Research Association, la American Psychological Association y el National Council on Measurement in Education de 1999.*

- *Code of Fair Testing Practices in Education.* Del Joint Committee on Testing Practices de 1998.
- *Ethical Standards of the American Educational Research.* De la *American Educational Research Association* de 1992.
- *Ethical Principles of Psychologist and Code of Conduct* de la *American Psychological Association* de 1992.
- *Code of Professional Responsibilities in Educational Measurement.* Del *National Council on Measurement in Education* de 1995.

Es relevante afirmar que estos documentos son creados en lengua inglesa y las consideraciones en muchos casos obedecen a realidades políticas, sociales y legales de los diversos comités que los elaboran; así que el uso o aplicación de cualquier estándar requiere de diversas consideraciones. Cronbach menciona, respecto al uso de los estándares para evaluar una prueba lo siguiente:

Los estándares no deben ser considerados individualmente. Por tanto, la valoración de la adecuación de un test implica el juicio de un profesional basado en los conocimientos de las ciencias de la conducta, de la psicometría y del campo en el que se apliquen los tests; el grado de cumplimiento de los estándares satisfecho por el editor del test; las alternativas disponibles y la evidencia experimental disponible. (pp.176).

Esta aseveración es particularmente relevante para quienes construyen pruebas en países con diferentes referentes nacionales y locales; la cuidadosa selección, adecuación o elaboración de estándares debe ser una tarea tomada con responsabilidad, profesionalismo y multitud de perspectivas. Pero es inevitable la

referencia a éstos porque son criterios utilizados de forma constante en el diseño de las pruebas.

## Pasos para elaborar una prueba

Es importante conocer los elementos que conforman una prueba que se puede considerar eficaz; si bien no siempre se presentan en el mismo orden, son elementos presentes en el proceso de construcción. En la tabla 2 se muestran doce pasos considerados esenciales para desarrollar una prueba eficaz, junto con algunas tareas que se realizan en cada uno de ellos y los estándares relacionados (acorde con los estándares). Estos son un referente de las tareas, de algunas actividades relacionadas y de los estándares en los que se pueden basar para su comprobación (Downing, 2012).

**Tabla 2**  
**Doce pasos para el desarrollo eficaz de una prueba**

<b>Pasos</b>	<b>Ejemplo de tareas en el desarrollo de la prueba</b>	<b>Estándares relacionados</b>
1. Plan general	Orientación sistemática sobre todas las actividades del desarrollo de la prueba: constructo; interpretaciones deseadas de la prueba; formatos; fuentes principales de la evidencia de la validez; propósito claro; inferencias deseadas; modelo psicométrico; cronología; seguridad; control de calidad.	1.1 3.2 3.9
2. Definición del contenido	Plan de muestreo del dominio o universo; métodos relacionados con el propósito de la evaluación; fuentes esenciales de la evidencia de validez de contenido	1.6 3.2 3.11 14.8
3. Especificaciones de la prueba	Definiciones operacionales del contenido; marco de referencia de la evidencia de la validez del muestreo sistemático y justificable del dominio de contenido; norma o criterio al que se hace referencia; características deseadas del reactivo.	1.6 3.2 3.3 3.4 3.11
4. Desarrollo de reactivos	Desarrollo de los estímulos eficaces; formatos; evidencia de validez relacionada con la adhesión a los principios basados en la evidencia; capacitación de los redactores y revisores de los reactivos; corrección adecuada de los reactivos; varianza irrelevante para el constructo (VIC) debida a defectos.	3.6 3.7 3.17 7.2 13.18
5. Diseño y	Diseño y elaboración de las formas de la prueba; selección de los	3.7

	ensamble de la prueba	reactivos para formas especificadas de la prueba; muestreo operacional de acuerdo con el proyecto planeado; consideraciones del piloteo.	3.8
6.	Producción de la prueba	Actividades de publicación; impresión o presentación de la evaluación basada en computadoras (CBT); cuestiones de seguridad; problemas de validez concernientes al control de calidad.	No aplican
7.	Aplicación de la prueba	Problemas de validez concernientes a la estandarización; problemas relacionados con la Ley para Estadounidenses con Discapacidades (ADA); supervisión; problemas relacionados con la seguridad y la sincronización.	3.18 3.19 3.20 3.21
8.	Calificación de las respuestas	Problemas de validez; control de calidad; validación de la clave; análisis de reactivos.	3.6 3.22
9.	Calificaciones de aprobación	Establecimiento de calificaciones de aprobación; relativas o absolutas, que puedan justificarse; problemas de validez concernientes a los puntos de corte, equivalencia de los estándares; mantenimiento de la constancia de la escala de calificación (equiparación, igualación).	4.10 4.11 4.19 4.20 4.21
10.	Reporte de los resultados de la prueba	Problemas de validez; exactitud; control de calidad; oportunidad; significado; problemas por el uso incorrecto; desafíos; repetición de la prueba.	8.13 11.6 11.12 11.15 13.19 15.10 15.11
11.	Banco de reactivos	Problemas de seguridad; utilidad; flexibilidad; principios para la formación eficaz del banco de reactivos.	6.4
12.	Reporte técnico de la prueba	Documentación sistemática, minuciosa, y detallada de la evidencia de validez; organización en doce pasos; recomendaciones.	3.1 6.5

*Nota:* Tomado y adaptado de Downing, S. Doce pasos para el desarrollo de pruebas eficaces. pp.27 y 28. En Downing S. M. & Haladyna T.M. (Eds.). (2012). *Manual para el desarrollo de pruebas a gran escala.* (2 vols.). (Centro Nacional de Evaluación para la Educación Superior Trad.). D.F., México.: Centro Nacional de Evaluación para la Educación Superior. (Trabajo original publicado en 2006).

Los elementos mostrados en la tabla nos dan una perspectiva en la que podemos ubicar diferentes tópicos de la construcción de la prueba, con los pasos que nos ayudan a ubicar a cada uno con relación a los demás. Es un mapa o guía que indica dónde se sitúa una tarea en relación con otras. En nuestro trabajo nos enfocaremos en el paso de los reportes de resultados que describiremos a continuación.

## **Los reportes de resultados de la prueba**

Los reportes de resultados de las pruebas son formas de comunicar, que incluyen un emisor, mensaje, medio, intención y audiencia; su eficacia radica en la capacidad de crear un vínculo efectivo entre el reporte, la audiencia intencional y el uso intencional de la prueba (Ryan, 2012), Los reportes se entregan a los estudiantes, sus padres, profesores, directores y autoridades educativas de nivel local o federal; que sea útil para una audiencia no implica que lo sea para otra y cada reporte debe ser evaluado en función del adecuado cumplimiento de su propósito.

## **Lineamientos para elaborar reportes**

Cuando se hace referencia al análisis de reportes (Davidson, 2005; Downing & Haladyna, 2012; Mc Intire & Miller, 2000 ; National Center for Education Statistics, 2012; National Education Goals Panel, 1998) es común encontrar dos documentos mencionados anteriormente: los estándares y el *Code of Fair Testing Practices in Education*; estos documentos, si bien están publicados en lengua inglesa, pueden ser utilizados como referentes adecuados sobre la manera en la que se deben abordar cuestiones particulares en los reportes de evaluación; en la literatura existen más documentos usados como referentes (como el *Code of Professional Responsibilities in Educational Measurement* o los documentos de la *National Education Goals Panel*) pero sus recomendaciones son en función de mandatos legales o políticos como documentos creados para servir como referentes en ámbitos nacionales, y por ende, aplicables y pertinentes, a dichos contextos.

De los estándares identificamos veintiséis estándares que consideramos útiles para diversos propósitos relacionados con los reportes de resultados, descritos de manera detallada en el anexo A. Cada uno de ellos hace referencia a diferentes partes o componentes de un reporte.

Del *Code of Fair Testing Practices in Education* identificamos trece recomendaciones que hacen referencia a los contenidos de los reportes de forma directa o indirecta. Se describe cada uno de ellos en el anexo B.

Los estándares y recomendaciones sirven como elementos que guían la elaboración, pero no mencionan la manera en la que se deben crear los reportes; los métodos y procedimientos para crear reportes eficaces es una tarea multidisciplinaria que agrupa cuestiones no solo relacionadas con la psicometría sino con disciplinas muy diversas.

### **Literatura sobre reportes de resultados en pruebas**

La literatura sobre los reportes de resultados de las pruebas es lo suficientemente extensa para permitir conocer el tema de forma apropiada. Nina Deng y Hanwook Yoo (Deng y Yoo, 2009) elaboraron un documento que enumera artículos, capítulos de libros y trabajos presentados en conferencias en idioma inglés, categorizados en diferentes tópicos como lineamientos, niveles de reportes y audiencias, contextos de reportes, despliegues gráficos, y políticas de reportes y rendición de cuentas. Es un documento indispensable para quien desee revisar material de manera especializada sobre nuestro tema.

Sobre el tema para elaborar reportes de resultados a gran escala eficaces, la literatura más detallada está en idioma inglés, especialmente enfocada a la adecuada presentación de información, que puede ir desde recomendaciones basadas en análisis de diversos reportes (Goodman y Hambleton, 2004), entrevistas a diversas personas involucradas en la evaluación (Hambleton y Salter, 1997), recomendaciones generales, basados en diferentes perspectivas (Allalouf, 2007; Zeninsky y Hambleton, 2012; Aschbacher y Herman, 1991; Forte y ASR SCASS, 2002; Roberts y Gierl, 2010; Roberts y Gierl, 2011; Ryan, 2012;

Watermann y Klieme, 2002), comunicación de acuerdo con la audiencia (Camara y Shaw, 2012; Sireci y Forte, 2012; Zapata-Rivera, 2011), despliegues gráficos (Cleveland y McGill, 1984; Gillan, Wickens, Hollands y Carswell, 1998; Gribbons, 1992; Hullman, Rhodes , Rodríguez y Shah, 2011; Simkin y Hastie, 1987; Wainer, 1984; Wainer, 1997). La literatura en español sobre estos tópicos es prácticamente nula; se encuentra información que describe la necesidad de hacer reportes eficaces y que en ocasiones muestra ejemplos al respecto (Froemel, 2009) pero de ninguna manera es suficiente para un adecuado análisis del tópico.

## **CAPÍTULO II. LAS PRUEBAS ESTANDARIZADAS DE APROVECHAMIENTO EN MÉXICO**

Es este capítulo se mencionan las instituciones que se abocan a diseñar y aplicar instrumentos de evaluación educativa, así como las pruebas internacionales que se aplican o han aplicado en México y el uso de la información que se le ha dado a los resultados.

### **Las instituciones de evaluación de la educación en México**

No hay muchos datos detallados sobre la historia de las instituciones de evaluación educativa en México. La falta de información da la apariencia de una trayectoria inconsistente (García, 2005). Se menciona al Instituto Nacional de Psicopedagogía creado en 1936 como el primer organismo que se enfocó en la evaluación del aprendizaje de los alumnos (Rizo, 2001; en García, 2005). Durante varias décadas las evaluaciones fueron limitadas a ámbitos locales sin proyección nacional, hasta la creación de instituciones independientes o desconcentradas de la Secretaría de Educación Pública (SEP) la evaluación nacional formó parte del sistema educativo.

En la década de los setenta del siglo XX se creó la Dirección General de Evaluación de la SEP, donde se conformó un grupo de trabajo a cargo de Raúl Talán para la creación de pruebas de rendimiento, desarrollo de manuales de elaboración, análisis de reactivos y la aplicación de pruebas de rendimiento a muestras nacionales; pero los resultados de este proceso de trabajo no fueron muy visibles para la comunidad educativa ni tuvieron impacto a nivel nacional (García, 2005).

En 1984 el Sistema Nacional de Investigadores vinculó la evaluación de los docentes o investigadores con compensaciones adicionales a su salario base; en 1989, la Comisión Nacional para la Evaluación de la Educación Superior dirigió la primera autoevaluación de Instituciones de Educación Superior; y, en 1991, se crearon los Comités Interinstitucionales para la Evaluación de la Educación

Superior (CIEES) para realizar evaluaciones diagnósticas de las instituciones de Educación Superior a través de comités de pares académicos (Vidal, 2009); pero estos esfuerzos aún no se concretaban en un modelo de diagnóstico de evaluación de gran alcance nacional.

Las instituciones que diseñan y elaboran pruebas estandarizadas tienen una historia reciente; en 1994 se creó el Centro Nacional de Evaluación para la Educación Superior (Ceneval) con el fin de elaborar pruebas de aplicación a gran escala estandarizadas para que los alumnos pudieran ingresar al nivel educativo medio superior por medio de la presentación de un solo examen y para evaluar los conocimientos y habilidades de alumnos egresados del nivel superior en diversas profesiones. Hasta este año el Ceneval elabora pruebas estandarizadas para instituciones privadas, particulares y públicas. Por la cantidad de instrumentos que ha creado (300, acorde con su página web) es la institución a nivel nacional con mayor número de pruebas estandarizadas para el nivel medio superior y superior.

En el 2002 se creó el Instituto Nacional para la Evaluación de la Educación (INEE) que abarca la evaluación de todo el sistema de educación básica que va desde el nivel preescolar hasta secundaria; elaboró dos instrumentos para evaluar algunos elementos del sistema educativo, la prueba de Evaluación Nacional del Logro Académico en Centros Escolares (Enlace) y los Exámenes de la Calidad y el Logro Educativos Excale (Cisneros-Cohernour, 2009; Vidal, 2009).

Para conocer qué tipo de evaluaciones se realizan en el país en las instituciones públicas de nivel medio superior, el INEE realizó en el 2003 un estudio llamado Acciones de Evaluación en las Instituciones Públicas de Educación Media Superior donde solicitó reportes a todas las instituciones de carácter público que imparten el bachillerato y analizó 140 estudios de evaluación enfocados a alumnos, profesores, sistemas, programas y currícula, aunque la mayoría dedicados a evaluar el aprendizaje, encontró que en éstos estudios había:

- Una evidente importancia de la función evaluativa expresada en el gran número de acciones o esfuerzos de evaluación reportados.
- Poca formación de cuadros técnicos e infraestructura institucional en torno a la evaluación.
- Una gran diversidad de esfuerzos en evaluación, que abarcaba desde evaluar estudiantes hasta componentes institucionales.
- Diferentes concepciones en algunos tipos de evaluación; incluso poca información sobre los marcos teóricos o conceptuales de la evaluación o modelos o enfoques evaluativos utilizados.
- Gran importancia en la evaluación de los estudiantes y en particular de la evaluación del aprendizaje.
- Poca evidencia de las acciones derivadas de los resultados o el seguimiento de estas acciones.
- Ninguna evidencia de que las instituciones educativas informen de los resultados a la comunidad educativa, padres de familia o sociedad, y al parecer sólo se reserva la información a las autoridades educativas.
- Diferentes instancias que norman y regulan las acciones de evaluación, variando en función del subsistema o institución.
- Ningún dato sobre acciones del análisis de costo-beneficio o costo-efectividad de los programas educativos.

Podemos concluir que aunque se practican en el país diversos ejercicios de evaluación en diversas instituciones, no se elaboran con el suficiente rigor metodológico ni con la utilidad de proporcionar información que permita emitir juicios empíricos sobre el valor de los sistemas y da la impresión de poca utilidad o relevancia de la evaluación.

## **Participación de México en evaluaciones internacionales**

Respecto a la participación de México en pruebas estandarizadas de aplicación internacional, la primera ocasión no fue afortunada. En 1994, la prueba TIMMS (Conocida en español como el Tercer Estudio Internacional de Matemáticas y Ciencias Naturales fue un proyecto de investigación patrocinado por la Asociación Internacional para la Evaluación del Logro Educativo, IEA y aplicado entre los años 1994 y 1995), pero el gobierno mexicano no permitió la divulgación de sus resultados (Vidal, 2009; Wolff, 2006). Posteriormente, la IEA cambió su reglamentación para evitar que una situación similar volviera a suceder; sin embargo, parte de los resultados los conservó la Dirección General de Evaluación Educativa (DGE) y el INEE elaboró años después reportes con esa información. Durante el año 2000 la SEP realizó una prueba con las mismas evaluaciones de TIMSS sin dar los resultados públicamente de nuevo (Backhoff y Solano, 2003). México no ha vuelto a participar en esa prueba.

Otra prueba internacional, que suele manejarse como un proyecto por la amplitud de los instrumentos que aplica (cuestionarios a diferentes usuarios) es PISA, que significa Programa para la Evaluación Internacional de Alumnos, es una prueba de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), que tiene el objetivo de evaluar la formación de los alumnos cuando llegan al final de la etapa de enseñanza secundaria hacia los 15 años; es la prueba más conocida en el país por la difusión que le dan los medios nacionales y por la cantidad de reportes donde se usan los datos. México sí participa y proporciona reportes desde el año 2000 trienalmente, incluso ha usado la prueba con intenciones más descriptivas que las usuales ofertadas por PISA, por ejemplo, México solicitó que se hiciera un muestreo nacional que permitiera hacer diagnósticos nacionales desde 2003 (OCDE, 2013). Al hacer un análisis de los resultados en 2000 con respecto al 2009 no hay diferencias significativas de un mejor desempeño general;

pero sí la hay en la mejora de resultados de alumnos de niveles inferiores de desempeño, probablemente asociada a la disminución de jóvenes en situación de rezago escolar grave. También se encontró que hay una excesiva proporción de jóvenes en niveles inferiores y una proporción muy pequeña en los superiores: muchos insuficientemente preparados y pocos preparados suficientemente. Las competencias lectoras en la subescala de *Acceder y recuperar* han mejorado, probablemente asociado a los esfuerzos de la Reforma y el Programa Nacional de Lectura. El análisis de los resultados evidencia que se pueden realizar estudios que proporcionen información para hacer reflexiones fundadas y con sentido.

En Latinoamérica se realiza una prueba coordinada por el Laboratorio Latinoamericano de la Evaluación de la Calidad de la Educación (LLECE) que funciona como una red de directores nacionales de evaluación educativa en América Latina y el Caribe que busca producir información y conocimiento para enriquecer las políticas educativa por medio de reportes de estudios que reflejen el estado de la calidad de la educación en la región. México participó en la primera implementación en 1997, pero, como en la prueba TIMSS las autoridades mexicanas no usaron los resultados argumentando la recepción de un informe técnico insuficientemente detallado sobre la metodología utilizada para el cálculo de los valores nacionales (UNESCO, 1998). Las aplicaciones posteriores, en 2006 y 2013, se han llevado a cabo sin incidentes.

## Pruebas estandarizadas de gran escala en México

En México se han elaborado diferentes pruebas estandarizadas que se aplican a grandes poblaciones a nivel nacional que han sido diseñadas por instituciones nacionales o internacionales; en la tabla 3 se muestran las más representativas (Vidal, 2009).

**Tabla 3.**  
**Principales instrumentos de evaluación que operan o han operado en el sistema educativo nacional.**

<b>Año</b>	<b>Instrumento</b>	<b>Descripción</b>
1989	IDANIS	Instrumento de evaluación al término de la educación primaria.
1994	EXANI	Prueba para ayudar a tomar las decisiones de admisión a la educación media superior
1994	EGEL	Prueba que proporciona información sobre los egresados de 34 carreras
1995	TIMSS	Tercer estudio internacional de matemáticas y ciencias (IEA).
1997	LLECE	Prueba ligada al currículo para tercero y sexto grados de educación básica. Desarrollada por UNESCO para los países latinoamericanos.
1998	Estándares nacionales	Primer esfuerzo por evaluar la educación básica nacional. En su lugar opera ahora la prueba Excale.
1998	PISA	Prueba internacional para evaluar las competencias básicas para la vida y el trabajo de los jóvenes de 15 años.
2000	ACREDITA-BACH	Prueba para acreditar los conocimientos equivalentes al bachillerato general.
2003	Excale	Prueba desarrollado por el INEE y destinada a evaluar al sistema educativo nacional.
2006	Enlace	Prueba que proporciona información a estudiantes, docentes y padres de familia acerca del nivel de cada alumno de educación básica y media superior.

*Nota:* Tomado y adaptado de Vidal, R. ¿Enlace, Exani, Excale o PISA?. (2009).: Centro Nacional de Evaluación para la Educación Superior (Ceneval).

Se aprecia que son pocas pruebas (si tomamos como referente la gran cantidad de pruebas que se elaboran en Estados Unidos de América), con una limitada capacidad para promover los resultados (García, 2005). Aunado a la tendencia de los medios de información que suelen proporcionar datos de tono alarmista sobre los resultados que arrojan los instrumentos (Dijk, 2009). Esto impide que la opinión pública forme una perspectiva más profesional y responsable sobre las pruebas.

### **La difusión de resultados de las pruebas estandarizadas en México**

La difusión de los resultados de las pruebas en México no es usual o eficaz; a pesar de ser un número pequeño de pruebas (y se esperaría mejor manejable respecto a los datos que proporcionan) no hay información concreta y profesional que mencione lo que se hace con los datos de las pruebas estandarizadas; hay una serie de análisis regionales que revisan el caso de Latinoamérica y enmarcan el caso de México en algunos temas de análisis, pero de manera general (UNESCO, 2003, 2007) o aplicado de forma muy particular a una prueba, especialmente en reportes del proyecto PISA o TIMMS. Aunque, en general las referencias sobre México hablan de una resistencia de las autoridades a publicar los resultados (Arregui, 2006; Dijk, 2009).

Ferrer (2003) menciona cinco acciones habituales de los resultados de evaluación en América Latina, en ocasiones simultáneos:

- Renunciar al uso de la información obtenida, donde los resultados no son publicados, difundidos o apenas utilizados.
- Devolver confidencialmente los resultados a las instituciones participantes, sin verse expuestos al escrutinio público.
- Publicar los resultados en forma de tablas de clasificación.

- Elaborar y difundir informes de evaluación; de diversa cobertura y para distintas audiencias, donde la importancia radica en la confianza en que la difusión pública de los resultados puede satisfacer las demandas de información planteadas por los diversos agentes educativos.
- Utilizar los resultados para proporcionar orientación pedagógica a instituciones y profesores.

Si bien estas prácticas son recomendables para las pruebas que se usan en el país, su aplicación es poca; las reticencias al ver un resultado de evaluación como castigo (Dijk, 2009) y los instrumentos diseñados con procesos poco claros crean una combinación difícil de superar para crear con los resultados de una evaluación acciones enriquecedoras.

No hay estudios en el ámbito nacional sobre los reportes de resultados de las pruebas; el INEE, que ha documentado las prácticas de evaluación desde talleres y concursos para que los usuarios utilicen la información, menciona que el problema es crear reportes de resultados útiles, pero no reporta ningún estudio sobre revisión o modificación de éstos (INEE, 2012), es decir, se reconoce el problema de la difusión de resultados pero no se ha hecho un ejercicio concreto de revisar los reportes en función de sí mismos o de analizar sistemáticamente cuál es la comprensión de los usuarios ante éstos. El Ceneval no tiene ningún documento público al respecto de este ejercicio.

## **CAPÍTULO III. LOS REPORTES DE RESULTADOS DE LAS PRUEBAS ESTANDARIZADAS**

## **Estándares, normas y recomendaciones sobre reportes de resultados**

El capítulo menciona cuáles, de todos los estándares y normas sobre diseño de pruebas, se enuncian sobre el reporte de resultados. También se muestran las recomendaciones de diferentes expertos sobre lo que debe y lo que no debe de contener un reporte de resultados eficaz.

Diversos documentos contienen la información que se espera obtener de los reportes de resultados, contienen sentencias llamadas estándares o recomendaciones y se mencionan constantemente en la literatura de pruebas psicológicas.

Los documentos formales en lengua española acerca de estándares o recomendaciones sobre qué se espera de un reporte de resultados son inexistentes; por ende, la mayoría de dichos estándares provienen de textos en inglés. De entre ellos, hay varios que se retoman habitualmente en la literatura de artículos especializados (Aiken, 1996; Anastasi, 1998; Cronbach, 1998; Goodman y Hambleton, 2004; Ryan, 2012), sin embargo, ninguno ha sido elaborado por mandato nacional o gubernamental. Se presentan los principales títulos:

- *Standards for Educational and Psychological Testing Measurement in Education*. Provee criterios para evaluar las pruebas, prácticas y efectos del uso de ellas, sin dar información psicométrica o metodológica sobre el desarrollo de las pruebas.
- *Code of Fair Testing Practices in Education*. Proporciona sentencias basadas en algunos de los datos contenidos en estándares *for Educational and Psychological Testing Measurement in Education*, pero redactadas de manera más sencilla, menos técnica y para audiencias particulares.

- *Estándares para la Evaluación de Programas*. Brinda orientaciones que mejoran la evaluación de programas en cuatro áreas.

