



ENLACE y Excale: Las pruebas que tenemos y las que necesitamos

México, D.F., Febrero de 2015



Instituto Nacional para la
Evaluación de la Educación

México

Contenido

1. Propósitos y criterios del estudio
2. Las pruebas que tenemos.
3. Las pruebas que necesitamos.
4. Una visión al futuro inmediato.

Propósitos y criterios del estudio

- I. Evaluar la validez técnica y usos de las pruebas ENLACE/básica, ENLACE/MS y Excale.
 - II. Hacer recomendaciones para que cumplan con el propósito de ayudar a mejorar la calidad del Sistema Educativo Nacional.
 - III. Aspectos analizados:
 1. Alineación a referentes (curriculares).
 2. Propiedades psicométricas.
 3. Sesgo cultural (atención a la diversidad).
 4. Aplicaciones.
 5. Usos y consecuencias.
 - IV. Se utilizaron **58** criterios y **97** sub-criterios para analizar las tres pruebas.
-

Las pruebas que tenemos



Alineación a los referentes

- *Planeación de la prueba:* Hay dificultades en alinear las pruebas al currículo mexicano en constante transformación.
- *Especificaciones de ítems:* Son demasiado generales en ENLACE/MS, mejor descritas en ENLACE/Básica y muy detalladas en Excale.
- *Elaboración y validación ítems:* Las tres pruebas tratan de seguir las mejores prácticas, pero hay diferencias en la calidad y suficiencia: mejor en Excale que en ENLACE.

ASPECTOS PSICOMÉTRICOS (1)

- *Manuales técnicos:* Es necesario actualizarlos y publicarlos. Excale tiene muchos manuales, pero no actualizados; ENLACE tiene pocos y no públicos.
 - *Desarrollo y construcción de pruebas:* Falta sistematizar los reportes estadísticos; cumplir con criterios de calidad.
 - *Calibración y análisis psicométricos:* Se requiere elaborar bancos de reactivos calibrados y especificar los métodos de anclaje.
 - *Confiabilidad:* Se debe mejorar este indicador con varios métodos y estimar el error de medida de cada prueba, dimensión y subescala.
-

ASPECTOS PSICOMÉTRICOS (2)

- *Calificación del nivel desempeño e interpretación de resultados:* Mejorar definición de puntos corte y niveles de desempeño. Describir método de calificación y correcciones con error de medida en puntos corte.
- *Bancos de reactivos y calidad psicométrica de ítems:* Sistematizar estudios de DIF y sesgo cultural; explorar fuentes de variación en estratos poblacionales específicos.
- *Evidencias de validez:* De la escala completa y de los niveles de desempeño.

SESGO CULTURAL

- *Marco conceptual*: Inexistencia documental sobre posible sesgo social, cultural y lingüístico.
- *Contextualización*: Especificaciones de ENLACE insuficientes para precisar características gráficas, textuales y contextuales.
- *Problemas lingüísticos*: En las tres pruebas hay deficiencias de contenido de los ítems.
- *Corrección de sesgo*: No hay procedimientos en ENLACE para eliminar reactivos con problemas de sesgo. Excale hace piloteos para detectar y evitar fuentes de sesgo.

APLICACIONES DE PRUEBAS (1)

- *Bases de datos*: Se observaron irregularidades en las BD; se debe asegurar su actualización.
- *Controles de calidad*: Las pruebas difieren en aseguramiento de la calidad; mejor en Excale que en ENLACE. Se documentan procedimientos de aplicación, pero no sobre su cumplimiento. Excale utiliza evaluadores externos; ENLACE no.
- *Capacitación a aplicadores*: Se apoya en materiales impresos y no en material digital. ENLACE debe asegurar que se lleve a cabo con tiempo suficiente.

APLICACIONES DE LAS PRUEBAS (2)

- *Integridad de la información*: Se debe considerar y anticipar formas diversas de fraude. ENLACE estuvo sujeta a ellas.
- *Cuestionarios