En la literatura se pueden encontrar más documentos, pero éstos son creados para responder a solicitudes gubernamentales que diseñan sus estándares de evaluación basados en las necesidades nacionales, por lo que no siempre corresponden a las circunstancias de otros países.

La tabla 4 muestra el número o código de los estándares aplicables a diversos aspectos de los reportes de resultados. La descripción de cada uno aparece en los Anexos A y B.

**Tabla 4**  
**Diversas consideraciones para diseñar, elaborar o evaluar un reporte de resultados acorde con *Standards for Educational and Psychological Testing* o el *Code of Fair Testing Practices in Education*.**

Consideración	Estándar o Código	
Para evitar las malas interpretaciones de los resultados	Standard	15.11
	Code	C2
	Code	C3
	Code	C6
Para informar de las interpretaciones adecuadas	Standard	11.20
	Standard	13.13
	Code	C1
	Code	C8
Para evitar las malas interpretaciones de los puntajes	Standard	11.15
Para asegurar interpretaciones adecuadas de los puntajes	Standard	5.10
	Standard	12.9
	Code	B6
	Standard	13.14
Para los puntajes en escenarios educativos	Standard	13.16
	Standard	7.8
	Standard	8.8
	Standard	12.19
	Standard	13.19
Para puntajes acumulados, desagregados o convertidos como categorías	Code	C4
	Standard	5.11
	Standard	6.12
	Standard	12.15
Sobre reportes entregados por formatos o medios electrónicos	Standard	5.12
Sobre los resultados para grupos diferentes	Standard	5.13
Sobre la confidencialidad de los resultados	Code	B.7
	Standard	7.10
Sobre la entrega de resultados dependiendo de la intención de la prueba (alto o bajo impacto)	Standard	13.5
	Standard	11.6
Sobre el tiempo en que se entregará el resultado	Standard	11.17
	Standard	11.6
	Code	C.7

*Fuente:* elaboración propia.

Los Estándares para la Evaluación de Programas tienen una sección llamada “Claridad del informe”, donde proporciona recomendaciones para el proceso total,

desde la planeación de una evaluación hasta la entrega y formatos de resultados; sobre este último punto expone:

La claridad es esencial para la comprensión de la audiencia, la credibilidad del informe y su aplicación. Para que una evaluación sea útil debe ser comprensible. Los interesados deberían entender fácilmente los propósitos de la evaluación, qué se evaluó, cómo se llevó a cabo la evaluación, qué información se obtuvo, cuáles fueron las conclusiones, y cuáles las recomendaciones que se realizaron, si las hubo. En este contexto, la *claridad* se refiere al discurso, los ejemplos y descripciones explícitas y nítidas (p. 67).

Además, contiene algunas consideraciones al respecto (la nomenclatura U5 significa que son estándares relacionados con la utilidad del informe con respecto a su claridad):

**U5 B.** Hacer que los reportes sean tan breves, sencillos y directos como sea posible, y centrados en dar respuesta a las preguntas de evaluación.

**U5 F.** Utilizar el lenguaje técnico en un modo limitado y ayudar a los interesados a comprender términos técnicos o poco familiares (ej. validez interna y externa, fiabilidad, desviación estándar, análisis de categorías, diseños de muestra) proporcionándoles un glosario, un resumen separado e informes técnicos, y ofreciéndoles oportunidades para la formación.

**U5 H.** Hacer que los clientes y representantes de las audiencias destinatarias revisen el/los informe(s), y juzguen su claridad, imparcialidad y comprensión, antes de emitir el informe final.

**U6 A.** Solicitar al cliente y a representantes de los interesados que indiquen cuáles son sus objetivos fundamentales, y el formato y lenguaje que esperan que tenga el informe, y los momentos en que el informe sería más útil.

También menciona errores comunes que se pueden cometer, y evitarlos:

- No dar al informe un formato que permita mejorar su legibilidad.

- Enfatizar demasiado la metodología a costa de los resultados.
- No describir el programa de un modo adecuado, asumiendo que todos los interesados conocen ya los aspectos más específicos del programa.
- Preocuparse del valor teórico a costa del práctico.
- No tener en cuenta los valores de los interesados al hacer las recomendaciones.
- Asumir que los clientes van a leer lo que se les entrega.
- No intervenir si se aprecia que los resultados de la evaluación se utilizan mal o son mal interpretados.
- Creer que cualquier trabajo de evaluación está completo una vez que se ha entregado o presentado el informe de evaluación.

Si bien los estándares y recomendaciones mencionados son relevantes para un reporte de resultados, resulta imposible cumplir con todos. Cada estándar se aplica en función de lo que va a desarrollar el test y de una valoración de expertos (Cronbach, 1998). Incluso se les ha descrito como ambiguos, poco aplicables en casos específicos o faltantes de significado práctico en contextos específicos (Koretz, 2006), no obstante siguen siendo un referente obligado. Es con base en un análisis puntual que se puede confirmar o no su viabilidad en contextos específicos.

### **Diseño de los reportes de calificación**

Los estándares, códigos y recomendaciones contienen lo que se espera de un reporte de resultados, pero no cómo hacerlo; para esto es necesario analizar diferentes trabajos enfocados en describir las fallas de los reportes y los componentes que debe integrar un reporte de resultados con calidad.

En Estados Unidos de América los datos sobre cómo reportar información son abundantes desde que se decretó la iniciativa *No Child Left Behind Act*, en el 2001, la cual determinó como necesidad nacional mejorar los puntajes de los alumnos en las áreas de ciencias, lengua y matemáticas. Para ello, cada estado elaboró sus reportes de calificaciones que resultaron ser muy diferentes entre sí, en términos de formato, contenido y calidad (Goodman y Hambleton, 2004). Así, fue necesario elaborar guías o directrices para elaborar reportes eficaces.

En ese contexto, Goodman y Hambleton (2004) publicaron una investigación donde analizaron los reportes de calificación de once estados (Estados Unidos de América y Canadá), enunciaron las características y fallas comunes, concluyendo con recomendaciones para crear reportes. En un ámbito más concreto, Joseph Ryan (2012), enumeró y describió ocho características que deben tener los buenos reportes (mostradas en la tabla 5); posteriormente, aplicó algunas de ellas a un reporte de resultados de un examen estatal. Por otra parte, Roberts y Gierl (2009) revisaron diversos reportes de calificación y propusieron un esquema de reporte para el estudiante en el ámbito de las matemáticas, mostrado en una sección posterior. Distinguieron la elaboración de diferentes reportes enfocados a sus audiencias.

Instituciones como el *National Council on Measurement in Education* elaboran información acerca de la correcta difusión de los conceptos de evaluación en los reportes de resultados (Sireci y Forte, 2012; Zenisky y Hambleton, 2012). En México, el Instituto Nacional de Evaluación Educativa (INEE) hace numerosas reflexiones en torno a lo que se debe reportar y elabora los documentos nacionales existentes para tales ejercicios (INEE, 2005).

## Componentes de un reporte de resultados

Joseph Ryan (2012) analizó varios reportes estatales y comerciales; esto le sirvió para definir ocho características comunes (véase Tabla 5) que conforman un reporte de resultados.

**Tabla 5**  
**Marco para el reporte de resultados con características, opciones y notas**

Características del reporte	Opciones y notas
<b>Destinatarios del reporte</b>	<ul style="list-style-type: none"> <li>• Estudiantes, padres, maestros, directores, administradores distritales y/o funcionarios del estado.</li> <li>• Los reportes se preparan para diversos destinatarios; su contenido y la manera en la que se presenta la información varía de acuerdo con el destinatario.</li> </ul>
<b>Escala o métrica del reporte</b>	<ul style="list-style-type: none"> <li>• Calificación cruda, porcentaje de aciertos, escala de calificaciones, percentiles, estaninas, equivalentes de grado y equivalentes de la curva normal.</li> <li>• La escala o escalas en que se reportan las calificaciones pueden añadir claridad o confusión al reporte de resultados.</li> </ul>
<b>Referencia para interpretación</b>	<ul style="list-style-type: none"> <li>• Referidas a normas, referida a estándares (niveles de logro) o ambos.</li> </ul>
<b>Unidad de evaluación</b>	<ul style="list-style-type: none"> <li>• Reactivo, área (por ejemplo, subescala o dominio) toda la prueba.</li> <li>• Es común que se reporte la información a nivel de área pero eso tiene limitaciones técnicas.</li> </ul>
<b>Unidad de reporte</b>	<ul style="list-style-type: none"> <li>• Estudiante, aula, escuela, distrito, estado, nación.</li> <li>• Para cada unidad, métrica y combinación del nivel de la prueba.</li> <li>• Es común que se reporte la precisión con la que se miden las calificaciones del desempeño estándar de toda la prueba. Sin embargo, no siempre se presenta el nivel estándar de medición cuando se informa el nivel de logro a nivel del área. Deben reportarse los errores de medición para las interpretaciones sugeridas por un reporte de resultados. Las gráficas y tablas que propician la comparación del logro de los estudiantes en las áreas deben reportar los errores estándar de las diferencias.</li> </ul>
<b>Error de medición</b>	<ul style="list-style-type: none"> <li>• Numérico, gráfico, narrativo.</li> <li>• Los resultados de la prueba pueden presentarse de manera numérica, gráfica o en una narración descriptiva, el mejor método para diferentes destinatarios no siempre es el mismo y es común ver en los reportes de resultados el uso de múltiples modos de presentación que incluyen cierta redundancia.</li> </ul>
<b>Modo de presentación</b>	<ul style="list-style-type: none"> <li>• Impreso, basado en la red (estático o interactivo).</li> </ul>
<b>Medio del reporte</b>	<ul style="list-style-type: none"> <li>• Impreso, basado en la red (estático o interactivo).</li> </ul>

Tomado de Ryan, J. (2012). Prácticas, problemas y tendencias en el reporte de resultados del estudiante, En Downing S. M. & Haladyna T.M. (Eds.). *Manual para el desarrollo de pruebas a gran escala: Vol. II.* (pp. 1005-1050). (Centro Nacional de Evaluación para la Educación Superior Trad.). D.F., México.: Centro Nacional de Evaluación para la Educación Superior. (Trabajo original publicado en 2006).

Aunque los elementos mostrados en la tabla no son exhaustivos, muestran los componentes de un reporte y las diferentes maneras en las que se pueden presentar. Si analizamos un reporte de resultados con esta tabla, es posible crear la ficha de datos de una prueba (como lo haremos más adelante en un caso aplicado) que permite verificar en un solo cuadro los componentes de un reporte de resultados.

### **Aspectos que hacen poco efectivo a un reporte de resultados**

Goodman y Hambleton encontraron, después de una detallada revisión de reportes nacionales, algunas fallas en la construcción de reportes de resultados:

- Dan demasiada información, provocando así que los lectores no encuentren lo que esperan extraer.
- Tienen un inapropiado nivel de conceptos estadísticos de alto nivel, incluso para audiencias bien educadas en esas temáticas.
- Proporcionan poca información con respecto al propósito de la evaluación.
- Brindan poca información sobre el uso de los resultados.
- Exponen poca información acerca de la precisión de los puntajes, provocando que los resultados parezcan más precisos de lo que son.
- Tienen imprecisiones de definición en términos clave, incluyendo los niveles de desempeño crítico, dejando su interpretación a elección de los usuarios, cuando muchos desconocen las interpretaciones adecuadas.
- Dan mucha información en un espacio pequeño, lo cual genera una percepción de información densa y desordenada.

- La utilización de jerga estadística confunde e intimida a algunos usuarios.
- Los símbolos y pies de páginas técnicas son mal interpretados o ignorados por los usuarios.
- Problemas relacionados con el diseño general, como el uso de fuentes tipográficas pequeñas en varios reportes, abreviaciones y gráficos que no incluyen escalas.

### **Elementos de un buen reporte de resultados acorde con investigaciones**

Goodman y Hambleton (2004) elaboraron varias recomendaciones de lo que resulta deseable en los reportes de instrumentos de aplicación a gran escala:

- Ser claros y concisos.
- Ser visualmente atractivos.
- Incluir texto fácil de leer, que apoye y mejore las interpretaciones de gráficos y tablas.
- Cuidar que, al mostrar datos, la presentación de éstos se haga con el fin de satisfacer un pequeño número de propósitos preestablecidos.
- Utilizar los apoyos visuales como cuadros y gráficos para destacar los principales hallazgos.
- Agrupar los datos en formas significativas.
- Evitar el uso de letras pequeñas y pies de página.
- Evitar la utilización de jerga estadística.
- Definir los términos claves, preferentemente en un glosario.
- Pilotear los reportes con miembros de la audiencia de interés.
- Considerar la creación de reportes diseñados para cumplir con necesidades particulares de diferentes usuarios.

Además, estos autores elaboraron recomendaciones para reportar resultados de calificación para estudiantes:

- Incluir la información esencial para una adecuada interpretación de los resultados (por ejemplo, las declaraciones explican el propósito de la evaluación, el significado de los niveles de desempeño y puntajes, cómo los resultados deben ser usados, además de la forma en que se deben interpretar los intervalos de confianza). Considerar la creación de reportes más grandes que puedan incluir esta información (un cartel de 11" x 17" es una buena medida).
- Contener información detallada sobre la evaluación y los resultados de los puntajes en una guía de interpretación separada, idealmente una en la que se pueda insertar el reporte del puntaje del estudiante.
- Personalizar los reportes del estudiante e interpretar las guías.
- Incluir un resumen narrativo de fácil lectura de los resultados del estudiante, al inicio del reporte del estudiante.
- Identificar las cosas que los padres pueden hacer para ayudar a mejorar a sus hijos. Idealmente estas sugerencias deben anexarse de manera separada, cerca del final del reporte de calificaciones y adaptarse al desempeño del estudiante. Recomendar a los padres o tutores hablar con sus hijos acerca de las maneras de mejorar el desempeño.
- Incluir ítems de muestra en las guías de apoyo que ejemplifiquen los tipos de desarrollo representado en cada nivel de desempeño.
- Contener una reproducción del reporte para el estudiante en las guías de apoyo para explicar claramente los diferentes elementos del reporte.

Hambleton y Slater entrevistaron a personajes involucrados en política, autoridades educativas y reporteros especializados en el tema, acerca de su

comprensión de los elementos contenidos en reportes de desempeño estudiantil de la *National Assessment of Educational Progress* (NAEP); y realiza un monitoreo a través del tiempo. Las respuestas permitieron saber que los entrevistados tenían una adecuada comprensión textual de los reportes, no obstante, el lenguaje estadístico les parecía confuso o incluso intimidante, además de la falta de claridad en los símbolos estadísticos, pues no eran comprensibles las tablas ni los gráficos (Hambleton y Slater, 1997). En ese tenor, sus recomendaciones para hacer mejores reportes puntualizan:

- Los gráficos, figuras y tablas deben ser comprensibles sin referencia al texto (al parecer, los lectores no ven la información para buscar interpretaciones).
- Se deben analizar con grupos de enfoque representativos de las audiencias las figuras, tablas y gráficos; muchos aspectos interesantes pueden ser aprendidos, como formatos que resultan confusos para los lectores.
- Hay que asegurar que los gráficos, figuras y tablas pueden ser impresos, reproducidos o modificados sin perder calidad visual (algunos resultados interesantes pueden ser distribuidos y el lector puede estar forzado a mirar malas copias).
- Los gráficos, las figuras y las tablas deben ser relativamente sencillas para minimizar la confusión y disminuir el tiempo en que los lectores identifican las principales tendencias de los datos.
- Con respecto a los reportes ejecutivos de la NAEP, se considera necesario proporcionar una introducción de la institución y sus escalas, incluir un glosario, enfatizar poco la jerga estadística, simplificar tablas, gráficos y figuras, así como usar más elementos para ilustrar los principales hallazgos.
- Se puede dar el caso de necesitar reportes diseñados especialmente para cada audiencia y si existen lectores interesados, deben ser expuestos a otros documentos para encontrar información adicional. Por ejemplo, para las autoridades educativas los reportes deben ser breves, usando marcadores que señalen los principales hallazgos o conclusiones, con tablas sencillas y

enfocadas en las implicaciones importantes; los datos técnicos (como el error estándar) y las discusiones técnicas, junto con detalles metodológicos, deben ser evitados para no entorpecer la comprensión; también es fundamental mantener el enfoque en las conclusiones y significados, para garantizar el reporte corto.

En una temprana investigación ya se mencionaban siete elementos generales y varios particulares que tenían como objetivo guiar la elaboración de reportes (Aschbacher y Herman, 1991). Por otra parte, Forte Fast y la *Accountability Systems and Reporting State Collaborative on Assessment and Student Standards* (ASR SCASS) elaboraron un documento de apoyo para las agencias estatales en el diseño de reportes eficientes para una adecuada rendición de cuentas (Forte y ASR SCASS, 2002).

Zeninsky y Hambleton revisaron los documentos que hacen referencia a las prácticas de reportar e identificaron cinco procesos para analizar los reportes así como los referentes en la literatura de éstos (Zeninsky y Hambleton, 2012). Los procesos y las consideraciones son:

1. Proceso para desarrollar el reporte:
  - Considerar la variedad de audiencias.
  - Probar el reporte con los individuos o en grupos de enfoque.
2. Diseñar y planear el formato del reporte:
  - Proporcionar la información de forma clara y sencilla.
  - Personalizar para conectar al usuario con el reporte.
  - Considerar a detalle el lenguaje utilizado (claro, conciso, simplificado y llano, que evite términos estadísticos o técnicos complejos).
  - Graficar, principalmente en formatos únicos con información comparativa.
3. Desarrollar los contenidos:

- Tener toda la información esencial para interpretar el reporte en su conjunto; incluir el propósito de la prueba, uso y rangos de confiabilidad.
  - Debe ser procesable por el lector, con información del estado actual del individuo y sugerencias sencillas que ayuden a mejorar el desempeño de los usuarios.
  - Contener resultados alineados de manera notoria y explícita a las categorías que reporte la prueba.
  - Reportar los resultados con detalle y de acuerdo con el nivel de confiabilidad posible.
4. Contener materiales auxiliares:
- Ejemplos detallados de los distintos apartados de un reporte.
  - Instrucciones para comprender y utilizar apropiadamente los resultados.
  - Muestras de ítems para ayudar a guiar las interpretaciones de los puntajes.
5. Difusión de los resultados:
- Deben proporcionarse en un tiempo satisfactorio.
  - Las páginas en internet donde se reporten los resultados deben soportar cualquier cantidad de visitas por parte de los usuarios.
  - Considerar las dificultades tecnológicas que implica hacer reportes en otro lenguaje y la posibilidad de que no estén disponibles en línea.

Como se aprecia, las consideraciones para crear reportes suelen estar construidas como listas, procesos o temáticas particulares, útiles para los expertos en evaluación con el objetivo de mejorarlos, pero no son una guía detallada de producción, lo cual entraña un trabajo más detallado y descriptivo. Es relevante comentar que varias consideraciones están relacionadas entre sí y comparten características, lo que hace que sean semejantes, esto en sí no es un problema, sino un referente que nos indica que hay consensos en los mismos tópicos.

## Cómo generar un reporte de resultados de una prueba

No hay un modelo único o general para crear reportes, aunque existen diversas propuestas para su creación con base en distintos enfoques y niveles de precisión, desde esquemas con pautas generales (Zenisky y Hambleton, 2012), descripciones de las técnicas de diseño y claves visuales (Gibbons, 1992) hasta los modelos de medición del desempeño para generar reportes más comprensibles o eficientes, como el Modelo de Atribución Jerárquica (Gierl, Wang y Zhou, 2008; Leighton, Gierl y Hunka, 2002). Todos tienen una noción en común: los reportes pueden ser eficaces y de calidad.

Zeninsky definió tres etapas y ocho pasos para elaborar reportes de resultados, mismos que se apuntan en la tabla 6.

**Tabla 6**  
**Etapas y pasos para elaborar reportes**

<b>Etapas</b>	<b>Pasos</b>
Preparación inicial	<ul style="list-style-type: none"><li>• Definir el propósito del reporte.</li><li>• Elaborar las necesidades de la medición.</li><li>• Identificar la audiencia de intención.</li></ul>
Desarrollo del reporte	<ul style="list-style-type: none"><li>• Desarrollar los reportes prototipos o preliminares.</li></ul>
Del reporte y revisión	<ul style="list-style-type: none"><li>• Comprobar empíricamente.</li><li>• Obtener información.</li><li>• Revisar y rediseñar.</li><li>• Repetir los ciclos cuantas veces sea necesario</li></ul>

Mantenimiento constante

---

*Fuente:* Tomado de Zenisky, A. L. & Hambleton, R. K. (2012). Developing test score reports that work: the process and best practices for effective communication. *Educational measurement: issues and practice*, 31 (2), 21-26.

## **Estructurando información en un reporte**

Otra manera de estructurar la información de un reporte es por medio de la consideración de éste como información presentada en mayor medida como un texto. Roberts y Gierl (2010) realizaron un esquema para el diseño del escrito en los reportes de resultados y consideraron que la información textual puede analizarse desde su estructura interna (redacción y presentación), o desde su estructura externa (el modo en que los componentes no textuales auxilian. color, tipografía, presentación de la información).

Una explicación de las estructuras se registra a continuación:

- Estructura interna. Son técnicas para organizar, secuenciar y proporcionar un esquema interno de contenido documental comprensible, tomando en cuenta el buen uso de los párrafos: iniciar con frases que incluyan un nuevo concepto; finalizar con otras que señalen su fin y un posible traslado a otro contenido.
  
- Estructura externa con acceso:
  - Global: Que proporcione una generalidad del contenido y ayude al lector a desarrollar una estrategia de lectura como las tablas de contenido, glosarios, objetivos y resúmenes.
  - Local: Con técnicas que señalen o identifiquen unidades particulares del texto, como encabezados, sistemas numéricos o listas precisas, específicas y concretas.
  
- Tipografía. Es la letra, número o cualquier otro carácter susceptible de impresión. La legibilidad es la consideración primaria; incluye claves como itálicas, negritas, subrayados, letras mayúsculas o colores.

- Disposición y organización del documento. Es la combinación del tipo, color y técnicas de organización espacial que estructuran de manera eficiente un documento. El uso de información en vertical u horizontal se determina por la intención e impacto deseado.
- Diseño efectivo de la información cuantitativa. Los puntajes pueden ser narrados o visualizados en una tabla, aunque los principios son similares a los de la organización de un texto: usar contrastes para señalar información relevante e incrementar legibilidad; utilizar redundancia visual para enfatizar información; usar proximidad para agrupar elementos similares; emplear una alineación similar para enfatizar la estructura visual de la organización.
- Elegir un formato para mostrar información. Cada uno con reglas de diseño:
  - Tablas. Recomendados para datos pequeños y breves que expresan números enteros, los cuales requieren comparación local.
  - Gráficos. Son preferibles cuando se hacen comparaciones de datos, comunican cantidades, cambios o tendencias con precisión y legibilidad. Su formato debe ser compatible con el propósito y forma intencional.

El recuento anterior permite entender el modo en que se han analizado los reportes de resultados de las pruebas, los elementos que debe contener un reporte eficiente y aquellos que no son deseables, así como la manera de considerar su diseño.

## **Metodología**

### **Tipo de investigación**

El proyecto de investigación es de tipo exploratorio porque el tema del análisis de los reportes de resultados de pruebas estandarizadas es nuevo en español, y como ya se ha mencionado, existe poca información al respecto. También es descriptiva, porque se describe y analiza ampliamente el tema de estudio.

## **Enfoque de la investigación**

La investigación presenta un enfoque cualitativo porque se evaluó la forma en la que se analizan los reportes de resultados de pruebas estandarizadas por medio de la revisión de distintas fuentes que hacen referencia a dichos reportes.

## **Población o universo de estudio**

El universo que se aborda en el estudio son los documentos que hacen referencia a los reportes de resultados de pruebas estandarizadas.

## **Unidad de análisis**

Esta investigación centró el análisis en documentos internacionales que norman la administración, uso y evaluación de pruebas estandarizadas, en investigaciones que analizaron la calidad de varios reportes y en recomendaciones de agencias internacionales y de expertos sobre cómo crear reportes de resultados o cómo crear documentos que comuniquen efectivamente sus propósitos a diferentes audiencias.

## **Técnicas de recolección de información**

La información bibliográfica para la investigación se obtuvo de un análisis documental.

## **Fuentes de información**

Las fuentes que se usaron como información para esta investigación permitieron elaborar un sustento teórico y metodológico que hicieron posible crear un conocimiento más amplio y desarrollado para el tema de los reportes de pruebas estandarizadas.

Manuales de procedimientos

Informes de manuales

Libros

Manuales

Investigaciones en revistas especializadas

## **CAPÍTULO IV. LA PROPUESTA DE TRABAJO**

En este capítulo se presenta la propuesta para revisar el reporte de resultados de una prueba estandarizada, con base en los criterios mencionados en los capítulos anteriores.

Tal como se ha visto hasta ahora, existen estándares e investigaciones sobre los contenidos esperados de un reporte de resultados, sin embargo, no hay protocolos formales que documenten el método para diseñarlo o para determinar si uno ya elaborado cumple con los elementos esperados. No es una sorpresa si consideramos que en México la cantidad de pruebas estandarizadas es poca y, además, que no son socializadas de manera adecuada. Así, es comprensible que la ocupación sobre la calidad de un reporte de resultados sea nula.

El diagnóstico adecuado de un reporte de resultados permite verificar la manera en la que se concibe la prueba, su alcance, intención y el cuidado general en su diseño. De este modo, el reporte es un punto culminante de la prueba como producto susceptible de análisis y mejora.

Esta propuesta tiene la intención de formar un bagaje en el lector con respecto a:

1. Los componentes que se esperan en un reporte de resultados de una prueba estandarizada.
2. Los estándares internacionales para un reporte de resultados en función de sus componentes.
3. Las recomendaciones de expertos para generar reportes de resultados eficaces.
4. La unión de los elementos anteriores con el fin de conformar un documento que sirva para diagnosticar un reporte de resultados.

La propuesta tiene cuatro pasos:

1. Elaborar la ficha de componentes del reporte de resultados de una prueba estandarizada.
2. Determinar los estándares pertinentes para la entrega de resultados.
3. Reconocer las consideraciones generales y particulares para cada elemento del reporte de una prueba estandarizada.
4. Elaborar las recomendaciones con el objetivo de verificar un reporte de resultados de una prueba estandarizada y determinar las mejoras.

En conjunto, una vez alcanzados los tres primeros pasos, se tiene conocimiento sobre lo que debe contener un reporte; el último es la recomendación que surge como producto de los pasos anteriores.

A continuación se describe, punto por punto, cada uno de los pasos.

## **Primer paso**

### **Elaborar la ficha de componentes de un reporte de resultados**

Se construye una tabla de datos con los componentes esperados de cualquier prueba estandarizada; esta tabla, al cruzarla con los datos reales, genera una ficha. Si algún elemento que se solicite en la tabla no está presente en una prueba, es necesario aclarar el por qué de su ausencia. En la tabla 7 se despliegan los componentes con sus diferentes posibilidades de expresión.

**Tabla 7**  
**Tabla del reporte de resultados con sus componentes y sus posibilidades**

<b>Componentes del reporte</b>	<b>Posibilidades</b>
<b>Objetivo de la prueba</b>	Generales y particulares para cada audiencia
<b>Destinatarios</b>	Estudiantes, padres de estudiantes, maestros, directores o funcionarios estatales.
<b>Uso de la información</b>	Se describirá lo que se pretende que hagan o consigan los usuarios con esa información.
<b>Referente de interpretación</b>	Referidas a normas, referida a estándares (niveles de desempeño) o ambos.
<b>Escala o métrica de medición</b>	Calificación cruda, porcentaje de aciertos, escala de calificaciones, percentiles, estatinas, equivalentes de grado y equivalentes de la curva normal.
<b>Modo de presentación de los puntajes</b>	Numérico, gráfico o narrativo.
<b>Nivel de agrupación para reportar ítems</b>	Reactivo, área, subescala o dominio; toda la prueba.
<b>Unidad de entrega de datos</b>	Estudiante, aula, escuela, distrito, estado, nación.
<b>Error de medición</b>	Para cada unidad, métrica y combinación del nivel de la prueba.
<b>Medio de entrega</b>	Impreso, basado en la red (estático) o basado en la red (interactivo).
<b>Documentos anexos</b>	Guías para usuarios, manuales, documentos técnicos o vínculos a instituciones (en la red).

La información contenida en esta tabla se convierte en la ficha de datos de la prueba; para ello se reemplaza la columna de posibilidad por la correspondiente de una prueba. Si una prueba tiene varios destinatarios se debe crear una ficha para cada uno.

Los datos de la ficha brindan información general de un reporte y si se traslada a preguntas, permite tener información bajo las consideraciones que se contemplan a continuación:

- ¿Cuál es el mensaje emitido en el reporte?
- ¿Cuáles son las audiencias meta del reporte?
- ¿Cuál es el uso esperado de los reportes por parte de las audiencias meta?
- ¿Qué indicadores se entregan de cada componente para las distintas audiencias?
- ¿Cuál es el mensaje general y cuáles los particulares?, ¿cómo se representan en el reporte?
- ¿Hay un balance entre los componentes del reporte y su diseño conceptual, es decir, el constructo con el que se concibió la prueba?
- ¿Hay alguna información adicional que deba ser incluida para facilitar las interpretaciones?
- ¿Qué documento es necesario vincular al reporte para proporcionar un contexto adecuado a las interpretaciones?
- ¿Es importante revisar algún otro documento para asegurar la cohesión de todo el sistema de reportes?

Este breve cuestionario, aunado a la tabla, permite abordar de forma inicial el modo en que se elaboró un reporte.

## **Segundo paso**

### **Determinar los estándares sobre la entrega de resultados y el uso de una prueba estandarizada**

Los estándares sobre el contenido de un reporte de resultados provienen de:

- Marco o regulación legal nacional.
- Gestores de la prueba, entendidos como los clientes.
- Documentos normativos o estándares internacionales.
- Documentos normativos de la institución que diseña la prueba.

Cada estándar establece algún criterio que puede ser general o aplicable a casos específicos. Con este fin, se utiliza el formato expuesto en la tabla 8 y se verifican los estándares o códigos que corresponden en los anexos A y B; las categorías en que se agrupan dichos estándares también se muestran en ella.

**Tabla 8**  
**Consideraciones para diseñar, elaborar o evaluar un reporte de resultados acorde con los documentos: *Standards for Educational and Psychological Testing* o *Code of Fair Testing Practices in Education*.**

Consideración	Estándar o Código
Evitar malas interpretaciones de los resultados	Standard 15.11
	Code C2
	Code C3
	Code C6
Informar de las interpretaciones adecuadas	Standard 11.20
	Standard 13.13
	Code C1
	Code C8
Evitar malas interpretaciones de los puntajes	Standard 11.15
Asegurar interpretaciones adecuadas de los puntajes	Standard 5.10
	Standard 12.9
	Code B6
Presentar puntajes en escenarios educativos	Standard 13.14
	Standard 13.16
Presentar puntajes acumulados, desagregados o en categorías	Standard 7.8
	Standard 8.8
	Standard 12.19
	Standard 13.19
	Code C4
Entregar resultados en formatos o medios electrónicos	Standard 5.11
	Standard 6.12
	Standard 12.15
Entregar resultados en grupos diferenciados	Standard 5.12
Garantizar la confidencialidad de los resultados	Standard 5.13
	Code B.7
Entregar resultados dependiendo del alto o bajo impacto de la prueba	Standard 7.10
	Standard 13.5
Considerar el tiempo de entrega de resultados	Standard 11.6
	Standard 11.17
	Code C.7

*Fuente:* elaboración propia.

Es importante reconocer que los estándares no son procedimientos para hacer los reportes, sino descripciones de calidad y, en este sentido, se usan como referentes o guías. Siendo así, los utilizamos como referentes antes de la creación de un reporte y como elementos a cotejar, una vez realizado éste.

El documento de los Estándares para la Evaluación de Programas proporciona dos recomendaciones muy representativas de los reportes:

**U5 B.** Hacer que los reportes sean tan breves, sencillos y directos como sea posible; centrados en dar respuesta a las preguntas de evaluación.

**U5 F.** Utilizar el lenguaje técnico en un modo limitado y ayudar a los interesados a comprender términos técnicos o poco familiares (validez interna y externa, fiabilidad, desviación estándar, análisis de categorías, diseños de muestra), suministrando glosarios, resumen separado e informes técnicos, ofreciendo también oportunidades para la formación.

### **Tercer paso**

#### **Reconocer consideraciones generales y particulares de cada elemento del reporte de una prueba estandarizada, acorde con los expertos**

En primer lugar, se deben conocer las recomendaciones generales del reporte:

- Contener toda la información esencial para interpretarlo, incluyendo el propósito de la prueba, uso y rangos de confiabilidad.
- Ser claro y conciso.
- Ser visualmente atractivo.
- Evitar uso de letras pequeñas y pies de página.

- Eludir la utilización de la jerga estadística.
- Los términos clave deben definirse, preferentemente en un glosario.

A continuación se usa la información de la tabla 7 del primer paso, que se complementa con recomendaciones para cada componente, y da como resultado las siguientes recomendaciones para cada componente del reporte

#### Objetivo de la prueba

- Proporcionar información clara y sencilla.
- Considerar el lenguaje de forma minuciosa.
- Personalizar con el fin de ayudar a conectar al usuario con el reporte.
- Incluir un resumen narrativo de fácil lectura con los resultados del estudiante al inicio del reporte.

#### Destinatarios

- Considerar la variedad de audiencias y probar el reporte, ya sea en individuos o grupos de enfoque.
- Personalizar con el fin de ayudar a conectar al usuario con el reporte.
- Considerar la creación de reportes diseñados especialmente con el objetivo de cumplir con las necesidades particulares de diferentes usuarios.
- Identificar las acciones de mejora que deben implementar los padres en beneficio de sus hijos, así como recomendar el diálogo sobre los medios para optimizar su desempeño. Idealmente, estas sugerencias deben incluirse por separado cerca del final del reporte de calificaciones y adaptarlas al desempeño del estudiante.
- Proporcionar una introducción de la institución que elabora la prueba y sus escalas, incluir un glosario, enfatizar poco la jerga estadística, simplificar tablas, gráficos y figuras favoreciendo el uso de cajas y esquemas para ilustrar los principales hallazgos.

- Con las autoridades educativas, los reportes deben ser breves mediante el uso de marcadores para señalar los principales hallazgos o conclusiones; las tablas deben ser sencillas y enfocadas a las conclusiones e implicaciones importantes. Los datos técnicos (como el error estándar) y discusiones técnicas, a la par de detalles metodológicos, deben ser evitados.
- Mantener el enfoque en las conclusiones y significados; sostener el reporte corto.

#### Uso de la información

- Proporcionar información clara y sencilla.
- Considerar el lenguaje de forma minuciosa.

#### Escala o métrica de medición

- Graficar esencialmente en formatos únicos y proporcionar información comparativa.
- Ofrecer información digerible del estado actual de un individuo con sugerencias sencillas que ayuden a mejorar el desempeño de los usuarios.

#### Modo de presentación de los puntajes

- Analizar en grupos de enfoque de las audiencias las figuras, tablas y gráficos.
- Graficar esencialmente en formatos únicos.
- Incluir textos de fácil lectura como apoyo a las interpretaciones de gráficos y tablas.
- No mostrar demasiados datos; la presentación debe hacerse para satisfacer un pequeño número de propósitos preestablecidos.
- Los apoyos visuales, como cuadros y gráficos, deben destacar los principales hallazgos.
- Los datos se agrupan en formas significativas.

- Los gráficos, figuras y tablas deben ser comprensibles sin referencias al texto (al parecer, los lectores no ven la información para buscar interpretaciones).
- Corroborar que las gráficas, figuras y tablas puedan ser impresas, reproducidas y modificadas sin perder calidad visual (algunos resultados interesantes pueden ser copiados y el lector estaría forzado a mirar malas copias en cada ocasión).
- Las gráficas, figuras y tablas deben ser relativamente sencillas para minimizar confusión y disminuir el tiempo en que los lectores identifican las principales tendencias de los datos.
- Las tablas se recomiendan para datos pequeños y breves, los cuales expresan números enteros que requieren comparación local.
- Las gráficas son preferibles cuando se hacen comparaciones de datos, pues comunican cantidades, cambios y tendencias de forma precisa y legible. Su formato debe ser compatible con el propósito y forma intencional.

#### Unidad de entrega de datos

- Proporcionar resultados alineados manifiesta y explícitamente a las categorías que reporte la prueba.

#### Error de medición

- Graficar esencialmente en formatos únicos.
- Proporcionar resultados de manera detallada, acorde con el nivel de confiabilidad posible.

#### Medio de entrega

- Proporcionar datos en tiempo apropiado.
- Las páginas de internet donde se reporten resultados deben soportar cualquier cantidad de visitas de los usuarios.

- Considerar las dificultades tecnológicas que implica hacer reportes en otro lenguaje, aunado a la posibilidad de que los reportes no estén disponibles en línea.

#### Documentos anexos

- Facilitar ejemplos detallados de las diferentes partes de un reporte.
- Proporcionar instrucciones para leer y usar los reportes.
- Incluir muestras de los ítems que ayudan a guiar las interpretaciones de los puntajes.
- Contener información detallada sobre la evaluación y los resultados de los puntajes en una guía de interpretación separada; idealmente, una en la que se pueda insertar el reporte del puntaje del estudiante.
- Insertar ítems de muestra en las guías de apoyo que ejemplifiquen los tipos de desarrollo representado en cada nivel de desempeño.
- Añadir una reproducción del reporte para el estudiante en las guías de apoyo con el fin de explicar claramente los diversos elementos del reporte.
- Los lectores interesados deben ser remitidos a otros documentos que les proporcionen información adicional.

Otra perspectiva concibe al reporte, ante todo, como información conformada por texto y propone las siguientes consideraciones:

- Calidad de la estructura de acceso global. El reporte, en su conjunto, proporciona una generalidad de contenido y ayuda al lector a desarrollar una estrategia de lectura basada en tablas de contenido, glosarios, objetivos y resúmenes.
- Calidad de la estructura de acceso local. Es el adecuado uso de técnicas que señalan o identifican unidades particulares del texto como encabezados, sistemas numéricos o listas que deben ser precisas, específicas y concretas.

- Tipografía. Entendida como una letra, número o cualquier otro carácter susceptible de impresión. La legibilidad es la consideración primaria, incluye el uso de itálicas, negritas, subrayados, letras mayúsculas o colores.
- Disposición y organización del documento. Es la combinación del tipo, color y técnicas de organización espacial que estructuran de manera eficiente un documento. El uso de información en vertical u horizontal se determina por la intención e impacto deseado.
- Diseño efectivo de la información cuantitativa. Los puntajes pueden ser narrados o visualizados en una tabla; los principios son similares a los de la organización de un texto: usar contrastes para señalar información relevante e incrementar legibilidad, utilizar redundancia visual para enfatizar información, manipular la proximidad para agrupar elementos similares y emplear una alineación similar para enfatizar la estructura visual de la organización.

#### **Cuarto paso**

##### **Recomendaciones para verificar el reporte de resultados de una prueba estandarizada y determinar las mejoras**

Se precisan los tres pasos anteriores, para realizar las recomendaciones que permitan verificar un reporte de resultados de una prueba estandarizada que reúna:

- Ficha descriptiva del reporte de resultados con sus características.
- Indicadores nacionales o internacionales en donde se especifique qué se espera del reporte en diferentes componentes.
- Conocimiento de recomendaciones detalladas para cada característica del reporte.
- Conocimiento de recomendaciones generales para elaborar un reporte.

Es en este punto que se tienen todos los elementos de un buen reporte de resultados contenidos en un documento completo que puede ser usado como referencia de un reporte de resultados ya elaborado. De este modo, lo que resulta del contraste entre lo que se espera y lo que existe en un reporte, es el diagnóstico del reporte de la prueba.

Con el diagnóstico del reporte de resultados se producen sugerencias de modificación y propuestas. Así, se elaboran tantas propuestas de modificación como se consideren necesarias y se realiza la tabla 9.

**Tabla 9**

**Ejemplo del uso de las fases para analizar los componentes del reporte de resultados de una prueba estandarizada**

**Lo que enuncia la prueba X**

- Que las autoridades educativas conozcan el nivel de conocimientos y habilidades de los estudiantes en el área de matemáticas a nivel de escuela, entidad y estado.
- Que las autoridades educativas tengan información de apoyo para las decisiones sobre el nivel de conocimientos y habilidades matemáticas en el que se encuentran los alumnos al nivel de escuela, región y estado.
- Proporcionar a las autoridades educativas información del nivel de conocimientos y habilidades matemáticas a lo largo de un periodo anual, de manera que puedan tener información de mediano y largo plazo acerca del desempeño en dicha área a nivel de escuela, región y estado.

**Lo que mencionan los indicadores internacionales**

- **Estándar 11.15** Quienes administran la prueba deben ser advertidos sobre potenciales malas interpretaciones y posibles consecuencias no intencionales de la prueba; los usuarios deben prever acciones que minimicen o eviten malas interpretaciones previsibles o consecuencias negativas no intencionales.
- **C6.** Proporcione información que permita a los usuarios interpretar y reportar los resultados de los subgrupos de forma precisa, incluyendo información de quién, cuándo y dónde no se incluyen en los diferentes grupos, así como los factores que pueden influir en la interpretación de los resultados.

**Lo que mencionan las consideraciones de expertos**

- Proporcionar la información de manera clara y sencilla.
- Considerar el lenguaje de forma minuciosa.

**Recomendación para la prueba X**

- Proporcionar información de las limitantes de los datos.
- Aportar información sobre una mala interpretación del uso de la prueba.

**Sugerencia de redacción para la prueba X**

Añadir a la sección sobre los usos de la prueba sentencias como las siguientes:

- Los resultados no permiten hacer comparaciones entre escuelas o estados.
- Los resultados no pueden usarse como evidencia de todo el conocimiento de los alumnos en matemáticas, solo comprende ámbitos aprendidos dentro del aula.

## **CAPÍTULO V. APLICACIÓN DE LA PROPUESTA A UNA PRUEBA ESTANDARIZADA**

Se muestra en este capítulo cómo se aplica a una prueba estandarizada real la propuesta de trabajo mostrada anteriormente.

## **La prueba enlace EMS**

### **El antecedente: porqué se decide crear una prueba de gran escala a nivel nacional**

En el año de 2007 el gobierno federal mexicano redactó un documento nominado *Programa Sectorial 2007-2012* con la intención de crear un modelo de políticas educativas acorde con nuevas expectativas sobre el sistema educativo; el documento enfatiza dos criterios: minimizar el rezago educativo y formar ciudadanos. Y para esto enuncia seis objetivos. De ellos, el primero tiene referencias directas de evaluación de alumnos. Se establece que, entre otros objetivos, la evaluación será un instrumento para analizar la calidad, para la relevancia y la pertinencia del diseño y la operación de las políticas públicas en materia de educación, donde la evaluación atenderá tres dimensiones:

- Para rendición de cuentas
- Como instrumento de difusión de resultados a padres de familia
- Como sustento del diseño de políticas públicas

El documento menciona que “los indicadores utilizados para evaluar deberán ser pertinentes y redundar en propuestas de mejora continua”. (p. 4), y proporciona la descripción de doce indicadores que tomando como inicio de referente datos de 2006 pretende que aumenten o mejoren para el 2012; en dos de ellos hacen referencia a las pruebas: la internacional de PISA donde a partir de los resultados de 2003 se tenía una media de 392 puntos y propone como meta en el año 2012, 435 puntos; y de la prueba Enlace para los niveles de primaria y secundaria,

donde en secundaria se pretende tener el nivel de “elemental” (hay cuatro niveles: insuficiente, elemental, bueno y excelente) para todos los alumnos egresados.

El documento también hace referencia a la creación de un sistema único de nivel bachillerato, expresado en el apartado 1.6. (las negritas son nuestras en todos los casos).

1.6 Alcanzar los acuerdos necesarios entre los distintos subsistemas y con instituciones de educación superior que operen servicios de educación media superior en el ámbito nacional, con la finalidad de **integrar un sistema nacional de bachillerato** en un marco de respeto a la diversidad de modelos, que permita dar pertinencia y relevancia a estos estudios, así como lograr el libre tránsito de los estudiantes entre subsistemas y contar con una certificación nacional de educación media superior.

- Homologar elementos comunes a las diferentes modalidades y subsistemas, estableciendo un **marco curricular común** que garantice que los alumnos cuenten con las competencias básicas y capacidades requeridas en este nivel que les permita transitar de una modalidad a otra. (p.13).

Posteriormente se describen los elementos que se consideran necesarios que tenga el alumno al final de su formación de bachillerato. Por último, se describe un apartado denominado Temas Transversales siendo la evaluación uno de ellos, donde, respecto a la evaluación de alumnos de educación media superior se enuncia:

E.7 Diseñar, aplicar y consolidar los instrumentos de evaluación del desempeño académico de los estudiantes como insumo para reorientar la acción pedagógica y para integrar los programas de nivelación académica.

- Aplicar la Evaluación Nacional de Logro Académico en la Educación Media Superior (ENLACE-EMS), cuyo propósito es determinar en **qué medida los jóvenes son capaces de aplicar los conocimientos y habilidades que les permitan hacer un uso apropiado del idioma –comprensión lectora– y las matemáticas, así como difundir sus resultados.** (p.39).

Así, el origen de la prueba está bajo una necesidad política concreta que determina su intención.

Hay dos momentos relevantes de la prueba, desde el 2007 hasta el 2010 donde se conformó la prueba con base en la recomendación de comités que revisaron los marcos teóricos de pruebas que miden componentes similares como las de PISA, los materiales disponibles de la encuesta ALL (*International Adult Literacy Survey*), los procedimientos de Cito (*Cito Instituut voor Toetsontwikkeling*, conocido en inglés como *Central Institute for Test Development*) y algunas aproximaciones a la evaluación de habilidades que han resultado exitosas en el contexto latinoamericano (SIMCE y TIMSS). También la revisión de literatura y de los planes y programas de las diferentes modalidades y subsistemas de la EMS sirvió como referente para definir las dos habilidades e identificar los contenidos y los procesos por evaluar. Una de las concepciones teóricas generales que guió la elaboración del perfil corresponde con el enfoque de las pruebas PISA y SIMCE (p. 27) (Ceneval, 2010).

Este referente cambió en la prueba a partir de 2010, cuando se usó como referente *el ACUERDO número 444 por el que se establecen las competencias que constituyen el Marco Curricular Común (MCC) del Sistema Nacional de Bachillerato (Acuerdo 444)* que determinó puntos muy importantes para el sistema bachillerato y para la prueba como:

- La noción de un marco curricular común para todo el sistema de bachillerato.
- La noción de competencias como componente rector
- La distinción de competencias aplicables a todo el sistema y particulares de apoyo o especialización (llamadas extendidas).

La tabla 10 muestra los objetivos generales de las competencias.

**Tabla 10**  
**Características de las competencias del Marco Curricular Común**

Competencias	Objetivo
Genéricas	Comunes a todos los egresados de la Educación Media Superior. Son competencias clave, por su importancia y aplicaciones diversas a lo largo de la vida; transversales, por ser relevantes a todas las disciplinas y espacios curriculares de la Educación Media Superior, y transferibles, por reforzar la capacidad de los estudiantes de adquirir otras competencias.
Básicas	Comunes a todos los egresados de la Educación Media Superior. Representan la base común de la formación disciplinar en el marco del Sistema Nacional del Bachillerato.
Disciplinares	No serán compartidas por todos los egresados de la Educación Media Superior. Dan especificidad al modelo educativo de los distintos subsistemas de la Educación Media Superior. Son de mayor profundidad o amplitud que las competencias disciplinares básicas.
Profesionales	Proporcionan a los jóvenes formación elemental para el trabajo
Extendidas	Preparan a los jóvenes con una calificación de nivel técnico para incorporarse al ejercicio profesional.

*Fuente:* adaptado del Acuerdo número 444 por el que se establecen las competencias que constituyen el marco curricular común del Sistema Nacional de Bachillerato. (2008, martes 21 de octubre). México.: *Diario Oficial. (1a. Sección).*

La tabla 11 describe los componentes de la competencia disciplinar, que nos permite ubicar conceptualmente de dónde parte la prueba Enlace, en color gris.

**Tabla 11**  
**Componentes de la competencia disciplinar del Marco Curricular Común**

Competencias	Componentes
Disciplinares	Básicas Matemáticas con una disciplina y ocho competencias Ciencias experimentales con cuatro disciplinas y catorce competencias Ciencias sociales con cinco disciplinas y diez competencias Comunicación con cuatro disciplinas y doce competencias Orientaciones de contenido
	Extendidas 1. Las competencias disciplinares extendidas se construyen a partir de la lógica de las distintas disciplinas en las que tradicionalmente se ha organizado el saber 2. Las competencias disciplinares extendidas se organizan en campos disciplinares amplios 3. Las competencias disciplinares extendidas son específicas de distintos modelos educativos y subsistemas de la EMS 4. Las competencias disciplinares extendidas amplían y profundizan los alcances de las competencias disciplinares básicas 5. Las competencias disciplinares extendidas dan sustento a la formación de los estudiantes en el Perfil de Egreso del Sistema Nacional Bachiller

*Fuente:* adaptado del Acuerdo número 444 por el que se establecen las competencias que constituyen el marco curricular común del Sistema Nacional de Bachillerato. (2008, martes 21 de octubre). México.: *Diario Oficial. (1a. Sección).*

La prueba en su nuevo diseño implicó la revisión y análisis de las competencias en las disciplinas de matemáticas y de comunicación para obtener solo aquellas que podían ser evaluadas en una prueba diagnóstica, objetiva, estandarizada, de bajo impacto y con reactivos de opción múltiple (Ceneval, 2012, 2013). De esta forma sólo quedaron elementos susceptibles de medición con ese instrumento. En realidad no hubo cambios radicales en la estructura nominal de la prueba, pero sí en la manera de concebir los contenidos. Las tablas 12 y 13 muestran el resultado final de los contenidos de la prueba.

**Tabla 12**  
**Reactivos por tipo de texto y proceso de la prueba Enlace EMS desde el 2011 en el campo de comunicación**

Tipo de texto	Proceso			Total de reactivos
	Extracción	Interpretación	Reflexión y evaluación	
<b>Argumentativo</b>	4	5	6	<b>15</b>
<b>Narrativo</b>	3	7	5	<b>15</b>
<b>Expositivo</b>	5	7	3	<b>15</b>
<b>Apelativo</b>	2	2	1	<b>5</b>
<b>Reactivos por proceso</b>	<b>14</b>	<b>21</b>	<b>15</b>	<b>50</b>

*Fuente:* Adaptado del Centro Nacional de Evaluación para la Educación Superior (Ceneval). (2013). *Manual técnico ENLACE media superior 2011-2012*. México.: Autor. p.36.

**Tabla 13**  
**Reactivos por contenido y proceso de la prueba Enlace EMS desde el 2011 en el campo de matemáticas**

Contenido matemático	Proceso			Total de reactivos
	Reproducción	Conexión	Reflexión	
<b>Cantidad</b>	6	7	7	<b>20</b>
<b>Cambios y relaciones</b>	5	8	7	<b>20</b>
<b>Espacio y Forma</b>	6	8	6	<b>20</b>
<b>Reactivos por proceso</b>	<b>17</b>	<b>23</b>	<b>20</b>	<b>60</b>

*Fuente:* Adaptado del Centro Nacional de Evaluación para la Educación Superior (Ceneval). (2013). *Manual técnico ENLACE media superior 2011-2012*. México.: Autor. p.36.

Estos datos evidencian el origen de la prueba desde el ámbito legal y permiten reconocer el carácter político que se debe considerar al evaluar el uso de la prueba.

### **Estatus actual de la prueba**

Durante el 2013 hubo una serie de críticas hacia la prueba, expresadas en los medios y con manifestaciones públicas en México; esto provocó que el Secretario de Educación Pública, Emilio Chuayffet, afirmara en el mes de julio la desaparición de la prueba Enlace (Hernández, 2013), aunque no hizo referencia concreta a si dejaba de aplicarse la prueba de nivel básico o medio superior, o ambas.

En marzo de 2014, sin embargo, el Secretario Chuayffet anunció que la prueba sí se aplicará, pero en el 2015 sin los “esquemas y estímulos perversos que llevaron a los maestros a no asumir compromisos con las mejores enseñanzas” (Becerril, 2014) y que el INEE será el encargado de hacer las correcciones necesarias a la prueba.

### **Reconocimiento de los componentes de un reporte de resultados de la prueba Enlace EMS**

Los componentes de un reporte de resultados que se describieron en la fase 1 aplicada a los contenidos de la prueba Enlace EMS dan como resultado la ficha de las características de los reportes de resultados de la prueba.

**Tabla 14**

**Ficha de las características de los reportes de resultados de la prueba Enlace EMS**

<b>Características del reporte</b>	<b>Descripciones</b>
Objetivo de la prueba	<ul style="list-style-type: none"> <li>• Determinar el dominio que los sustentantes poseen sobre diferentes competencias básicas asociadas a los campos disciplinares de Comunicación (Comprensión Lectora) y Matemáticas.</li> </ul>
Destinatarios	<ul style="list-style-type: none"> <li>• Estudiantes.</li> <li>• Padres de familia.</li> <li>• Maestros.</li> <li>• Directores.</li> <li>• Funcionarios del estado (no es explícito sino a través de las modalidad de los reportes).</li> </ul>
Uso de la información	<ul style="list-style-type: none"> <li>• Generar información diagnóstica para cada alumno.</li> <li>• Proveer elementos que contribuyan a la mejora del sistema educativo.</li> <li>• Proveer información útil para el plantel, profesores y los padres de familia.</li> </ul>
Referente de interpretación	<ul style="list-style-type: none"> <li>• Referida a estándares expresada en niveles de dominio.</li> </ul>
Escala o métrica del reporte	<ul style="list-style-type: none"> <li>• Calificación cruda (acierto o error por cada pregunta en el reporte del alumno y de la escuela).</li> <li>• Porcentaje de aciertos (para todos los reportes).</li> <li>• No hay escalas en los reporte de resultados.</li> </ul>
Unidad de evaluación	<ul style="list-style-type: none"> <li>• Reactivo.</li> <li>• Nivel de dominio de cada habilidad matemática.</li> <li>• Nivel de dominio de cada habilidad lectora.</li> </ul>
Unidad de reporte	<ul style="list-style-type: none"> <li>• Estudiante.</li> <li>• Escuela por turno.</li> <li>• Entidad, (con dos opciones de desagregación, municipal y localidad).</li> <li>• Grado de marginación (muy alto, alto, medio, bajo, muy bajo).</li> <li>• Modalidad del bachillerato (general, tecnológico, técnico).</li> <li>• Tipo de sostenimiento (público o privado).</li> </ul>
Error de medición	<ul style="list-style-type: none"> <li>• Reportado solo en un Manual técnico y para una sola ocasión en toda la prueba.</li> <li>• No se reporta al nivel de reactivo ni por área para los usuarios.</li> </ul>
Modo de presentación	<ul style="list-style-type: none"> <li>• Numérico (expresado en porcentajes).</li> <li>• Gráfico (tablas).</li> <li>• Narrativo.</li> </ul>
Medio del reporte	<ul style="list-style-type: none"> <li>• Impreso.</li> <li>• basado en la red de manera estática.</li> </ul>
Documentos de apoyo	<ul style="list-style-type: none"> <li>• Dos manuales técnicos (2008-2011 y 2011-2012).</li> <li>• Dos documentos de Marcos de referencia (para La habilidad matemática y lectora).</li> <li>• Un manual para docentes y directivos.</li> </ul>

## **Estándares sobre la entrega de resultados y el uso de la prueba Enlace EMS**

No hay estándares nacionales de evaluación que especifiquen lo que debe de reportarse para una prueba estandarizada.

No hay estándares nacionales que especifiquen lo que debe de reportarse para la prueba de Enlace EMS.

Los estándares que pueden usarse son provenientes de los *Standards for Educational and Psychological Testing* y el *Code of Fair Testing Practices in Education*, que en consideración de los elementos mostrados en la ficha, pueden ser los siguientes:

**Estándar 5.10** Cuando se entregue información de puntajes a estudiantes, padres, representantes legales, maestros, clientes o medios de comunicación, los responsables de realizar la evaluación deben proporcionar interpretaciones adecuadas. Las interpretaciones deben describir en un lenguaje sencillo lo que abarca la prueba, el significado de los puntajes y su precisión, las malas interpretaciones más comunes y cómo interpretar adecuadamente.

**Estándar 5.13** La entrega de resultados a usuarios autorizados o instituciones debe realizarse asegurando la confidencialidad de los puntajes.

**Estándar 6.12** Los editores y servicios de calificación que ofrecen interpretaciones de calificaciones generadas por computadora deben proporcionar un resumen de la evidencia que apoya las interpretaciones.

**Estándar 7.8** Cuando los puntajes se reportan al público de forma desagregada como género, etnicidad, edad, lengua, eficacia o discapacidad; se deben incluir advertencias en los reportes de investigación acerca de la incomparabilidad de los grupos.

**Estándar 8.8** Cuando los reportes incluyen asignar individuos a ciertas categorías, estas deben elegirse cuidadosamente y describirse con precisión; deben elaborarse descripciones que no estigmaticen y que sean representaciones exactas.

**Estándar 11.15** Quienes administran la prueba deben ser advertidos sobre potenciales malas interpretaciones y posibles consecuencias no intencionales de la prueba; los usuarios deben prever acciones que minimicen o eviten malas interpretaciones previsibles o consecuencias negativas no intencionales.

**Standard 12.9** Los profesionales responsables de supervisar los programas de evaluación grupal deben asegurarse de que los individuos que interpretan las calificaciones han recibido una instrucción apropiada sobre los métodos adecuados para dicha tarea.

**Estándar 13.3** Cuando una prueba se usa como indicador de logro o desempeño respecto de un dominio o estándar curricular, se debe proporcionar evidencia acerca del rango o extensión de conocimiento que contiene la prueba sobre ese dominio o estándar. Estos últimos deben describirse con detalle suficiente para poder evaluar su relación con el contenido de la prueba. Los análisis deben explicitar los aspectos del dominio que la prueba abarca y aquellos que no.

**Estándar 13.13** Los responsables de las evaluaciones deben asegurar que los individuos que interpretan los resultados y toman decisiones en el contexto escolar están calificados para hacerlo o tienen asistencia de personas que sí lo están.

**Estándar 13.14** En escenarios educativos los reportes de puntajes deben acompañarse de una clara descripción del error de medida asociado a cada puntaje o nivel de clasificación y la forma de interpretar los puntajes (p. 148).

**Estándar 13.16** En escenarios educativos donde se reporte un puntaje, debe proporcionarse el dato de la fecha de administración de la prueba. Esta información y el tiempo de las normas o referentes utilizados para hacer las interpretaciones debe considerarse al hacer inferencias (p. 149).

**C1.** Proporcione información para apoyar las interpretaciones de los resultados, incluyendo la naturaleza del contenido, las normas o grupos comparativos y evidencia técnica. Aconseje a los usuarios de los beneficios y límites de los resultados y su interpretación. Advierta sobre los problemas que se pueden generar si se presume de una precisión mayor a la garantizada.

**C3.** Especifique los usos apropiados de la prueba y advierta a los usuarios sobre los potenciales usos inadecuados.

**C4.** Cuando existan estándares o clasificaciones, proporcione las razones, procedimientos y evidencia del ajuste de los estándares de desempeño o puntos de corte. Evite el uso de frases que sean negativas o estigmas.

**C6.** Proporcione información que permita a los usuarios interpretar y reportar los resultados de los subgrupos de forma precisa, incluyendo información de quién, cuándo y dónde no se incluyen en los diferentes grupos, así como los factores que pueden influir en la interpretación de los resultados.

**C8.** Proporcione guías para los usuarios acerca de cómo verificar que la prueba cumple con su propósito.

## **Reconocimiento de consideraciones particulares de cada elemento del reporte de una prueba estandarizada**

El primer elemento considerado para un reporte de resultados es que debe informar qué es lo que mide la prueba, es decir, el objetivo.

- Proporcionar la información clara y sencillamente.
- Considerar cuidadosamente el lenguaje.
- Personalizar para ayudar a conectar al usuario con el reporte
- Incluir un resumen narrativo de fácil lectura de los resultados del estudiante al inicio del reporte del estudiante.

Y consideraciones generales de estilo como:

- Ayudar al lector a desarrollar una estrategia de lectura como tablas de contenido, glosarios, objetivos y resúmenes.
- Usar técnicas que señalan o identifican unidades particulares del texto como encabezados, sistemas numéricos o listas que deben ser precisas, específicas y concretas.
- Tipografía. Entendida como una letra, número o cualquier otro carácter susceptible de impresión. La legibilidad es la consideración primaria, incluyendo el uso de itálicas, negritas, subrayados, letras mayúsculas o colores.
- La combinación del tipo, color y técnicas de organización espacial que estructuran de manera eficiente un documento. El uso de información en vertical u horizontal se determina por la intención e impacto deseado.

## **Recomendaciones para elaborar el reporte de resultados de la prueba Enlace EMS**

### **Lo que proporciona la prueba respecto a su información sobre su objetivo**

#### **El reporte de resultados de la prueba Enlace EMS para el alumno**

El reporte de la prueba Enlace EMS para el alumno se proporciona únicamente en línea y está dividido en tres pestañas o accesos separados:

- La primera es una descripción de las condiciones contextuales de escuelas, una descripción de lo que mide la habilidad matemática y la habilidad lectora y una introducción a los reportes de cada resultado.
- La segunda pestaña es un reporte con la descripción de cuatro niveles de dominio, indicando en cuál se encuentra el sustentante, y una tabla de datos horizontal que indica el porcentaje de alumnos en cada nivel de dominio ajustado a condiciones similares de la escuela del alumno.
- las tercera y cuarta pestañas son los reactivos de toda la prueba y el resultado que tuvieron en cada uno de ellos.

En este esquema se presenta la primera pestaña, de presentación, donde el alumno accede con un número de registro o su CURP (Clave Única de Registro Poblacional) y aparece esta información:

<b>SEP</b>		<b>ENLACE</b> Evaluar para Aprender			
SECRETARÍA DE EDUCACIÓN PÚBLICA		<b>PRUEBA ENLACE MEDIA SUPERIOR 2008 INFORME DE RESULTADOS</b>			
		FOLIO: <input type="text"/>			
<b>DATOS GENERALES</b>					
Nombre del Alumno:	<input type="text"/>				
CURP:	<input type="text"/>				
Nombre de la Escuela:	<input type="text"/>				
CCT:	<input type="text"/>		Turno:	<input type="text"/>	
Modalidad:	<input type="text"/>				
Tipo de Sostenimiento:	<input type="text"/>				
Grado de Marginación (CONAPO):	<input type="text"/>				
Entidad:	<input type="text"/>		Municipio:	<input type="text"/>	
Localidad:	<input type="text"/>				
<p>El propósito de la prueba ENLACE Media Superior es determinar en qué medida eres capaz de aplicar a situaciones del mundo real los conocimientos y habilidades básicas que adquiriste durante la educación media superior, que te permiten hacer un uso apropiado de la lengua –habilidad lectora– y de las matemáticas –habilidad matemática–.</p>					
<b>Áreas que mide la prueba ENLACE</b>					
<p><b>Habilidad Lectora.</b> Es la capacidad de un individuo para comprender, utilizar y analizar textos escritos, con el fin de alcanzar sus metas, desarrollar el conocimiento, el potencial personal, y participar en la sociedad.</p> <p><b>Habilidad Matemática.</b> Es la aptitud de un individuo para identificar y comprender el papel que desempeñan las matemáticas en el mundo, alcanzando razonamientos bien fundados, utilizando y participando de las matemáticas en función de las necesidades diarias como ciudadano constructivo, comprometido y reflexivo.</p> <p>El siguiente esquema muestra en qué Nivel de Dominio te encuentras en cada área y la descripción de las habilidades que lo conforman. Considera que, en la prueba, demostraste poseer esos conocimientos y los de todos los niveles anteriores al que tú te encuentras. Te recomendamos revisar los demás niveles para que conozcas los aspectos que debes mejorar.</p>					
<b>Ahora sí, ¡mira tu reporte de resultados!</b>					

## Lo que mencionan los indicadores y las recomendaciones para reportar la intención de una prueba

- Proporcionar la información clara y sencillamente.
- Considerar cuidadosamente el lenguaje.
- Personalizar para ayudar a conectar al usuario con el reporte
- Incluir un resumen narrativo de fácil lectura de los resultados del estudiante al inicio del reporte del estudiante.

## Sugerencias de modificaciones

A continuación se muestra la recomendación de cómo debería ser la pestaña de presentación. Posteriormente se explica la modificación y su argumento.

INTRODUCCIÓN	Resultados	Tu desempeño en Comunicación	Tu desempeño en Matemáticas
<p style="text-align: center;"><b>PRUEBA ENLACE MEDIA SUPERIOR 201X</b></p> <p style="text-align: center;"><b>INFORME DE RESULTADOS</b></p> <p><b>Alberto Aguilar Rojas</b></p> <p style="text-align: center;"><b>¿Qué es la prueba ENLACE Media Superior?</b></p> <p>Es una prueba diseñada para saber en qué nivel eres capaz de poner en práctica en situaciones del mundo real las competencias disciplinares básicas de Comunicación (o comprensión lectora) y de matemáticas que adquiriste durante tu trayectoria escolar.</p> <p style="text-align: center;"><b>¿Qué se evalúa en esta prueba?</b></p> <ul style="list-style-type: none"><li>• <b>Comprensión lectora:</b> Es la capacidad para comprender, analizar, interpretar, reflexionar, evaluar y utilizar textos escritos, mediante la identificación de su estructura, funciones y elementos con el fin de desarrollar una competencia comunicativa y construir nuevos conocimientos que permitan intervenir activamente en la sociedad.</li><li>• <b>Matemáticas:</b> Es la capacidad para identificar, interpretar, aplicar, sintetizar y evaluar matemáticamente el entorno, usando la creatividad y el pensamiento lógico y crítico que permite solucionar problemas cuantitativos, con diferentes herramientas matemáticas.</li></ul> <p style="text-align: center;">Cada pestaña en la parte superior contiene el resultado que obtuviste</p>			

Uno de los principales lineamientos internacionales mencionados en diferentes documentos es asegurar que los usuarios conocen los usos apropiados y las interpretaciones que no deben hacerse. Proponemos una sección (puede ser una sección

interna en la misma pestaña) que haga evidente en un mismo bloque de información (la introducción) lo que se puede o no hacer con la prueba. El siguiente cuadro lo ilustra.

INTRODUCCIÓN	Resultados	Tu desempeño en Comunicación	Tu desempeño en Matemáticas
<p><b>PRUEBA ENLACE MEDIA SUPERIOR 201X</b></p> <p><b>INFORME DE RESULTADOS</b></p> <p><b>Considera que la prueba...</b></p> <ul style="list-style-type: none"><li>✓ está hecha sólo para evaluar las disciplinas básicas de comunicación y matemáticas.</li><li>✓ está diseñada para conocer tu nivel de dominio y saber qué puedes realizar o qué falta por hacer en estas dos disciplinas.</li><li>✓ te proporciona las preguntas que contestaste correcta e incorrectamente.</li><li>× <b>no</b> está hecha para dar conclusiones de toda la Enseñanza Media Superior, o los subsistemas, las escuelas ni de los estados.</li><li>× <b>no</b> está hecha para influir en tus calificaciones ni para penalizar.</li><li>× <b>no</b> se diseñó para hacer comparaciones entre alumnos o entre escuelas.</li><li>× <b>no</b> informa si tu plantel o maestros son mejores o peores respecto a otra escuela.</li></ul>			

## Consideraciones al diseñar estas pestañas

- Eliminar la sección de los datos generales por considerar que no es información relevante o útil para el alumno; o al menos de manera evidente no es información que se vincule al proceso de producir información que produzca conocimiento en el sustentante.
- Utilizar los encabezados de manera más eficiente. A los encabezados y logos institucionales que siempre estarán presentes, se añade el nombre del usuario y son evidentes dos mensajes centrales, qué es y qué evalúa la prueba.
- En la primera pestaña la narración está redactada en segunda persona en la sección de lo que mide la prueba; pero no en la sección que describe los campos disciplinares.
- En cada pestaña se distribuyó en el espacio de manera que se visualizara que hay solo dos secciones importantes.
- La redacción tiene sintaxis muy concreta, aunque la sentencia original no se puede modificar debido a que es un trabajo colegiado. Es importante considerar que toda la información que se presente en estas pestañas se complementa o reafirma con la que se presente en las siguientes pestañas.
- Se usó un solo tipo de letra, con dos colores, negro para indicar títulos o notas y azul intenso para el resto del texto. El lector visualiza en cada color intenciones claramente diferentes: texto o encabezado. Dentro del texto se usó el formato de negrita para identificar un concepto inicial o importante.
- En la segunda pestaña se usaron seis advertencias, tres positivas y tres negativas, indicadas con un signo que reforzaba el uso o no de alguna actividad, enfatizando con el formato de negrita la palabra *no*.
- Las consideraciones fueron redactadas de manera breve y contundente. Las afirmaciones son de tono descriptivo y las negaciones en tono imperativo. No es posible hacer recomendaciones más allá de la descripción mientras no se realice una investigación o entrevistas que aseguren las inferencias que se pueden realizar.

## Reportes individuales del nivel de dominio que el alumno obtuvo en la prueba Enlace EMS

### Lo que proporciona la prueba respecto al nivel de dominio que obtuvo un alumno en cada campo disciplinario

El segundo reporte es el nivel de dominio que tiene el sustentante en ambas competencias; el siguiente cuadro expone, para la comprensión lectora, sus componentes, el nivel de dominio del alumno, la explicación de todos los dominios y un cuadro horizontal que indica los porcentajes de cada dominio de la escuela, la entidad y el país.

HABILIDAD LECTORA	
<b>INSUFICIENTE</b>	Sólo eres capaz de identificar elementos que se encuentran de manera explícita en textos narrativos y expositivos, ya sean acciones, hechos, episodios, personajes o sus características. Realizas inferencias sencillas sobre las acciones de los personajes y estableces relaciones entre dos o más elementos. Identificas si la estructura de algunas partes del texto es adecuada al contenido que presenta.
<b>ELEMENTAL</b>	Ubicas e integras diferentes partes de un texto. Reconoces la idea central y comprendes relaciones del tipo: problema-solución, causa-efecto, comparación-contraste. Inferes el significado de palabras, así como la relación entre párrafos e ideas. Reconoces la postura del autor. Relacionas la información que se presenta en el texto y la que se encuentra en tablas.
<b>BUENO</b>	Relacionas elementos que se encuentran a lo largo del texto y/o en glosarios. Comprendes el texto de forma completa y detallada, y sintetizas su contenido global. Inferes relaciones del tipo: problema-solución, causa-efecto, comparación-contraste. Estableces relaciones entre la postura del autor y la información que apoya su punto de vista, por ejemplo: un hecho, un dato, el contexto, etcétera. Reconoces la función que cumplen las tablas en los textos expositivos. Evalúas la estructura del texto en relación con su contenido.
<b>EXCELENTE</b>	Haces inferencias complejas para construir una interpretación global del texto. Comprendes la información contenida en tablas y esquemas y la relacionas con el contenido del texto. Estableces relaciones entre argumentos y contraargumentos. Analizas si la organización, las expresiones y los recursos que utiliza el autor son adecuados al tipo de texto y a su destinatario.



En la siguiente tabla se presenta el porcentaje de alumnos del último grado de la escuela, entidad y país, en cada nivel de dominio y con los mismos *tipo de sostenimiento, modalidad y grado de marginación*

HABILIDAD LECTORA	INSUFICIENTE			ELEMENTAL			BUENO			EXCELENTE		
	Escuela	Entidad	País	Escuela	Entidad	País	Escuela	Entidad	País	Escuela	Entidad	País

Número de alumnos programados para evaluar del último grado en tu escuela:

Número de alumnos evaluados en habilidad lectora:

El cuadro siguiente es el reporte de la sección de matemáticas, que se presenta posteriormente al de comprensión lectora.

Ambos reportes utilizan los mismos colores para cada nivel de dominio, la misma presentación gráfica, narrativa y disposición espacial de los contenidos, solo hay un cambio de los contenidos en ambos reportes.

HABILIDAD MATEMÁTICA	
	<b>INSUFICIENTE</b> Sólo resuelves problemas donde la tarea se presenta directamente. Identificas información en esquemas o gráficas y realizas estimaciones. Efectúas sumas y restas con números enteros y traduces del lenguaje común al lenguaje algebraico. Resuelves problemas en los que se requiere identificar figuras planas y tridimensionales.
	<b>ELEMENTAL</b> Realizas multiplicaciones y divisiones con números enteros, y sumas que los combinan con números fraccionarios. Calculas porcentajes, utilizas fracciones equivalentes, ordenas y comparas información numérica. Estableces relaciones entre variables y resuelves problemas que combinan datos en tablas y gráficas. Aplicas conceptos simples de probabilidad y estadística. Construyes expresiones equivalentes a una ecuación algebraica y resuelves ejercicios con sistemas de ecuaciones lineales. Manejas conceptos sencillos de simetría y resuelves problemas que involucran un razonamiento viso-espacial.
	<b>BUENO</b> Resuelves problemas que involucran más de un procedimiento. Realizas multiplicaciones y divisiones combinando números enteros y fraccionarios. Calculas raíz cuadrada, razones y proporciones, y resuelves problemas con números mixtos. Analizas las relaciones entre dos o más variables de un proceso social o natural y resuelves los sistemas de ecuaciones que las representan. Identificas funciones a partir de sus gráficas para estimar el comportamiento de un fenómeno. Construyes una figura tridimensional a partir de otras e identificas características de una figura transformada. Utilizas fórmulas para calcular superficies y volumen, y reconoces los elementos de una cónica a partir de su representación gráfica.
	<b>EXCELENTE</b> Empleas operaciones con fracciones para solucionar problemas y resuelves combinaciones con signos de agrupación. Conviertes cantidades de sistema decimal a sexagesimal. Identificas la relación existente entre gráficas y funciones lineales o cuadráticas, y expresas algebraicamente una representación gráfica. Aplicas conceptos avanzados de probabilidad. Solucionas problemas con series de imágenes tridimensionales y aplicas conceptos de simetría. Utilizas fórmulas para calcular el perímetro de composiciones geométricas. Determinas los valores de los elementos de la circunferencia, la parábola y la elipse a partir de su ecuación y viceversa; identificas la ecuación de una recta a partir de sus elementos y la aplicas para encontrar la distancia entre dos puntos. Solucionas problemas donde se aplican funciones y leyes trigonométricas.

En la siguiente tabla se presenta el porcentaje de alumnos del último grado de la escuela, entidad y país, en cada nivel de dominio y con los mismos *tipo de sostenimiento, modalidad y grado de marginación*

HABILIDAD MATEMÁTICA	INSUFICIENTE			ELEMENTAL			BUENO			EXCELENTE		
	Escuela	Entidad	País	Escuela	Entidad	País	Escuela	Entidad	País	Escuela	Entidad	País

Número de alumnos programados para evaluar del último grado en tu escuela:

Número de alumnos evaluados en habilidad matemática:

## Lo que mencionan los indicadores y las recomendaciones para reportar el nivel de dominio

- **Estándar 8.8** Cuando los reportes incluyen asignar individuos a ciertas categorías, estas deben elegirse cuidadosamente y describirse con precisión; deben elaborarse descripciones que no estigmaticen y que sean representaciones exactas.
- **Estándar 13.3** Cuando una prueba se usa como indicador de logro o desempeño respecto de un dominio o estándar curricular, se debe proporcionar evidencia acerca del rango o extensión de conocimiento que contiene la prueba sobre ese dominio o estándar. Estos últimos deben describirse con detalle suficiente para poder evaluar su relación con el contenido de la prueba. Los análisis deben explicitar los aspectos del dominio que la prueba abarca y aquellos que no.
- Incluir texto fácil de leer que apoye y mejore las interpretaciones de gráficos y tablas.
- Cuidar no hacer demasiado al mostrar datos (la presentación de éstos deben estar hechos para satisfacer un pequeño número de propósitos preestablecidos).
- Los apoyos visuales como cuadros y gráficos deben ser utilizados para destacar los principales hallazgos.
- Asegurar que las gráficas, figuras y tablas pueden ser impresas, reproducidas y reducidas sin perder calidad visual (algunos resultados interesantes pueden ser copiados y distribuidos y el lector puede estar forzado a mirar malas copias en cada ocasión).

## Sugerencias de modificaciones

A continuación se muestran los reportes con las modificaciones, primero el de dominio en comprensión lectora, posteriormente el de matemáticas.

INTRODUCCIÓN	RESULTADOS	Tu desempeño en Comunicación	Tu desempeño en Matemáticas										
<p><b>PRUEBA ENLACE MEDIA SUPERIOR 201X</b></p> <p><b>LOS NIVELES DE DOMINIO EN COMPRENSIÓN LECTORA</b></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr style="background-color: #003366; color: white;"> <th style="width: 20%;">Dominio</th> <th style="width: 80%;">¿Qué haces en cada nivel?</th> </tr> </thead> <tbody> <tr> <td style="vertical-align: top; padding: 5px;"> <p><b>INSUFICIENTE</b></p> <p style="text-align: center;"></p> </td> <td style="padding: 5px;"> <ul style="list-style-type: none"> <li>✓ Identificas elementos que se encuentran de manera explícita en textos narrativos y expositivos. Esos elementos pueden ser acciones, hechos, episodios, personajes o sus características.</li> <li>✓ Realizas inferencias sencillas sobre las acciones de los personajes y estableces relaciones entre dos o más elementos.</li> <li>✓ Identificas si la estructura de algunas partes del texto es adecuada al contenido que presenta.</li> </ul> </td> </tr> <tr> <td style="vertical-align: top; padding: 5px;"> <p><b>ELEMENTAL</b></p> <p style="text-align: center;"></p> </td> <td style="padding: 5px;"> <ul style="list-style-type: none"> <li>✓ Utilizas e integras diferentes partes de un texto.</li> <li>✓ Reconoces la idea central y comprendes relaciones. Por ejemplo relaciones que expresan el problema y su solución, la causa y el efecto, o una comparación y su contraste.</li> <li>✓ Infieres el significado de las palabras, así como la relación entre párrafos e ideas.</li> <li>✓ Reconoces la postura del autor.</li> <li>✓ Relacionas la información que se presenta en el texto y la que se encuentra en tablas.</li> </ul> </td> </tr> <tr> <td style="vertical-align: top; padding: 5px;"> <p><b>BUENO</b></p> <p style="text-align: center;"></p> </td> <td style="padding: 5px;"> <ul style="list-style-type: none"> <li>✓ Relacionas elementos que se encuentran a lo largo del texto y en glosarios.</li> <li>✓ Comprendes el texto de forma completa y detallada, y sintetizas su contenido global.</li> <li>✓ Infieres problemas que expresan el problema y su solución, la causa y el efecto, una comparación y su contraste.</li> <li>✓ Estableces relaciones entre la postura del autor y la información que apoya su punto de vista, por ejemplo: un hecho, un dato, o el contexto.</li> <li>✓ Reconoces la función que cumplen las tablas en los textos expositivos.</li> <li>✓ Evalúas la estructura del texto en relación con su contenido.</li> </ul> </td> </tr> <tr> <td style="vertical-align: top; padding: 5px;"> <p><b>EXCELENTE</b></p> </td> <td style="padding: 5px;"> <ul style="list-style-type: none"> <li>✓ Haces inferencias complejas para construir una interpretación global del texto.</li> <li>✓ Comprendes la información contenida en tablas y esquemas y las relacionas con el contenido del texto.</li> <li>✓ Estableces relaciones entre argumentos y contraargumentos.</li> <li>✓ Analizas si la organización, las expresiones y los recursos que utiliza el autor son adecuados al tipo de texto y a su destinatario.</li> </ul> </td> </tr> </tbody> </table>				Dominio	¿Qué haces en cada nivel?	<p><b>INSUFICIENTE</b></p> <p style="text-align: center;"></p>	<ul style="list-style-type: none"> <li>✓ Identificas elementos que se encuentran de manera explícita en textos narrativos y expositivos. Esos elementos pueden ser acciones, hechos, episodios, personajes o sus características.</li> <li>✓ Realizas inferencias sencillas sobre las acciones de los personajes y estableces relaciones entre dos o más elementos.</li> <li>✓ Identificas si la estructura de algunas partes del texto es adecuada al contenido que presenta.</li> </ul>	<p><b>ELEMENTAL</b></p> <p style="text-align: center;"></p>	<ul style="list-style-type: none"> <li>✓ Utilizas e integras diferentes partes de un texto.</li> <li>✓ Reconoces la idea central y comprendes relaciones. Por ejemplo relaciones que expresan el problema y su solución, la causa y el efecto, o una comparación y su contraste.</li> <li>✓ Infieres el significado de las palabras, así como la relación entre párrafos e ideas.</li> <li>✓ Reconoces la postura del autor.</li> <li>✓ Relacionas la información que se presenta en el texto y la que se encuentra en tablas.</li> </ul>	<p><b>BUENO</b></p> <p style="text-align: center;"></p>	<ul style="list-style-type: none"> <li>✓ Relacionas elementos que se encuentran a lo largo del texto y en glosarios.</li> <li>✓ Comprendes el texto de forma completa y detallada, y sintetizas su contenido global.</li> <li>✓ Infieres problemas que expresan el problema y su solución, la causa y el efecto, una comparación y su contraste.</li> <li>✓ Estableces relaciones entre la postura del autor y la información que apoya su punto de vista, por ejemplo: un hecho, un dato, o el contexto.</li> <li>✓ Reconoces la función que cumplen las tablas en los textos expositivos.</li> <li>✓ Evalúas la estructura del texto en relación con su contenido.</li> </ul>	<p><b>EXCELENTE</b></p>	<ul style="list-style-type: none"> <li>✓ Haces inferencias complejas para construir una interpretación global del texto.</li> <li>✓ Comprendes la información contenida en tablas y esquemas y las relacionas con el contenido del texto.</li> <li>✓ Estableces relaciones entre argumentos y contraargumentos.</li> <li>✓ Analizas si la organización, las expresiones y los recursos que utiliza el autor son adecuados al tipo de texto y a su destinatario.</li> </ul>
Dominio	¿Qué haces en cada nivel?												
<p><b>INSUFICIENTE</b></p> <p style="text-align: center;"></p>	<ul style="list-style-type: none"> <li>✓ Identificas elementos que se encuentran de manera explícita en textos narrativos y expositivos. Esos elementos pueden ser acciones, hechos, episodios, personajes o sus características.</li> <li>✓ Realizas inferencias sencillas sobre las acciones de los personajes y estableces relaciones entre dos o más elementos.</li> <li>✓ Identificas si la estructura de algunas partes del texto es adecuada al contenido que presenta.</li> </ul>												
<p><b>ELEMENTAL</b></p> <p style="text-align: center;"></p>	<ul style="list-style-type: none"> <li>✓ Utilizas e integras diferentes partes de un texto.</li> <li>✓ Reconoces la idea central y comprendes relaciones. Por ejemplo relaciones que expresan el problema y su solución, la causa y el efecto, o una comparación y su contraste.</li> <li>✓ Infieres el significado de las palabras, así como la relación entre párrafos e ideas.</li> <li>✓ Reconoces la postura del autor.</li> <li>✓ Relacionas la información que se presenta en el texto y la que se encuentra en tablas.</li> </ul>												
<p><b>BUENO</b></p> <p style="text-align: center;"></p>	<ul style="list-style-type: none"> <li>✓ Relacionas elementos que se encuentran a lo largo del texto y en glosarios.</li> <li>✓ Comprendes el texto de forma completa y detallada, y sintetizas su contenido global.</li> <li>✓ Infieres problemas que expresan el problema y su solución, la causa y el efecto, una comparación y su contraste.</li> <li>✓ Estableces relaciones entre la postura del autor y la información que apoya su punto de vista, por ejemplo: un hecho, un dato, o el contexto.</li> <li>✓ Reconoces la función que cumplen las tablas en los textos expositivos.</li> <li>✓ Evalúas la estructura del texto en relación con su contenido.</li> </ul>												
<p><b>EXCELENTE</b></p>	<ul style="list-style-type: none"> <li>✓ Haces inferencias complejas para construir una interpretación global del texto.</li> <li>✓ Comprendes la información contenida en tablas y esquemas y las relacionas con el contenido del texto.</li> <li>✓ Estableces relaciones entre argumentos y contraargumentos.</li> <li>✓ Analizas si la organización, las expresiones y los recursos que utiliza el autor son adecuados al tipo de texto y a su destinatario.</li> </ul>												
<p><b>Tu nivel</b> </p>													

INTRODUCCIÓN	RESULTADOS	Tu desempeño en Comunicación	Tu desempeño en Matemáticas
<b>LOS NIVELES DE DOMINIO EN MATEMÁTICAS</b>			
<b>Dominio</b>		<b>¿Qué haces en cada nivel?</b>	
<b>INSUFICIENTE</b> 	<ul style="list-style-type: none"> <li>✓ Resuelves problemas donde la tarea se presenta directamente.</li> <li>✓ Identificas información en esquemas o gráficas y realizas estimaciones.</li> <li>✓ Efectúas sumas y restas con números enteros y traduces del lenguaje común al lenguaje algebraico.</li> <li>✓ Resuelves problemas en los que se requiere identificar figuras planas y tridimensionales.</li> </ul>		
 <b>Tu nivel</b> <b>ELEMENTAL</b> 	<ul style="list-style-type: none"> <li>✓ Realizas multiplicaciones y divisiones con números enteros, y sumas que los combinan con números fraccionarios.</li> <li>✓ Calculas porcentajes, utilizas fracciones equivalentes, ordenas y comparas información numérica.</li> <li>✓ Estableces relaciones entre variables y resuelves problemas que combinan datos en tablas y gráficas.</li> <li>✓ Aplicas conceptos simples de probabilidad y estadística.</li> <li>✓ Construyes expresiones equivalentes a una ecuación algebraica y resuelves ejercicios con sistemas de ecuaciones lineales.</li> <li>✓ Manejas conceptos sencillos de simetría y resuelves problemas que involucran un razonamiento viso-espacial.</li> </ul>		
<b>BUENO</b> 	<ul style="list-style-type: none"> <li>✓ Resuelves problemas que involucran más de un procedimiento.</li> <li>✓ Realizas multiplicaciones y divisiones combinando números enteros y fraccionarios.</li> <li>✓ Calculas raíz cuadrada, razones y proporciones, y resuelves problemas con números mixtos.</li> <li>✓ Analizas las relaciones entre dos o más variables de un proceso social o natural y resuelves los sistemas de ecuaciones que las representan.</li> <li>✓ Identificas funciones a partir de sus gráficas para estimar el comportamiento de un fenómeno.</li> <li>✓ Construyes una figura tridimensional a partir de otras e identificas características de una figura transformada.</li> <li>✓ Utilizas fórmulas para calcular superficies y volumen y reconoces los elementos de una cónica a partir de su representación gráfica.</li> </ul>		
<b>EXCELENTE</b>	<ul style="list-style-type: none"> <li>✓ Empleas operaciones con fracciones para solucionar problemas y resuelves combinaciones con signos de agrupación.</li> <li>✓ Conviertes cantidades de sistema decimal a sexagesimal.</li> <li>✓ Identificas la relación existente entre gráficas y funciones lineales o cuadráticas, y expresas algebraicamente una representación gráfica.</li> <li>✓ Aplicas conceptos avanzados de probabilidad.</li> <li>✓ Solucionas problemas con series de imágenes tridimensionales y aplicas conceptos de simetría.</li> <li>✓ Utilizas fórmulas para calcular el perímetro de composiciones geométricas.</li> <li>✓ Determinas los valores de los elementos de la circunferencia, la parábola y la elipse a partir de su ecuación y viceversa.</li> <li>✓ Identificas la ecuación de una recta a partir de sus elementos y la aplicas para encontrar la distancia entre dos puntos.</li> <li>✓ Solucionas problemas donde se aplican funciones y leyes trigonométricas.</li> </ul>		

## Consideraciones al diseñar los reportes del nivel de dominio del estudiante

- Eliminar el reporte horizontal de porcentajes. Dada la necesidad de proveer al sustentante sólo la información que es relevante para su diagnóstico. Los porcentajes no brindan información útil al alumno para saber o complementar su nivel de habilidad.
- Usar encabezados. Debe ser identificada cada columna o sección, como las referidas a dominios y sus descripciones.
- Eliminar los colores en cada nivel de dominio. Los colores pueden mejorar un reporte al enfocar la atención sobre información crítica, especialmente en gráficos. En este caso un cuadro de cuatro niveles evidentemente diferenciados hace innecesario el uso de distintos colores, que son plena y claramente diferenciables visualmente.
- Evidenciar que cada nivel es jerárquico. Cada nivel de dominio incluye a los anteriores, y es necesario indicarlo de manera gráfica con el uso de señales, como las flechas descendentes.
- La tipografía de las letras es más clara y la descripción de cada nivel debe ser más concreto.
- Considerar los diferentes niveles de comprensión de análisis textual de los usuarios. El texto debe redactarse de manera que pueda ser comprendido claramente por la audiencia con un nivel de comprensión elemental. Las palabras técnicas o demasiado académicas contribuyen, en el mejor de los casos, a provocar que el lector no comprenda lo que es capaz o no de hacer. También palabras como “solo eres capaz” pueden demeritar o ser poco sensibles. Sin alterar las descripciones de dominio, se eliminaron algunas palabras redundantes, sin quitar ninguna descripción.
- En cada nivel de dominio, identificar su descripción de forma clara y separada de las otras. En cada nivel de dominio original hay de tres hasta nueve descripciones agrupadas en un solo párrafo, con diferentes mensajes y significados, esto provoca que el lector no pueda identificar un mensaje común,

ni pueda reconocer de manera sencilla los elementos que componen cada nivel.

- Todas las descripciones están identificadas con un signo positivo. Si asumimos que tienen una connotación positiva (independientemente del nivel que se tenga en general) es favorable el uso de signos que lo refuercen, como las viñetas de “palomitas”.

### **Reportes individuales con los reactivos por cada campo disciplinario**

#### **Lo que mencionan los indicadores y las recomendaciones para reportar resultados para el alumno**

- **Estándar 5.10** Cuando se entregue información de puntajes a estudiantes, padres, representantes legales, maestros, clientes o medios de comunicación, los responsables de realizar la evaluación deben proporcionar interpretaciones adecuadas. Las interpretaciones deben describir en un lenguaje sencillo lo que abarca la prueba, el significado de los puntajes y su precisión, las malas interpretaciones más comunes y cómo interpretar adecuadamente.
- **Estándar 11.15** Quienes administran la prueba deben ser advertidos sobre potenciales malas interpretaciones y posibles consecuencias no intencionales de la prueba; los usuarios deben prever acciones que minimicen o eviten malas interpretaciones previsibles o consecuencias negativas no intencionales.
- **Estándar 13.3** Cuando una prueba se usa como indicador de logro o desempeño respecto de un dominio o estándar curricular, se debe proporcionar evidencia acerca del rango o extensión de conocimiento que contiene la prueba sobre ese dominio o estándar. Estos últimos deben describirse con detalle suficiente para poder evaluar su relación con el contenido de la prueba.

Los análisis deben explicitar los aspectos del dominio que la prueba abarca y aquellos que no.

- **Estándar 15.11** Cuando los resultados se proporcionan al público o a los que definen políticas educativas, quienes proporcionen los resultados deben proporcionar toda la información complementaria que minimice las malas interpretaciones de los datos.
- **C8.** Proporcione guías para los usuarios acerca de cómo verificar que la prueba cumple con su propósito.

Después de indicar el nivel de dominio del sustentante, se pueden consultar todos los reactivos con las respuestas que el estudiante eligió y compararlas con la opción correcta (Ceneval, 2013) Al elegir desplegar un reactivo se muestra éste y el resultado obtenido por el alumno, así como una descripción de lo que el reactivo mide. En la tabla siguiente se muestra la manera en la que se despliega la información en los reportes que sí pueden ser públicos, como los resultados por escuela, pero el formato de presentación es el mismo para el estudiante.

Se presentan todas las preguntas del examen contenidas en una pantalla, en una forma visualmente resumida por medio de pequeños bloques donde cada uno tiene el número de la pregunta en el examen; al acceder a un bloque se muestra la pregunta completa con su respuesta correcta y si el alumno la contestó correctamente o no. Están agrupadas por tipos de textos (apelativo, argumentativo, narrativo o expositivo) o por contenido matemático (cantidad, cambios y relaciones, espacio y forma).

Este despliegue gráfico es la única manera de acceder a todos los reactivos de la prueba. Es importante mencionar que el número de reactivos mostrados en este reporte no coincide con los que oficialmente expresa, es decir, para los textos argumentativos y narrativos los reactivos mostrados son 14, pero deberían ser 15 según la estructura oficial. Para cada contenido matemático de forma oficial se reportarán 20 reactivos, pero en lo expresado en el despliegue gráfico se reportan

17 en Cantidad y 18 en Espacio y forma y en Cambios y relaciones. No hay de forma evidente una explicación que indique la razón de esas diferencias, al menos no en este despliegue gráfico o en el portal (donde se menciona la estructura).

En los documentos que hacen referencia a los reportes que se proporcionan a los alumnos (manuales técnicos y manual para el profesorado) no hay referencias de la existencia de más resultados para el alumno que el nivel de dominio y las preguntas con las respuestas que obtuvo el estudiante (Ceneval, 2013) y las descripciones relativas a los porcentajes de los resultados de la misma población que el alumno.

Se muestran los despliegues gráficos de los reactivos, en Comunicación y Matemáticas.

## Respuestas de Comunicación distribuidas por contenido

**Respuestas dadas por la Escuela en la Sección de Comunicación**

Dé clic en el número de pregunta para conocer el detalle de su resultado.

Tema	Preguntas
Apelativo	001 002 003 004 005
Argumentativo	006 007 008 009 010 011 012 013 014 015 016 017 018 019
Narrativo	020 021 022 023 024 025 026 027 028 029 030 031 032 033 034
Expositivo	035 036 037 038 039 040 041 042 043 044 045 046 047 048 049 050

**Simbología Utilizada**

- Preguntas que contestaron incorrectamente menos del 40% de los alumnos de la Escuela
- Preguntas que contestaron incorrectamente entre el 40% y el 60% de los alumnos de la Escuela
- Preguntas que contestaron incorrectamente más del 60% de los alumnos de la Escuela

## Respuestas de Matemáticas distribuidas por contenido.

**Respuestas dadas por la Escuela en la Sección de Matemáticas**

Dé clic en el número de pregunta para conocer el detalle de su resultado.

Tema	Preguntas
Cantidad	021 022 023 024 025 026 027 028 029 030 031 032 033 034 035 036 037 038 039 040
Espacio y Forma	041 042 043 044 045 046 047 048 049 050 051 052 053 054 055
Cambios y Relaciones	056 057 058 059 060 061 062 063 064 065 066 067 068 069 070 071 072 073 074 075

**Simbología Utilizada**

- Preguntas que contestaron incorrectamente menos del 40% de los alumnos de la Escuela
- Preguntas que contestaron incorrectamente entre el 40% y el 60% de los alumnos de la Escuela
- Preguntas que contestaron incorrectamente más del 60% de los alumnos de la Escuela

### Ventajas de estos reportes

- El despliegue de los datos está en una sola pantalla, lo que facilita la visualización completa de todos los reactivos.
- Se indica con uno de dos colores la categoría de la respuesta ya sea que ésta haya sido correcta o incorrecta.
- Los reactivos están agrupados por tipo de texto o contenido de manera muy clara.
- La instrucción para ver el reactivo es concreta y clara.
- Se da por separado el reporte para Comunicación y Matemáticas.

### Sugerencias de modificaciones

A continuación se muestran el reporte de los reactivos con las recomendaciones; sólo se muestra el de comprensión lectora que tendría la misma forma que el de matemáticas.

Introducción	Resultados	Tu desempeño en Comunicación	Tu desempeño en Matemáticas															
<b>PRUEBA ENLACE MEDIA SUPERIOR 201X</b>																		
<b>INFORME DE RESULTADOS</b>																		
<b>Alberto Aguilar Rojas</b>																		
<b>RESUMEN DE RESULTADOS</b>																		
<table border="1"><thead><tr><th>Tipo de texto</th><th>Total de preguntas</th><th>Contestaste correctamente</th></tr></thead><tbody><tr><td>Apelativo</td><td>5</td><td>5</td></tr><tr><td>Argumentativo</td><td>14</td><td>8</td></tr><tr><td>Narrativo</td><td>14</td><td>10</td></tr><tr><td>Expositivo</td><td>15</td><td>11</td></tr></tbody></table>				Tipo de texto	Total de preguntas	Contestaste correctamente	Apelativo	5	5	Argumentativo	14	8	Narrativo	14	10	Expositivo	15	11
Tipo de texto	Total de preguntas	Contestaste correctamente																
Apelativo	5	5																
Argumentativo	14	8																
Narrativo	14	10																
Expositivo	15	11																

La siguiente sección es el resultado que obtuviste en cada reactivo, con el proceso que representa; para ver la pregunta haz clic en cada uno de ellos.

PROCESO DE CADA PREGUNTA	Número de pregunta	Correcta	Incorrecta
<b>TEXTO APELATIVO</b>		<b>5</b>	<b>0</b>
Identificar el asunto central que se plantea en una carta formal.	001	●	
Distinguir la frase de despedida que sea congruente con el contenido de una carta formal.	002	●	
Distinguir dos elementos de información que compitan con otros semejantes y enlistados en un texto.	003	●	
Identificar el remitente de una carta formal.	004	●	
Identificar el responsable de solucionar un problema planteado en una carta formal.	005	●	
<b>TEXTO ARGUMENTATIVO</b>		<b>8</b>	<b>6</b>
Localizar dos elementos de información vinculados y presentados de manera explícita en el texto.	006	●	
Identificar dos explicaciones presentadas en un artículo de opinión.	007	●	
Identificar el asunto central que trata el autor del texto en un párrafo específico.	008		●
Identificar el tipo de relación que existe entre dos ideas presentadas en un texto.	009	●	
Identificar la frase que sintetiza el contenidos de un artículo de opinión	010	●	
Localizar un elemento de información relevante que se presenta de manera explícita en un artículo de opinión.	011	●	
Localizar tres elementos de información vinculados y presentados de manera explícita en un texto.	012	●	
Identificar dos explicaciones utilizadas para apoyar un argumento del autor de un texto.	013		●
Identificar un contraargumento que presenta el autor de un artículo de opinión.	014		●
Identificar tres recursos discursivos empleados por el autor de un texto para apoyar su punto de vista.	015		●
Inferir cuál es la postura del autor de un artículo de opinión.	016		●
Distinguir entre un hecho y una opinión que se presentan en un párrafo.	017	●	
Identificar la estructura en que se presenta el contenido de un artículo de opinión.	018	●	
Identificar una premisa o una conclusión derivadas de un texto.	019		●
<b>TEXTO NARRATIVO</b>		<b>10</b>	<b>4</b>
Identificar el sentido de un enunciado connotativo.	051	●	
Identificar el ambiente en el que están ocurriendo los hechos en un texto narrativo.	052	●	
Identificar el motivo por el que el personaje de un texto narrativo actuó de determinada manera.	053	●	
Identificar el personaje de un texto narrativo al que corresponda una descripción física.	054	●	
Identificar el enunciado que represente el contenido de un fragmento de un relato	055		●
Identificar el enunciado que expresa la acción principal en un fragmento de un relato.	056		●
Identificar dos hechos que se relacionen casualmente a lo largo de un texto.	057	●	
Identificar tres acciones realizadas por un mismo personaje de un texto narrativo.	058	●	
Identificar la relación entre la intención del autor y la forma de un relato.	059	●	
Identificar el sentido con que se utiliza una analogía, metáfora u otra figura retórica.	060	●	

Identificar la forma en que el autor relaciona las acciones de una historia para lograr el propósito comunicativo.	061	●	
Identificar los elementos de un relato que retoma el autor de una reseña crítica del mismo.	063		●
Identificación de una acción que corresponda con las características psicológicas de un personaje de un texto narrativo.	064		●
Identificar el tema central de un texto narrativo.	065	●	
<b>TEXTO EXPOSITIVO</b>		<b>11</b>	<b>4</b>
Identificar dos elementos de información ubicados a lo largo de un artículo de divulgación.	096	●	
Identificar tres elementos de información relacionando una tabla y el artículo de divulgación que la contiene.	097	●	
Identificar un elemento de información relacionando el contenido de un texto, una tabla y su respectivo glosario.	098	●	
Identificar el tema central de un párrafo en un artículo de divulgación científica.	099	●	
Identificar el enunciado que sintetiza un apartado de un texto.	100		●
Identificar el organizador gráfico que representa la estructura de un artículo de divulgación.	101	●	
Identificar una acepción de una palabra utilizada por el autor de un texto.	102	●	
Identificar el propósito de la tabla que se incluye en un texto.	103	●	
Identificar la relación entre las ideas principales de dos párrafos.	104	●	
Identificar la razón por la que son diferentes dos fenómenos o hechos descritos en un artículo de divulgación científica.	105		●
Identificar la solución a un problema teniendo en cuenta el contenido global de un texto.	106	●	
Identificar un término utilizado en un artículo de divulgación científica que corresponde a una definición.	107		●
Identificar dos datos relacionados que estén incluidos en la tabla que acompaña un artículo de divulgación científica.	108	●	
Identificar el propósito comunicativo del autor de un artículo de divulgación científica.	109	●	
Identificar una conclusión derivada de un texto y relacionada con información externa al mismo.	110		●

#### RECOMENDACIONES GENERALES PARA APROVECHAR TUS RESULTADOS

**Ante las respuestas correctas.**

Lee detenidamente el proceso de acuerdo con el tipo de texto, eso es lo que puedes hacer ante un texto. Los procesos siempre actúan en conjunto, un proceso ayuda y complementa a los demás. Complementa tu reflexión viendo qué procesos contestaste incorrectamente

**Ante las respuestas incorrectas.**

Analiza cada proceso de acuerdo con el tipo de texto, léelo detenidamente. Cada proceso es complementario con otros, debes revisar los que hiciste bien y los que no. Revisa el reactivo hasta que estés seguro de lo que significa cada proceso

Si quieres tener más información de cómo usar los resultados para mejorar tus habilidades lectoras ve a la sección [¿cómo puedo utilizar mejor mis resultados?](#)

### Consideraciones para los reportes por ítem para el estudiante

- Hacer el reporte personalizado con el nombre del sustentante de manera visible y evidente.

- Crear un reporte resumido, visualmente accesible de las respuestas que el alumno contestó; es un puntaje crudo, que se puede contrastar solo con el número total de reactivos, pero no permite comparaciones y es un preámbulo a la siguiente sección.
- Reportar de forma evidente el proceso de cada pregunta y permitir que cada una se visualice al darle un clic. Esta forma permite que el lector conozca el proceso involucrado de cada reactivo y su resultado ante ese proceso, a la vez que tiene la posibilidad de analizar el reactivo.
- Permitir tener una versión impresa en una sola hoja. Si bien en este formato no es viable por el modelo de presentación, la versión impresa de manera horizontal permite tener en una sola hoja el resultado global.
- Ser un referente para realizar más actividades. No es la intención original, pero conocer los contenidos de los procesos ayuda al alumno a reconocer qué procesos contestó correcta o incorrectamente y basar en este documento para realizar algunas actividades.
- Elaborar de manera clara la división entre cada tipo de texto, el encabezado y los contenidos de cada texto por medio de la disposición general de los colores
- Crear una tabla sencilla, con los elementos suficientes para entender cada columna. Solo dos líneas internas marcan el número de los reactivos; pero no hay líneas en la sección que marca la diferencia entre reactivos correctos e incorrectos porque consideramos importante que se visualice como un bloque único de información, es decir, es tanto lo que contesta correctamente como lo que no es clara la ubicación y da un resumen visual de la posición de cada punto.
- Las recomendaciones al final de la sección se consideran un puente o un mecanismo que puede proporcionarles a los alumnos una actividad basada en la premisa: ¿y qué hago con estos resultados? Que puede proporcionarse como frases que generen reflexión o comprensión, o transferencia a más información, como una sección especialmente diseñada para usar los

resultados o con estrategias de aprendizaje basadas en los distintos niveles e dominio.

Estos reportes para los sustentantes son considerados prioritarios dada la intención manifiesta en los documentos donde se menciona que:

La posibilidad de conocer no solo la definición del nivel en que se está ubicado, sino las de los inferiores y superiores es la herramienta más importante que brinda Enlace EMS para que el sustentante haga uso de sus resultados para autodirigir sus procesos de aprendizaje, una vez que reconoce sus fortalezas y debilidades. (Manual Técnico, 2013, p.75).

Por tanto, mejorar el reporte para hacerlo más claro es una de las tareas fundamentales para mejorar los reportes.

### **Reportes para los profesores**

En la prueba Enlace EMS no hay reportes dirigidos a los maestros. Se menciona en los manuales técnicos y en el portal de internet que la información sirve para ellos; pero no se produce ningún reporte.

Los reportes para los maestros son, después del proporcionado a los alumnos, el insumo principal acerca del desempeño de los alumnos y son un indicador de la manera en la que los alumnos de un grupo o una escuela se apropian de los contenidos expresados en la prueba. Es muy importante que la información que se les proporcione tenga los elementos que les permitan saber qué hacer con ellos, las malas interpretaciones son comunes, particularmente con los reportes que involucran puntajes y lenguaje estadístico (Goodman, 2004).

Conocer el grado en el que los alumnos han comprendido un concepto es de un gran valor, y a los maestros se les debe proporcionar información que les permita realizar una retroalimentación de su práctica docente (INEE, 2005). Con estos fines los reportes deben realizarse y entregarse a los maestros.

Los datos disponibles por la prueba de Enlace no proporcionan datos al nivel de grupos, sólo se proporciona el reporte al nivel del plantel por turno, y muestra el número de cada reactivo con un color que representa si los alumnos de ese plantel y turno contestaron de forma incorrecta el reactivo en rangos menores al 40%, entre 40% y 60% y más del 60%.

En primer lugar, es necesario para esta prueba proporcionar información introductoria que dé la información en un lenguaje claro, sencillo y sin jerga técnica lo que se puede o no hacer con la prueba como profesores; una de las principales críticas que surgen y expresan a través de los medios de comunicación es la resistencia de un sector de profesorado que se niegan a la prueba por considerarla un instrumento de control, vigilancia o castigo. Estas creencias se pueden minimizar a través de un método sistemático de información acerca de las posibilidades de la prueba, en una forma transparente y demostrando la utilidad de los resultados o la intención de hacer que resulten.

Los resultados que no están concebidos para ese fin se pueden utilizar para generar reportes para los maestros, como se muestra en el siguiente reporte para los profesores.

INTRODUCCIÓN	Resultados	El desempeño en comunicación	El desempeño en matemáticas
<p><b>PRUEBA ENLACE MEDIA SUPERIOR 201X</b></p> <p><b>INFORME DE RESULTADOS</b></p> <p><b>Considere que la prueba...</b></p> <ul style="list-style-type: none"> <li>✓ está diseñada sólo para evaluar las disciplinas básicas de comunicación y matemáticas y conocer el nivel de dominio de los alumnos en estas para saber qué pueden realizar o qué falta por hacer.</li> <li>✓ proporciona las preguntas que presentaron los alumnos y el proceso mental que se realiza en cada una de ellas.</li> <li>✓ Está diseñada para ser una herramienta que ayude a los profesores a diseñar estrategias de aprendizaje en ciertos temas donde pueden ayudar a los alumnos a mejorar sus procesos.</li> <li>× <b>no</b> está hecha para dar conclusiones de toda la enseñanza media superior, los subsistemas, las escuelas o los estados.</li> <li>× <b>no</b> está hecha para influir en las calificaciones ni para penalizar.</li> <li>× <b>no</b> se diseñó para hacer comparaciones entre alumnos o entre escuelas.</li> <li>× <b>no</b> informa si el plantel o maestros son mejores o peores respecto a otra escuela.</li> <li>× <b>no</b> está concebida para penalizar o castigar a partir de los resultados.</li> </ul>			

## Los niveles de dominio de la prueba

La prueba proporciona un reporte a cada alumno indicándole el nivel de dominio que obtuvo, cada nivel tiene una serie de descripciones, que son las mismas presentadas aquí.

El siguiente gráfico indica el porcentaje de alumnos del plantel en cada nivel de dominio; en la parte inferior de cada uno se describen las descripciones de lo que son capaces de hacer en ese nivel.

La calificación se realizó con 350 alumnos.



### Más sobre los niveles de dominio

Son jerárquicos e inclusivos, es decir, conforme se avanza en nivel, se incluye al anterior.

No tienen impacto en las calificaciones, solo es un diagnóstico.

Permiten reconocer cuáles son los pasos que siguen al revisar otros niveles.

Introducción	Resultados	El desempeño en Comunicación	El desempeño en Matemáticas		
<b>PRUEBA ENLACE MEDIA SUPERIOR 201X</b> <b>INFORME DE RESULTADOS PARA PROFESORES</b>					
<p>La siguiente sección es el resultado que el plantel obtuvo en cada reactivo, con el proceso que representa; para ver la pregunta dé clic en cada uno de ellos.</p>					
PROCESO DE CADA PREGUNTA		Número de pregunta	Porcentaje de respuesta correctas		
			Hasta el 40	De 40 a 60	Más del 60
<b>TEXTO APELATIVO</b>					
Identificar el asunto central que se plantea en una carta formal.		001			
Distinguir la frase de despedida que sea congruente con el contenido de una carta formal.		002			
Distinguir dos elementos de información que compitan con otros semejantes y enlistados en un texto.		003			
Identificar el remitente de una carta formal.		004			
Identificar el responsable de solucionar un problema planteado en una carta formal.		005			
<b>TEXTO ARGUMENTATIVO</b>					
Localizar dos elementos de información vinculados y presentados de manera explícita en el texto.		006			
Identificar dos explicaciones presentadas en un artículo de opinión.		007			
Identificar el asunto central que trata el autor del texto en un párrafo específico.		008			
Identificar el tipo de relación que existe entre dos ideas presentadas en un texto.		009			
Identificar la frase que sintetiza el contenidos de un artículo de opinión		010			
Localizar un elemento de información relevante que se presenta de manera explícita en un artículo de opinión.		011			
Localizar tres elementos de información vinculados y presentados de manera explícita en un texto.		012			
Identificar dos explicaciones utilizadas para apoyar un argumento del autor de un texto.		013			
Identificar un contraargumento que presenta el autor de un artículo de opinión.		014			
Identificar tres recursos discursivos empleados por el autor de un texto para apoyar su punto de vista.		015			
Inferir cuál es la postura del autor de un artículo de opinión.		016			
Distinguir entre un hecho y una opinión que se presentan en un párrafo.		017			
Identificar la estructura en que se presenta el contenido de un artículo de opinión.		018			
Identificar una premisa o una conclusión derivadas de un texto.		019			
<b>TEXTO NARRATIVO</b>					
Identificar el sentido de un enunciado connotativo.		051			
Identificar el ambiente en el que están ocurriendo los hechos en un texto narrativo.		052			
Identificar el motivo por el que el personaje de un texto narrativo actuó de determinada manera.		053			
Identificar el personaje de un texto narrativo al que corresponda una descripción física.		054			
Identificar el enunciado que represente el contenido de un fragmento de un relato.		055			

Identificar el enunciado que expresa la acción principal en un fragmento de un relato.	056	
Identificar dos hechos que se relacionen casualmente a lo largo de un texto.	057	
Identificar tres acciones realizadas por un mismo personaje de un texto narrativo.	058	
Identificar la relación entre la intención del autor y la forma de un relato.	059	
Identificar el sentido con que se utiliza una analogía, metáfora u otra figura retórica.	060	
Identificar la forma en que el autor relaciona las acciones de una historia para lograr el propósito comunicativo.	061	
Identificar los elementos de un relato que retoma el autor de una reseña crítica del mismo.	063	
Identificación de una acción que corresponda con las características psicológicas de un personaje de un texto narrativo.	064	
Identificar el tema central de un texto narrativo.	065	
<b>TEXTO EXPOSITIVO</b>		
Identificar dos elementos de información ubicados a lo largo de un artículo de divulgación.	096	
Identificar tres elementos de información relacionando una tabla y el artículo de divulgación que la contiene.	097	
Identificar un elemento de información relacionando el contenido de un texto, una tabla y su respectivo glosario.	098	
Identificar el tema central de un párrafo en un artículo de divulgación científica.	099	
Identificar el enunciado que sintetiza un apartado de un texto.	100	
Identificar el organizador gráfico que representa la estructura de un artículo de divulgación.	101	
Identificar una acepción de una palabra utilizada por el autor de un texto.	102	
Identificar el propósito de la tabla que se incluye en un texto.	103	
Identificar la relación entre las ideas principales de dos párrafos.	104	
Identificar la razón por la que son diferentes dos fenómenos o hechos descritos en un artículo de divulgación científica.	105	
Identificar la solución a un problema teniendo en cuenta el contenido global de un texto.	106	
Identificar un término utilizado en un artículo de divulgación científica que corresponde a una definición.	107	
Identificar dos datos relacionados que estén incluidos en la tabla que acompaña un artículo de divulgación científica.	108	
Identificar el propósito comunicativo del autor de un artículo de divulgación científica.	109	
Identificar una conclusión derivada de un texto y relacionada con información externa al mismo.	110	
<b>ALGUNAS RECOMENDACIONES GENERALES PARA APROVECHAR LOS RESULTADOS</b>		
<p>Lea detenidamente cada proceso, analice y reflexione sobre él y vea el reactivo.  Los procesos actúan en conjunto ante un texto, cada proceso complementa a los demás y está asociado a un ejercicio de extracción, interpretación o reflexión y evaluación.</p> <p>Si quiere tener más información de cómo se estructura la prueba y sus contenidos y más información de cómo se estructura la prueba y cómo utilizar los resultados en su formación acceda al <a href="#">Manual de ENLACE EMS para el profesorado</a></p>		

## Consideraciones al diseñar los reportes para los maestros

- Proporcionar al maestro las interpretaciones adecuadas de la prueba. En la primera sección se proporcionan interpretaciones importantes y generales que expresan para qué es la prueba y también las malas interpretaciones, una de las más comunes y evidentes es que los profesores serán despedidos o

penalizados con la prueba; así como la comparación con otras escuelas para estigmatizar.

- Proporcionar información que apoye las interpretaciones de los resultados. Los niveles de dominio no son una descripción concreta de lo que puede lograr el sustentante, son varias actividades y deben describirse como tal al describir el dominio.
- Entregar información que apoye a las interpretaciones. Si se afirma que la prueba está diseñada para conocer el nivel de dominio y lo que pueden o no hacer lo sustentantes, proporcionar el proceso que subyace a cada reactivo es muy importante. De esta manera se proporciona información a los maestros acerca de la capacidad que tienen los alumnos de responder a una acción, más que presentar el reactivo y hacer inferencias sobre él.
- Redactar información breve, sencilla y directa.
- Cambiar la manera en la que el profesor puede revisar el reporte, no por medio del número de reactivo y un porcentaje de respuesta, como lo está actualmente; sino por uno basado en el proceso que implica contestar el reactivo, en conjunto con la probabilidad de respuesta. Esto hace, que antes de ver un reactivo, se analiza la tarea para realizarla.
- Hacer que la información dé respuesta a los objetivos de generar información diagnóstica de los alumnos, proveer elementos que contribuyan a la mejora de sistema educativo e información útil para el plantel y los profesores.
- Redactar el texto que precede a los gráficos fácil de leer y para apoyar su interpretación.

- Los gráficos resaltan los principales resultados para los profesores: el nivel de dominio de los alumnos en ese año, con las descripciones del dominio diferenciadas en columnas con una división clara y en los anteriores.
- Agregar columnas que indiquen el porcentaje de las respuestas que se contestan correctamente. En realidad se pensó en un formato que reportara el número exacto, pero la prueba lo reporta en tres categorías y se respetó eso. La mejor manera fue una barra horizontal que fuera completando cada categoría, en lugar de puntos o marcas o sólo rellenar el cuadro correspondiente. La lógica de las barras horizontales es hacer que visualmente se cree un esquema que no necesite demasiada explicación y que no compita con las otras barras, como puede suceder con otros elementos como los puntos o marcas.
- La información complementaria mostrada al final de las secciones es para proporcionar algunas consideraciones sobre los datos, o como en el caso de las recomendaciones de la última sección, proporcionar algunas recomendaciones de qué hacer con esos datos e invitar a los maestros a ver una sección con más datos.

### **Reportes para autoridades educativas**

Es innegable que la información de una evaluación es muy importante para las autoridades educativas (INEE, 2004), que la información es utilizada y fomentada por ellos; y es sumamente importante que conozcan las fortalezas, beneficios o limitaciones de las pruebas educativas, aunque proporcionar tal información puede ser una tarea muy compleja, lo que hace la comunicación de conceptos clave difícil (Sireci, 2012). Autores como Stephen Sireci, de la Universidad de Massachusetts Amherst abogan por una estrategia de comunicación que involucre conceptos claves como validez, error de medida, equidad, puntos de corte,

comparación de puntajes, versiones de la prueba y proporciona sugerencias para comunicar de manera más eficiente tales conceptos, como el uso de metáforas.

Los reportes para los administradores no son sencillos, tienen que estar pensados en función de su utilidad para ayudar a la toma de decisiones, aunque la naturaleza de sus responsabilidades provoca que en ocasiones las decisiones se tomen con poca evidencia (Zapata, 2011); con el fin de hacer reportes útiles para este sector es necesario hacer una revisión de las funciones que desempeñan.

Los datos que existen de la prueba Enlace EMS para poblaciones ajenas a los estudiantes son tablas en formato de Excel en donde se reportan los niveles de dominio por año y reportados en diferentes categorías:

- Por grado de marginación (utilizando los criterios del Consejo Nacional de Población (CONAPO) que usa cinco categorías de marginación (muy alta, alta, media, baja y muy baja).
- Por el tipo de sostenimiento, es decir, la manera en la que obtiene recursos la institución, de tipo público o privado.
- Por la modalidad del plan de estudios del plantel y el tipo de subsistema, clasificados como bachillerato general, bachillerato tecnológico o modalidad técnica.

Los datos siempre se presentan clasificados para comunicación o matemáticas y están divididos por pestañas de acuerdo con las clasificaciones; la siguiente tabla es un ejemplo de una sección de una hoja para la modalidad de bachillerato general en la sección de comunicación.

MODALIDAD BACHILLERATO GENERAL																				
COMUNICACIÓN																				
PORCENTAJE DE ALUMNOS DEL ÚLTIMO GRADO EN CADA NIVEL DE DOMINIO																				
NIVEL DE DOMINIO																				
	INSUFICIENTE					ELEMENTAL					BUENO					EXCELENTE				
ENTIDAD	2008	2009	2010	2011	2012	2008	2009	2010	2011	2012	2008	2009	2010	2011	2012	2008	2009	2010	2011	2012
<b>Aguascalientes</b>	12.7	10.5	7.4	12.4	10.9	33.5	25.9	24.4	27.7	30.1	45.9	49.4	50.5	48.0	46.6	7.9	14.2	17.7	11.9	12.4
<b>Baja California</b>	7.4	9.1	6.3	10.6	6.9	29.0	26.0	24.6	25.3	24.7	53.6	51.3	52.8	49.4	54.2	10.0	13.6	16.3	14.7	14.2
<b>Baja California Sur</b>	7.2	14.4	8.3	12.3	12.8	28.0	30.3	28.6	28.7	32.1	53.1	46.1	51.1	48.5	45.7	11.8	9.1	11.9	10.5	9.3
<b>Campeche</b>	5.0	10.6	7.9	8.5	8.3	25.2	31.6	30.1	28.8	29.9	59.4	50.2	52.2	51.5	51.9	10.5	7.7	9.9	11.2	9.9

Por el espacio visual disponible, solo se muestra esta sección; el final de la tabla es el número de alumnos evaluados para cada año (cinco columnas más).

Así se componen todas las tablas y a partir de éstas se elaboraron distintas propuestas para realizar los reportes, se utilizó esta información para hacer el reporte, que en realidad se da para cada entidad por separado, para no hacer comparaciones entre estados. Los reportes se muestran a continuación.

## Los niveles de dominio de la prueba

La prueba proporciona un reporte a cada alumno indicándole el nivel de dominio que obtuvo, cada nivel tiene una serie de descripciones, que son las mismas presentadas aquí.

 <b>Insuficiente</b>	 <b>Elemental</b>	 <b>Bueno</b>	<b>Excelente</b>
<p>Identifica elementos que se encuentran de manera explícita en textos narrativos y expositivos. Esos elementos pueden ser acciones, hechos, episodios, personajes o sus características.</p> <p>Realiza inferencias sencillas sobre las acciones de los personajes y establece relaciones entre dos o más elementos.</p> <p>Identifica si la estructura de algunas partes del texto es adecuada al contenido que presenta.</p>	<p>Utiliza e integra diferentes partes de un texto.</p> <p>Reconoce la idea central y comprende relaciones. Por ejemplo relaciones que expresan el problema y su solución, la causa y el efecto, o una comparación y su contraste.</p> <p>Infiere el significado de las palabras, así como la relación entre párrafos e ideas.</p> <p>Reconoce la postura del autor.</p> <p>Relaciona la información que se presenta en el texto y la que se encuentra en tablas.</p>	<p>Relaciona elementos que se encuentran a lo largo del texto y en glosarios.</p> <p>Comprende el texto de forma completa y detallada, y sintetiza su contenido global.</p> <p>Infiere problemas que expresan el problema y su solución, la causa y el efecto, una comparación y su contraste.</p> <p>Establece relaciones entre la postura del autor y la información que apoya su punto de vista.</p> <p>Reconoce la función que cumplen las tablas en los textos expositivos.</p> <p>Evalúa la estructura del texto en relación con su contenido.</p>	<p>Hace inferencias complejas para construir una interpretación global del texto.</p> <p>Comprende la información contenida en tablas y esquemas y las relaciones con el contenido del texto.</p> <p>Establece relaciones entre argumentos y contraargumentos.</p> <p>Analiza si la organización, las expresiones y los recursos que utiliza el autor son adecuados al tipo de texto y a su destinatario.</p>

Los reportes estatales contienen gráficos que indican el porcentaje de alumnos de cada nivel de dominio por cada grado de marginación (muy alto, alto, medio, bajo o muy bajo) tipo de sostenimiento de los planteles (privados o públicos) o modalidad (bachillerato general, bachillerato tecnológico o tecnológico).

La prueba no está diseñada para hacer comparaciones entre estados o subsistemas, por esa razón solo se reportan los datos que se pueden conformar en una categoría.

Cada reporte indica el número de alumnos con los que se realizó la prueba, considérelolo para complementar sus decisiones.

Considere los contextos en los que se genera el reporte, es decir, las características de cada categoría, (grado de marginación, tipo de sostenimiento y modalidad) para sustentar la información que maneja.

### Más sobre los niveles de dominio

Son jerárquicos e inclusivos, es decir, conforme se avanza en nivel, se incluye al anterior.

No tienen impacto en las calificaciones, solo son diagnósticos.

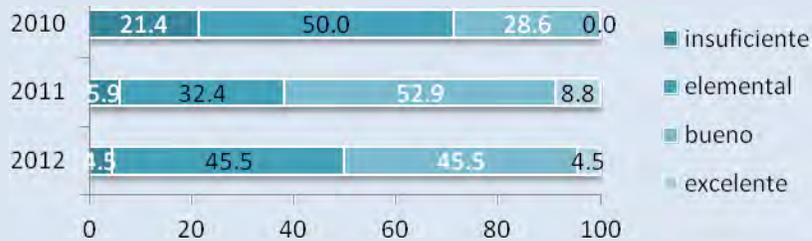
Permiten reconocer cuáles son los pasos que siguen al revisar otros niveles.

## PRUEBA ENLACE MEDIA SUPERIOR 201X

### INFORME DE RESULTADOS POR GRADO DE MARGINACIÓN PARA CADA NIVEL DE DOMINIO DEL ESTADO DE XXXXX

Se muestra el desempeño de los alumnos a nivel estatal clasificados por año, agrupando a las escuelas de acuerdo con cinco grados de marginación (clasificación de la CONAPO), para ver los detalles de cada grado, de clic en el título del cuadro.

#### Porcentaje de alumnos de grado de marginación muy alto



Año	Evaluados
2010	14
2011	34
2012	44

#### Porcentaje de alumnos de grado de marginación alto



Año	Evaluados
2010	3,821
2011	4,342
2012	4,996

#### Porcentaje de alumnos de grado de marginación medio



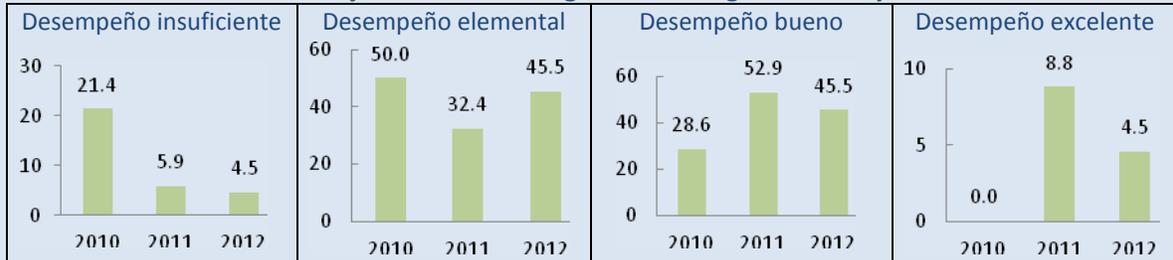
Año	Evaluados
2010	3,336
2011	3,731
2012	3,959

## PRUEBA ENLACE MEDIA SUPERIOR 201X

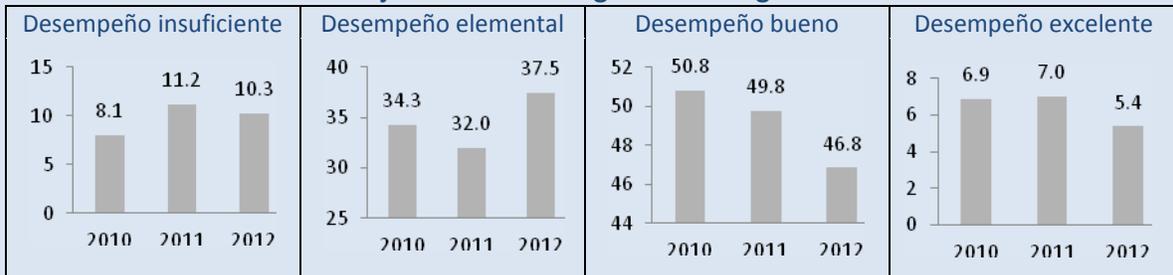
### INFORME DE RESULTADOS POR GRADO DE MARGINACION PARA CADA NIVEL DE DOMINIO DEL ESTADO DE XXXXX

Es necesario revisar las características de cada grado de marginación acorde con la CONAPO. Al revisar cada gráfico tome en cuenta que las escalas no siempre inician en cero se ajustan a los datos de su grupo.

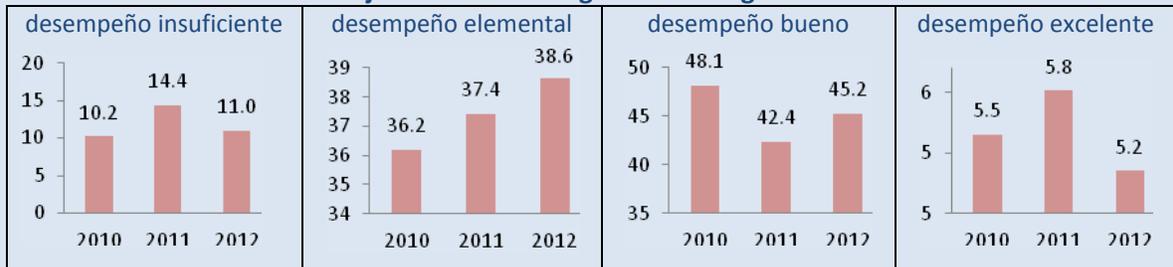
#### Porcentaje de alumnos de grado de marginación muy alto



#### Porcentaje de alumnos de grado de marginación alto



#### Porcentaje de alumnos de grado de marginación medio



#### Alumnos evaluados por grado de marginación

Año	Muy alto	Alto	Medio	Bajo	Muy bajo
2010	14	3,821	3,336	19,803	80,403
2011	34	4,342	3,731	19,976	82,595
2012	44	4,996	3,959	22,727	86,624

## Consideraciones al diseñar los reportes para los maestros

- No hay reportes para las autoridades educativas. Una de las afirmaciones de la prueba es proporcionar información a las autoridades educativas.
- Proporcionar datos de tablas, pero en formatos de lectura visualmente más comprensibles y atractivos. La lectura de datos en formato gráficos proporciona la misma información, pero en una forma más comprensible dada la forma en la que están los datos.
- Proporcionar gráficos que permitan una comparación de los datos. Es un cambio de formato, las tablas proporcionan la misma información, pero en una forma que es visualmente más comprensible para apreciar cifras que se pueden comparar entre ellas.
- Hacer dos formatos de gráficos, dependiendo de las consideraciones de los interesados. Uno en forma de gráficos de columnas en forma vertical, que permite la comparación entre datos del mismo grupo; el otro en forma de barras horizontales que no permiten tal comparación sino que agrupan a toda la población y marcan la proporción de cada categoría en relación a un total; esto permite apreciar el conjunto en un año pero no una comparación evidente entre distintos años. Cada uno de los formatos atiende a dos intereses diferentes: la comparación o no de datos.
- Claridad en los datos gráficos. En el formato de gráfico de barras horizontales se muestra cada grupo de datos, con el nivel de dominio, el porcentaje y la cantidad de población por año. Todas las barras tienen los mismos colores, para no discriminar visualmente o indicar con los colores que los grupos son diferentes. Es la información que se proporciona en las tablas de Excel.

- Contraste en los datos gráficos. Los gráficos con columnas permiten hacer comparaciones, se representan por categoría de desempeño, año y grupo de marginación, en cada grupo horizontal hay cuatro gráficos lo más sencillos posibles donde se aclara qué grupo puede ser comparable. La división en cuadros agrupa o que se puede comparar entre si, aunque todos los gráficos horizontales usan el mismo color, para indicar que forman parte de un grupo (grado de marginación), pero distinto de otros grados de marginación que están en otros colores.
- Proporcionar información breve sobre el tipo de datos que se verán a continuación. El primer reporte menciona los niveles de dominio, su composición, límites y consideraciones generales. No se satura con información técnica y es muy limitada la información que se proporciona.
- Separar los datos para un sector. Los datos son estatales, en lugar de presentar datos en tablas de todas las entidades, lo que es acorde con las limitaciones mencionadas en los manuales técnicos que afirman la imposibilidad de reportar datos entre estados, en estos formatos no es posible la comparación estatal, o al menos su comparación requiere procesar información.

Los trabajos anteriores tuvieron la intención de mostrar los reportes e información existente, y sus posibilidades de modificación acorde con reglas claras y precisas. En los casos donde no hay reportes, como los de profesores y directivos, se crearon uniendo datos públicos de Enlace con las reglas de cómo hacer esos reportes.

## **CAPÍTULO VI. CONCLUSIONES Y RECOMENDACIONES**

En este apartado se exponen las reflexiones y enseñanzas derivadas de la creación de la propuesta y su aplicación a la prueba de Enlace de Educación Media Superior.

## **Conclusiones**

La creación de reportes de resultados de pruebas estandarizadas no es una tarea simple, pero en muchas pruebas es un proceso que recibe poca atención de los profesionales en su diseño; debido a que muchos procesos reciben la mayor parte de la atención y estrategia técnica el reporte de resultados, que es mucho menos técnico y con información para un usuario no experto, no suele ser visto o tratado con el mismo interés. Esto da como resultado que muchos reportes no estén elaborados de la misma forma o calidad y que el nivel o tipo de información sea diferente incluso entre aquellos de diferentes años.

En México las pruebas estandarizadas de aplicación a gran escala en el ámbito educativo son pocas con un impacto limitado con una pobre difusión, pero las pruebas que tienen mayor promoción mediática como la de Enlace son constantemente criticadas, politizadas y suelen ser objeto de juicio desde diferentes ámbitos, pero raramente desde el técnico. Esto no es extraño, la reticencia a instrumentos de evaluación es bien conocida en la mayor parte del mundo donde se aplican y los ciudadanos mexicanos suelen tener reserva hacia los instrumentos, particularmente si provienen de instituciones gubernamentales. Ante esto, una obligación de quienes se dedican a elaborar pruebas estandarizadas es informar a sus usuarios qué esperar de ellas y cómo utilizar la información que generan, y el reporte de resultados es el escenario perfecto para hacer esto. La intención de este trabajo fue articular información sobre los elementos que se deben esperar en un buen reporte de resultados de una prueba estandarizada y aplicarlos a una prueba.

Uno de los propósitos de este trabajo fue proporcionar un andamiaje teórico sobre los elementos que se necesitan conocer para entender cómo se estructura un reporte de resultados; esto generó un capítulo que resume varios aspectos de la elaboración de pruebas con un enfoque poco técnico y una mirada amplia hacia la forma en la que se conforma una prueba estandarizada. También uno de los propósitos fue reconocer las pruebas estandarizadas nacionales que se aplican en México en el ámbito de evaluación del desempeño estudiantil para mostrar la oferta actual. Para esto, parte del trabajo fue realizar una búsqueda de información en idioma español e inglés.

Así, uno de los principales descubrimientos fue encontrar que hay poca información útil en español o al menos en la región de América Latina concerniente a la elaboración de reportes de resultados en pruebas estandarizadas; no es una sorpresa si analizamos con cuidado la cantidad de pruebas existentes y su duración en esta región. La mayor parte de los datos se encuentran en documentos de lengua inglesa, lo que no es un problema en sí mismo, excepto por el hecho de que varias recomendaciones que realizan están enmarcadas bajo aspectos legales o condicionales de sus países de origen y hay que tenerlo en cuenta constantemente al realizar recomendaciones basadas en ellos.

Por lo anterior, el propósito de cómo valorar reportes estuvo dirigido por la búsqueda de estándares que no estuvieran relacionados con decretos, actas o legislaciones nacionales, y ésta fue una actividad poco usual porque se da por hecho que todos los estándares son útiles, pero no se atiende a su naturaleza o propósito legal. La elección de los estándares para evaluar reportes de resultados estuvo en función de ser de documentos que fueran referentes internacionales, ese es su valor y su funcionalidad. Pero los estándares no mencionan la manera en la que se elabora bien un reporte ni sus componentes detallados y por ello se volvió necesaria la búsqueda de diversas investigaciones que hayan realizado esa tarea que es monumental por la cantidad de recursos y gran capacidad técnica

para hacerlo adecuadamente; esas investigaciones resumían en forma de recomendaciones los aciertos y defectos en los reportes de resultados, incluso llegando a hacer listas sumarias en forma de sugerencias breves y muy útiles. Pero también se encontró que un reporte de resultados contiene elementos no tan obvios, como la presentación adecuada en función de imágenes, colores, texto o gráficos y en atención a diferentes audiencias.

La experiencia de aplicar la propuesta de nuestro instrumento como diagnóstico de una prueba fue revelador, aunque parezca obvio: si no hay una declaración clara y concreta de lo que se evalúa, es difícil saber qué esperar de un reporte de resultados; si los documentos de acceso público no detallan cómo interpretar y utilizar los resultados de un reporte la utilidad de éstos será limitada. Si bien la propuesta que generamos no se concibe como un diagnóstico de la prueba, sí ayuda a evaluar componentes que no son claros o bien definidos. En este punto es importante aclarar que la intención no es evidenciar las fallas en las que incurre una prueba, sino poner en claro que ciertos elementos no considerados inicialmente se ven reflejados en defectos del reporte.

Hay diversas conclusiones que se pueden mencionar y que son factores que deben tomarse en cuenta para elaborar reportes de resultados; son de diferente nivel de impacto y algunos tienen que ver directamente con sus diseñadores y otros incluso con el país donde se encuentren.

- Es necesario que se definan las reglas para elaborar pruebas a nivel nacional. No hay documentos que normen o regulen su producción o diseño, esto puede provocar que se elaboren pruebas sin asegurar criterios de gran calidad técnica.
- Con relación al punto anterior, es necesario crear estándares que delimiten lo que se espera de reportes de resultados de pruebas estandarizadas.
- Los reportes de resultados deben concebirse de manera detallada desde el inicio de la prueba.

- Para saber qué esperar de un reporte de resultados los propósitos de la prueba deben estar perfectamente definidos.
- Los reportes de resultados, como parte de su diseño, deben involucrar a los agentes que impactan, ya sea como grupo de trabajo o como un grupo de enfoque al que se apliquen las pruebas.
- Las audiencias importan, así que cada reporte se debe concebir en función de cada una.
- Un buen reporte debe complementarse de documentos que ayuden a interpretar sus contenidos de forma más amplia.
- La aplicación e interpretación de un instrumento diagnóstico involucra a diferentes expertos, no es una tarea solitaria.

Los puntos anteriores, al no cumplirse de forma evidente en las pruebas evidencian que hace falta un gran camino por recorrer respecto a éstas como instrumentos adecuados para ofrecer información a los usuarios y por ende, donde los reportes sean insumos útiles a las diferentes audiencias.

### **Limitaciones de la propuesta y sugerencias para trabajos futuros**

Hay algunas limitaciones en la propuesta:

- Aunque se ha mencionado de manera constante, no es una propuesta para valorar toda una prueba, sólo el componente que atañe al reporte de resultados es puesto en juicio, la valoración de una prueba es compleja y sumamente técnica.
- No es un diseño para hacer un reporte de resultados desde la etapa inicial de una prueba. Se concibe como un apoyo a la mejora de un reporte ya planificado.

- No está hecho para mejorar los procesos anteriores al reporte. Si bien el reporte es un indicador del estatus de diferentes etapas previas, no proporciona suficiente información para corregir procesos anteriores a éste.
- No reemplaza las acciones de revisión o diagnóstico en las diferentes fases de la prueba.
- Es la propuesta de una persona; no hay soluciones individuales para problemas colectivos; debe ser un tema de discusión y propuestas de diferentes expertos, fuentes e interesados.
- No está avalada por agencias o instituciones de evaluación educativa, la propuesta es resultado de experiencia práctica en el diseño y elaboración de pruebas.
- Es limitada en recursos financieros y técnicos, se debe valorar con diferentes audiencias y probar su utilidad de manera documentada.

Por otro lado, hay mucho por realizar, entre otras propuestas; quien se interese por el tema, debe abordar:

- La existencia de leyes sobre evaluación educativa y diseño de pruebas.
- La definición de estándares con referentes nacionales.
- La creación de lineamientos técnicos para crear reportes de resultados.
- El análisis de reportes de resultados de las pruebas nacionales.
- El uso que se les ha dado de forma histórica a los resultados de las pruebas.
- El mal uso que se ha dado a los resultados de pruebas estandarizadas.
- El involucramiento de diferentes agentes de la evaluación.
- La importancia de verificar si se usa o no el resultado de una prueba.
- La optimización de los resultados por personas en todo el país de diferentes niveles sociales o educativos.
- El uso de datos de contexto o de aquellos que añaden valor a los reportes de resultados de pruebas estandarizadas.

Si bien el proceso es complejo, de ninguna manera es limitado, hace falta más interés y responsabilidad por adentrarse en estos temas, pero es muy importante hacerlo: las pruebas estandarizadas y sus reportes de resultados nos conciernen a todos, expertos o no, porque al final, el sistema educativo está conformado por todos nosotros, y no podemos ser indiferentes a eso.

## **REFERENCIAS**

## REFERENCIAS

- (2003/2004, Invierno). Michael Scriven on the differences between evaluation and social science research. *The evaluation exchange IX (4)*. pp. 2-3. (Entrevista de Julia Coffman)
- Aboites, H. (2013, 22 de junio). La nueva crisis de la evaluación. *La jornada*. Recuperado el 06 de julio de 2013, de <http://www.jornada.unam.mx/2013/06/22/opinion/017a1pol>
- Acuerdo número 1/CD/2009 del Comité Directivo del Sistema Nacional de Bachillerato*. (2011, viernes 19 de agosto). México.: *Diario Oficial*.
- Acuerdo número 12/CD/2011 del Comité Directivo del Sistema Nacional de Bachillerato*. (2011, viernes 19 de agosto). México.: *Diario Oficial*.
- Acuerdo número 444 por el que se establecen las competencias que constituyen el marco curricular común del Sistema Nacional de Bachillerato. (2008, martes 21 de octubre). México.: *Diario Oficial*. (1a. Sección).
- Alkin, M.C. & Christie, C. A. (2004). An evaluation theory tree. En Alkin, M. C. (Ed.) *Evaluation roots: tracing Theorists' views and influences*. Thousand Oaks, CA, EE. UU.: Sage Publications.
- Álvarez, L. & De la O, C. (2005, mayo-agosto). Evaluación y certificación de competencias laborales en México: el caso de las dependencias del gobierno federal. *Contaduría y administración*, 216, 13-34.
- American Educational Research Association & American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC.: American Educational Research Association.
- Arregui, P. (Ed.). (2006). *Sobre estándares y evaluaciones en América Latina*. Santiago, Chile.: Programa de Promoción de la Reforma Educativa en América Latina y el Caribe. (PREAL).
- Aschbacher, P. R. & Herman, J. L. (1991). *Guidelines for effective score reporting*. (Reporte técnico de la CSE 326). Los Angeles, CA.: National Center for Research on Evaluation, Standards, and Student Testing.

- Asociación Nacional de Universidades e Instituciones de Educación Media Superior. (ANUIES). El 26 de junio de 2013 en <http://www.anui.es.mx/content.php?varSectionID=2>
- Backhoff, E. & Solano, F. (2003, diciembre) *Tercer estudio internacional de matemáticas y ciencias naturales (TIMSS): resultados de México en 1995 y 2000*. (Informe técnico). México.: Instituto Nacional para la Evaluación de la Educación.
- Becerril, A. & Ballinas, V. (2014, 6 de marzo) La prueba Enlace no desaparece, afirma Chuayffet; “volverá a aplicarse en 2015” *La jornada*. Recuperado el 12 de marzo de 2014, de <http://www.jornada.unam.mx/2014/03/06/sociedad/050n1soc>
- British Columbia, Ministry of education. (2001). *Interpreting and communicating British Columbia foundations skills assessment results*. Columbia Británica, Canada.: Autor.
- Camara, W. J. & Shaw, E. J. (2012, verano). The media and educational testing: in pursuit of the truth or in pursuit of a good story? *Educational measurement: issues and practice*, 31 (2), 33-37.
- Centro Nacional de Evaluación para la Educación Superior (Ceneval). (2012). *Manual técnico ENLACE media superior 2008-2010*. México.: Autor.
- Centro Nacional de Evaluación para la Educación Superior (Ceneval). (2013). *Manual técnico ENLACE media superior 2011-2012*. México.: Autor.
- Centro Nacional de Evaluación para la Educación Superior. (Ceneval). (s.f.). *Manual para docentes y directivos 2012: ENLACE media superior 2012*. México.: Autor.
- Cisneros-Cohernour, E. J. & Grayson, T. E. (2009). Exemplary case: implementing large-scale assessment of education in Mexico. En Ryan, K. E. & Cousins, J.B. (Eds.). *The SAGE international handbook of educational evaluation*. (pp. 231-251). Thousand Oaks, CA, EE. UU.: Sage Publications.
- Cleveland, W. S. & McGill, R. (1984, septiembre). Graphical perception: theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79 (387), 531-554.
- Cohen, R. J. & Swerdlik, M. E. (2006). *Pruebas y evaluación psicológicas. (Sexta edición)*.

México.: Mc. Graw Hill.

Comité conjunto de estándares para la evaluación educativa. (1998). *Estándares para la evaluación de programas* (Solabarrieta J. Trad.). Bilbao, España.: Ediciones mensajero. (Trabajo original publicado en 1994).

Cronbach, J.L. (1998). *Fundamentos de los test psicológicos*. Madrid, España.: Biblioteca Nueva. (Trabajo original publicado en 1984).

Davidson, E. J. (2005). *Evaluation methodology basics: the nuts and bolts of sound evaluation*. Thousand Oaks, CA, EE. UU.: Sage Publications.

Deng, N., & Yoo, H. (2009). *Resources for Reporting Test Scores: A Bibliography for the Assessment Community*. Elaborado para el National Council on Measurement in Education. Recuperado el 12 de mayo de 2013 de <http://ncme.org/linkservid/98ADCCAD-1320-5CAE-6ED4F61594850156/showMeta/0/>

Dijk, S. V. (2009, abril). Breve revisión histórica de la evaluación en México. *Educatio: Revista regional de Investigación Educativa* 7, 51-72.

Downing S. M. & Haladyna T.M. (Eds.). (2012). *Manual para el desarrollo de pruebas a gran escala*. (2 vols.). (Centro Nacional de Evaluación para la Educación Superior Trad.). D.F., México.: Centro Nacional de Evaluación para la Educación Superior. (Trabajo original publicado en 2006).

Dunn, D. D. (1999). Mixing Elected and Nonelected Officials in Democratic Policy Making: Fundamentals of Accountability and Responsibility, en Manin, Bernard, Adam Przeworski y Susan C. Stokes, *Democracy, Accountability and Representation*, Cambridge.: Cambridge University Press.

En Alkin, M. C. (Ed.). (2004). *Evaluation roots: tracing Theorists' views and influences*. Thousand Oaks, CA, EE. UU.: Sage Publications.

Fernández-Ballesteros, R. (2004): *Evaluación psicológica: Concepto, métodos y estudio de casos*. Madrid: Pirámide.

Forte, E. & ASR SCASS. (2002). *A guide to effective accountability reporting: designing*

*public reports that effectively communicate accountability, assessment and other quantitative education indicators in an easily understood format.* Washington, DC. EE.UU.: The Council of Chief State School Officers.

Frade, L. (2009). *La evaluación por competencias* (3a. ed.). México.: Laura Gloria Frade Rubio.

Frade, L. (2009). *Planeación por competencias* (2a. ed.). México.: Laura Gloria Frade Rubio

Froemel, J.E. (2009). La efectividad y la eficacia de las mediciones estandarizadas y de las evaluaciones en educación. *Revista Iberoamericana de Evaluación Educativa*, 2 (1), 10-28.

García, J. M. (2005, octubre-diciembre). El avance de la evaluación en México y sus antecedentes. *Revista mexicana de investigación educativa*, 10 (027), 1275-1283.

Georgia Department of Education. (2012). *Georgia high school graduation test: Interpretative guide for score reports. English language arts, mathematics, science, social studies.* Georgia, EE.UU.: Georgia Department of Education.

Gierl, M.J., Wang, C. & Zhou, Jiawen. (2008, febrero). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills in algebra on the SAT. *The journal of technology, learning, and assessment*, 6 (6), 1-53.

Gonzalez F.M. (2007). *Instrumentos de evaluación psicológica.* La Habana, Cuba.: Editorial Ciencias Médicas.

Goodman, D. P. & Hambleton, R.K. (2004). Student test score reports and interpretative guides: review of current practices and suggestions for future research. *Applied measurement in education*, 17 (2), 145-220.

Gibbons, W. M. (1992). Organization by design: some implications for structuring information. *Journal of technical writing and communication*, 22 (1), 57-75.

Grupo de Evaluación de las Naciones Unidas. (UNEG). (2005, 29 de abril). *Normas de evaluación en el sistema de las Naciones Unidas.* Recuperado el 2 de abril de 2013 de [http://www.uneval.org/documentdownload?doc\\_id=22&file\\_id=129](http://www.uneval.org/documentdownload?doc_id=22&file_id=129).

- Guba, E. G. & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Thousand Oaks, CA, EE. UU.: Sage Publications.
- Harvard Family Research Project. (Invierno 2005). *The evaluation exchange*, XI (3).
- Hernandez, G. (2013, 18 de Julio). Anuncia Chuayfett fin de la prueba Enlace. *Proceso.com.mx* Disponible en <http://www.proceso.com.mx/?p=347774>
- House, E. (2005, otoño). The many forms of democratic evaluation. *The evaluation exchange*, XI (3), 7.
- House, E. R. (1994). *Evaluación, ética y poder*. Madrid, España.: Ediciones Morata. (Trabajo original publicado en 1980).
- INEE 2013. *Panorama educativo de México*. Disponible en <http://www.inee.edu.mx/index.php/publicaciones/informes-institucionales/panorama-educativo>
- Instituto Vasco de Evaluación e Investigación Educativa. (2005). *Evaluación internacional de matemáticas y ciencias: segundo informe de resultados TIMSS 2003* Bilbao, España.: Autor.
- Joint Committee on Testing Practices (JCTE). (2004). *Code of fair testing practices in education*. Washington, DC.: Autor. Recuperado el 15 de mayo de 2013, de <http://www.apa.org/science/programs/testing/fair-testing.pdf>
- Kerlinger, F. N. & Lee, H. B. (2002). *Investigación del comportamiento* (McGraw-Hill Trad.). D. F., México.: McGraw-Hill. (Trabajo original publicado en 2000).
- Leighton, J. P., Gierl, M. J. & Hunka, S. M. (2002, abril). The attribute hierarchy model for cognitive assessment. Presentado en la *Annual Meeting of the National Council on Measurement in Education (NCME)*. Nueva Orleans, Luisiana, EE.UU.
- López, M. J. (2013, 5 de junio). La prueba Enlace ¿mejora la educación? *La jornada*, Disponible en <http://www.jornada.unam.mx/2013/06/05/opinion/018a1pol>
- Lundgren, U. P. (2009). Evaluation and educational policymaking. En Ryan, K. E. &

- Cousins, J.B. (Eds.). *The SAGE international handbook of educational evaluation*. (pp. 501-509). Thousand Oaks, CA, EE. UU.: Sage Publications.
- Mark, M. (2005, verano). Evaluation theory or what are evaluation methods for? *The evaluation exchange*, XI (2), 2-3.
- Mark, M. M. & Greene J. C & Shaw I. F. (Eds.). (2006). The evaluation of policies, programs, and practices. En Shaw, I. F. & Greene, J. C. & Mark, M. M. *The SAGE handbook of evaluation*. (pp. 1-30). Inglaterra: Sage Publications.
- Mark, M. M. (2009). Evaluation, methods choices, and pathways to consequences: trying to make sense of how evaluation can contribute to sensemaking. En Ryan, K. E. & Cousins, J.B. (Eds.). *The SAGE international handbook of educational evaluation*. (pp. 55-73). Thousand Oaks, CA, EE. UU.: Sage Publications.
- Martín J. (2013, 15 de junio). Experiencia y evaluación en el trabajo de los docentes. *La jornada*, Disponible en <http://www.jornada.unam.mx/2013/06/15/politica/016a1pol?partner=rss>
- Martínez Arias, R. (1995). *Psicometría: Teoría de los tests psicológicos y educativos*. Madrid.: Síntesis.
- Martínez, F. (2001). Evaluación educativa y pruebas estandarizadas: elementos para enriquecer el debate. *Revista de Educación Superior*, 30 (120), 71-85.
- Martínez, F. (2005). *Sobre la difusión de resultados por escuela: cuaderno No. 15*. México.: Instituto Nacional para la Evaluación de la Educación.
- Martínez, F. (2006). PISA en América Latina: lecciones a partir de la experiencia en México de 2000 a 2006. *Revista de educación. (No. Extraordinario)*. pp.153-167.
- Martínez, R. (2006) La metodología de los estudios PISA. *Revista de educación. (No. extraordinario)*. pp.111-129.
- Martinic, S. (2008). Información, participación y enfoque de derechos. En Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (Ed.). *Reflexiones en torno a la evaluación de la calidad educativa en América Latina*. Santiago, Chile.: Autor.

- Mc Intire, S. A. & Miller, L. A. (2000). *Foundations of psychological Testing*. EE. UU.: McGraw-Hill.
- Méndez A. (s.f.). *Terminología pedagógica específica al enfoque por competencias: el concepto de la competencia*. Recuperado el 7 de mayo de 2013, de <http://redeca.uach.mx/concepto/Terminologia%20pedagogica%20especifica%20al%20enfoque%20por%20competencias.El%20concepto%20de%20competencia.pdf>
- Monroy, L. & Herrera, M. & García, P. (2009). *Cuestionarios de contexto: una herramienta medular para el desarrollo de investigaciones educativas*. México.: Centro Nacional de Evaluación para la Educación Superior.
- National Center for Education Statistics. (2012). *The nation's report card: writing 2011 (NCES 2012-470)*. Washington, DC.: Institute of Education Sciences, U.S. Department of Education.
- National Council on Measurement in Education. (1995). *Code of professional responsibilities in educational measurement*. Washington, DC.: Autor.
- National Education Goals Panel (NEGP). (1998). *Talking about tests: an idea book for state leaders*. Washington, DC. EE. UU.: Government Printing Office.
- Navarro, C. (2006). Las reformas educativas como reformas de estado: América Latina en las dos últimas décadas. En Lora, E. (Ed.). *El estado de las reformas del estado en América Latina*. Bogotá, Colombia.: Banco Interamericano de Desarrollo.
- New Mexico Public Education Department. (2012, abril). 2012 Interpretive guide to standards based assessment reports for schools and districts. Disponible en <http://www.ped.state.nm.us/assessmentaccountability/assessmentevaluation/sba/InterpretiveGuide2012.pdf>
- Niremberg, O. & Brawerman, J. & Ruis V. (2000). *Evaluar para la transformación: investigaciones en la evaluación de programas y proyectos sociales*. Argentina: Paidós.
- Ohio department of Education. (2008, primavera). Ohio graduation tests: interpretative guide educator reports. Ohio. EE.UU.: Autor.

- Opfer, V. D. (2006, enero y marzo). Evaluating equity: a framework for understanding action and inaction on social justice issues. *Educational policy*, 20 (1), 271-290.
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (1986). *Evaluación en América latina y el Caribe: experiencias concretas*. Francia.: Autor.
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2008). *Situación de América Latina y el Caribe: garantizando la educación de calidad para todos*. (Informe regional de revisión y evaluación del progreso de América Latina y el Caribe hacia la educación para todos en el marco del proyecto regional de educación EPT/PRELAC-2007). Santiago, Chile.: Autor.
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (Ed.). (2003). *Evaluar las evaluaciones: una mirada política acerca de las evaluaciones de la calidad educativa*. Buenos Aires, Argentina.: Autor.
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (Ed.) (2008). *Reflexiones en torno a la evaluación de la calidad educativa en América Latina*. Santiago, Chile.: Autor.
- Organización para la Cooperación y Desarrollo Económico. (OCDE). (2009). *Informe TALIS: la creación de entornos eficaces de enseñanza y aprendizaje. Síntesis de los primeros resultados*. España.: Santillana.
- Organización para la Cooperación y Desarrollo Económico. (OCDE). (2011, noviembre). ¿Qué pueden hacer los padres para ayudar a sus hijos a tener éxito en los centros educativos? *PISA in focus*, 10, 1-4.
- Organización para la Cooperación y Desarrollo Económico. (OCDE). (Octubre 2011). Autonomía y rendición de cuentas en los centros educativos: ¿están relacionadas con el rendimiento de los estudiantes? *PISA in focus*, 9, 1-4.
- Organización para la Cooperación y Desarrollo Económico. (OCDE). (s.f.). *El programa PISA de la OCDE: qué es y para qué sirve*. Recuperado el 12 de abril de 2013, de <http://www.oecd.org/pisa/39730818.pdf>

- Patelis, T. & Matos-Elefonte, H. (2009, abril). Efforts to produce relevant score reports to school, district, and state officials on national tests. Presentación en la *Annual meeting of NCME*. San Diego, CA.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3a. ed.). Thousand Oaks, CA, EE. UU.: Sage Publications.
- Popham, W. J. (1999, marzo). ¿Por qué las pruebas estandarizadas no miden la calidad educativa? Tomado y traducido de *Educational leadership*, 56, 6. por el Programa de promoción de la reforma educativa en América Latina y el Caribe (PREAL).
- Programa Sectorial de Educación 2007-2012. (2008, 17 de enero). *Diario Oficial*.
- Reyes, S., Castillo, A., Zuñiga, B. & Llarena, R. (2009). *La evaluación objetiva de la habilidad lectora en la educación media superior: marcos de referencia 2*. D.F., México.: Centro Nacional de Evaluación para la Educación Superior.
- Reyes, S., Castillo, A., Zuñiga, B. & Llarena, R. (2012). *Niveles de dominio en habilidad matemática: la estrategia de evaluación de ENLACE media superior. Marcos de referencia 5*. D.F., México.: Centro Nacional de Evaluación para la Educación Superior.
- Ríos, R. A. & García, E. (2003). *Acciones de evaluación en las instituciones públicas de educación media superior*. (Informe). México.: Instituto Nacional para la Evaluación de la Educación (INEE). Disponible en <http://www.inee.edu.mx/index.php/publicaciones/resultados-de-investigacion/reportes-de-investigacion/79-publicaciones/reportes-de-investigacion-capitulos/437-acciones-de-evaluacion-en-las-instituciones-publicas-de-educacion-media-superior-informe>
- Roberts, M. R. & Gierl, M. J. (2008, abril). Developing and evaluating score reports for a diagnostic mathematics assessment. Simposio efectuado en la Annual meeting of the American Educational research Association, Nueva Orleans, Luisiana. EE.UU.
- Roberts, M. R. & Gierl, M. J. (2009, otoño). Developing score reports for cognitive diagnostic assessments. *Educational measurement: issues and practice*, 29 (3), 25-38.
- Rosenkvist, M. A. (2010). Using student test results for accountability and improvement: a

- literature review. *OECD Education working papers, 54*. OECD publishing.
- Rossi, P. H. & Lipsey, M. W. & Freeman H. E. (2004). *Evaluation: a systematic approach*. (7a. ed.). Thousand Oaks, CA, EE.UU.: Sage Publications.
- Ryan, J. M. (2012). Prácticas, problemas y tendencias en el reporte de resultados del estudiante. En Downing S. M. & Haladyna T.M. (Eds.). *Manual para el desarrollo de pruebas a gran escala: Vol. II*. (pp. 1005-1050). (Centro Nacional de Evaluación para la Educación Superior Trad.). D.F., México.: Centro Nacional de Evaluación para la Educación Superior. (Trabajo original publicado en 2006).
- Ryan, K. E. & Cousins, J.B. (Eds.). (2009). *The SAGE international handbook of educational evaluation*. Thousand Oaks, CA, EE. UU.: Sage Publications.
- Ryan, K. E. & Feller, I. (2009). Evaluation, accountability, and performance measurement in national education systems: trends, methods and issues. En Ryan, K. E. & Cousins, J.B. (Eds.). *The SAGE international handbook of educational evaluation*. (pp. 171-189). Thousand Oaks, CA, EE. UU.: Sage Publications.
- Scriven, M. (1991). *Evaluation thesaurus*. (4a. ed.). Thousand Oaks, CA, EE. UU.: Sage Publications.
- Scriven, M. (2007). *Key evaluation list*. Western Michigan University. Recuperado el 12 de mayo de 2013, de <http://www.wmich.edu/evalctr/checklists/evaluation-checklists/>
- Shaw, F. I. (1999) *La evaluación cualitativa: introducción a los métodos cualitativos*. España: Paidós.
- Simkin, D. & Hastie, R. (1987, junio). An information-processing analysis of graph perception. *Journal of the American Statistical association*, 82 (398), 454-465.
- Sireci, S.G. & Forte, E. (2012, verano). Informing in the information age: how to communicate measurement concepts to education policy makers. *Educational Measurement: issues and practice*, 31 (2), 27-32.
- Stake, R. E. (2010). *Qualitative research: studying how things work*. NY, EE.UU.: The Guilford Press.

Subsecretaría de Educación Media Superior. (2009). *Acuerdo número 1/CD/2009 del Comité Directivo del Sistema Nacional de Bachillerato*. Diario Oficial.

Tamassia, C. V. & Adams, R. J. (2009). International assessments and indicators: how will assessments and performance indicators improve educational policies and practices in a globalized society. En Ryan, K. E. & Cousins, J.B. (Eds.). *The SAGE international handbook of educational evaluation*. (pp. 213-229). Thousand Oaks, CA, EE. UU.: Sage Publications.

The board of regents of the University of Wisconsin System. (2011). *Access for ELLs: interpretative guide for score reports. Spring 2011*. Wisconsin, EE.UU.: Autor.

The Danish Evaluation Institute. (2003). *Educational evaluation around the world: an international anthology*. Copenhagen, Dinamarca.: Vester Kopi.

Tiana, A. (2003). ¿Qué pretendemos evaluar, qué evaluamos y qué conclusiones podemos extraer de la evaluación? En Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (Ed.). *Evaluar las evaluaciones: una mirada política acerca de las evaluaciones de la calidad educativa*. (pp.111-132). Buenos Aires, Argentina.: Autor.

Torrance, H. (2009). Pursuing the wrong indicators? The development and impact of test-based accountability. En Ryan, K. E. & Cousins, J.B. (Eds.). *The SAGE international handbook of educational evaluation*. (pp. 483-498). Thousand Oaks, CA, EE. UU.: Sage Publications.

Ugalde, C. (2002). *Rendición de cuentas y democracia: el caso de México*. México.: Instituto Federal Electoral.

Utah State Board of Education. (2013). *Utah academic language proficiency assessment: score reports Interpretation guide*. EE. UU.: Autor.

Vidal, R. (2009). *¿Enlace, exani, excale o pisa?* México.: Centro Nacional de Evaluación para la Educación Superior.

Vidal, R. *¿Enlace, Exani, Excale o PISA?*. (2009): Centro Nacional de Evaluación para la Educación Superior (Ceneval).

- Walsh, W. B. & Betz, N. E. (2001). *Tests and assessment*. (4a. ed.). Upper Saddle River, NJ, EE. UU.: Prentice- Hall.
- Weiss, C. (2003/2004, invierno). On theory-based evaluation: winning friends and influencing people. *The evaluation exchange*, IX (4), 2-3.
- Weiss, C. H. (2008). *Investigación evaluativa: métodos para determinar la eficiencia de los programas de acción* (3ª ed.). (Trillas Trad.) D.F., México.: Trillas. (Trabajo original publicado en 1996).
- Wolff, L. (2006). Las evaluaciones educacionales en América Latina: avance actual y futuros desafíos. En Arregui, P. (Ed.). *Sobre estándares y evaluaciones en América Latina*. Santiago, Chile.: Programa de Promoción Regional de Educación en América Latina (PREAL).
- Zapata-Rivera, D. & Zwick, R. (Eds.). (2011). *Tests score reporting: perspectives from the ETS score reporting conference (research report ETS RR-11-45)*. Princeton, Nueva Jersey, EE. UU.: Educational Testing Services.
- Zenisky, A. L. & Hambleton, R. K. (2012). Developing test score reports that work: the process and best practices for effective communication. *Educational measurement: issues and practice*, 31 (2), 21-26.

## **ANEXOS**

## **Anexo A. Estándares de los *Standards for Educational and Psychological Testing* pertinentes con el tema de los reportes de resultados**

**Estándar 5.10** Cuando se entregue información de puntajes a estudiantes, padres, representantes legales, maestros, clientes o medios de comunicación, los responsables de realizar la evaluación deben proporcionar interpretaciones adecuadas. Las interpretaciones deben describir en un lenguaje sencillo lo que abarca la prueba, el significado de los puntajes y su precisión, las malas interpretaciones más comunes y cómo interpretar adecuadamente (p. 65)

**Estándar 5.11** Cuando se reportan interpretaciones por computadora, deben hacerse accesibles sus fuentes, razones y base empírica, así como describirse sus limitaciones (p.65).

**Estándar 5.12** Cuando se reporta información grupal a partir de la integración de diferentes pruebas, se debe reportar la validez y la confiabilidad en el nivel de integración final. No se deben reportar puntajes individuales a menos que la validez, comparabilidad y confiabilidad de dichos puntajes se establezca (p.65).

**Estándar 5.13** La entrega de resultados a usuarios autorizados o instituciones debe realizarse asegurando la confidencialidad de los puntajes (p.66).

**Estándar 6.12** Los editores y servicios de calificación que ofrecen interpretaciones de calificaciones generadas por computadora deben proporcionar un resumen de la evidencia que apoya las interpretaciones (p.70).

**Estándar 7.8** Cuando los puntajes se reportan al público de forma desagregada como género, etnicidad, edad, lengua, eficacia o discapacidad; se deben incluir advertencias en los reportes de investigación acerca de la incomparabilidad de los grupos (p.83).

**Estándar 7.10** Cuando los resultados de una prueba afectan las oportunidades educativas de los examinados; debe examinarse la evidencia de las diferencias de las medias de los puntajes entre subgrupos, de ser posible, con los subgrupos donde reportes de investigación encuentren diferencias en pruebas similares. Se debe realizar una investigación para averiguar si las diferencias son o no atribuibles a una baja representación de los constructos o a varianza irrelevante al constructo. El usuario de la prueba tiene la responsabilidad de usarla con los grupos con los cuales se diseñó la prueba (p. 83).

**Estándar 8.8** Cuando los reportes incluyen asignar individuos a ciertas categorías, estas deben elegirse cuidadosamente y describirse con precisión; deben

elaborarse descripciones que no estigmaticen y que sean representaciones exactas (p.88).

**Estándar 11.6** A menos que las circunstancias obliguen a no proporcionar los resultados, existe la obligación de reportarlos en un tiempo razonable que sea comprensible para quienes sustentan la prueba y para aquellos que reciben información sobre esta (p. 114).

**Estándar 11.15** Quienes administran la prueba deben ser advertidos sobre potenciales malas interpretaciones y posibles consecuencias no intencionales de la prueba; los usuarios deben prever acciones que minimicen o eviten malas interpretaciones previsibles o consecuencias negativas no intencionales (p. 116).

**Estándar 11.17** En circunstancias donde el público recibe resúmenes de los resultados, se debe formular una política respecto al tiempo de entrega de éstos y aplicarla de forma constante (p. 117).

**Standard 11.18** Cuando los resultados de la prueba se entregan al público o a los políticos, los responsables de la entrega deben proporcionar y explicar cualquier información complementaria que pueda minimizar las malas interpretaciones de los datos (p. 118).

**Estándar 11.20** Excepto en casos legales o en circunstancias donde el que administra la prueba es el empleador, los profesionales deben compartir los resultados e interpretaciones con quien respondió la prueba. Tal información debe expresarse en un lenguaje que puede comprender éste o su representante legal. (p. 135).

**Standard 12.9** Los profesionales responsables de supervisar los programas de evaluación grupal deben asegurarse de que los individuos que interpretan las calificaciones han recibido una instrucción apropiada sobre los métodos adecuados para dicha tarea (p 132).

**Estándar 12.15** Aquellos que usan interpretaciones de pruebas basadas en computadora deben evaluar la calidad de las interpretaciones y de ser posible la relevancia y propiedad de las normas sobre las cuales se basan estas (p.134).

**Standard 12.19** La interpretación de las calificaciones obtenidas en la prueba o de los patrones de resultados obtenidos en una batería de prueba debe prestar atención a los factores que pueden influir en el resultado de una evaluación particular. Cuando sea apropiado, deben incluirse en el reporte una descripción y un análisis de las hipótesis o explicaciones alternativas que pueden haber contribuido al patrón de resultados. (p. 134).

**Standard 13.1** Cuando las evaluaciones son elaboradas a petición de escuelas, distritos, estados u otras autoridades, las diferentes formas en las que los resultados pueden ser usados deben ser descritas claramente. Es responsabilidad

de los que solicitaron la realización de la prueba monitorear su impacto e identificar y minimizar las potenciales consecuencias negativas. Las consecuencias que resulten del uso de las pruebas, intencionales o no, deben ser examinadas por los usuarios de la prueba (p. 143).

**Estándar 13.3** Cuando una prueba se usa como indicador de logro o desempeño respecto de un dominio o estándar curricular, se debe proporcionar evidencia acerca del rango o extensión de conocimiento que contiene la prueba sobre ese dominio o estándar. Estos últimos deben describirse con detalle suficiente para poder evaluar su relación con el contenido de la prueba. Los análisis deben explicitar los aspectos del dominio que la prueba abarca y aquellos que no (p.145).

**Estándar 13.5** Cuando los resultados se usan para tomar decisiones sobre promoción de grado o el egreso de los estudiantes debe existir evidencia de que la prueba cubre adecuadamente el contenido específico, el general o las habilidades que los estudiantes tuvieron oportunidad de adquirir (p. 146).

**Estándar 13.13** Los responsables de las evaluaciones deben asegurar que los individuos que interpretan los resultados y toman decisiones en el contexto escolar están calificados para hacerlo o tienen asistencia de personas que sí lo están (p. 148).

**Estándar 13.14** En escenarios educativos los reportes de puntajes deben acompañarse de una clara descripción del error de medida asociado a cada puntaje o nivel de clasificación y la forma de interpretar los puntajes (p. 148).

**Estándar 13.16** En escenarios educativos donde se reporte un puntaje, debe proporcionarse el dato de la fecha de administración de la prueba. Esta información y el tiempo de las normas o referentes utilizados para hacer las interpretaciones debe considerarse al hacer inferencias (p. 149).

**Estándar 13.19** En escenarios educativos donde se reporten puntajes promedios o acumulados de manera grupal, deben complementarse con información adicional del tamaño de la muestra y forma de dispersión de la distribución de puntajes (p. 149).

**Estándar 15.11** Cuando los resultados se proporcionan al público o a los que definen políticas educativas, quienes proporcionen los resultados deben proporcionar toda la información complementaria que minimice las malas interpretaciones de los datos (p. 169).

## **Anexo B. Recomendaciones del *Code of Fair Testing in Practices in Education* pertinentes con el tema de los reportes de resultados**

**B6.** Corrija errores que afecten la interpretación de los puntajes y comunique los resultados rápidamente.

**B7.** Desarrolle e implemente procedimientos que aseguren la confidencialidad de los puntajes.

**C1.** Proporcione información para apoyar las interpretaciones de los resultados, incluyendo la naturaleza del contenido, las normas o grupos comparativos y evidencia técnica. Aconseje a los usuarios de los beneficios y límites de los resultados y su interpretación. Advierta sobre los problemas que se pueden generar si se presume de una precisión mayor a la garantizada.

**C2.** Proporcione apoyo y guía respecto a las interpretaciones de los resultados para aquellos que administran las pruebas cuando tiene modificaciones. Informe a los usuarios de problemas potenciales al interpretar los resultados de la prueba cuando la prueba o sus procedimientos son modificados.

**C3.** Especifique los usos apropiados de la prueba y advierta a los usuarios sobre los potenciales usos inadecuados.

**C4.** Cuando existan estándares o clasificaciones, proporcione las razones, procedimientos y evidencia del ajuste de los estándares de desempeño o puntos de corte. Evite el uso de frases que sean negativas o estigmas.

**C5.** Fomente en los usuarios la toma de decisiones sobre la prueba desde diversas y apropiadas fuentes de información, no solo en el puntaje de la prueba.

**C6.** Proporcione información que permita a los usuarios interpretar y reportar los resultados de los subgrupos de forma precisa, incluyendo información de quién, cuándo y dónde no se incluyen en los diferentes grupos, así como los factores que pueden influir en la interpretación de los resultados.

**C7.** Proporcione resultados en un tiempo considerable que sea comprensible para el receptor.

**C8.** Proporcione guías para los usuarios acerca de cómo verificar que la prueba cumple con su propósito.

**D4.** Proporcione a los usuarios o a los padres o tutores información sobre las responsabilidades que tienen como usuarios, como conocer los propósitos intencionales de la prueba, su uso, desempeño, seguimiento de instrucciones y no exponer los reactivos o interferir con otros usuarios de la prueba.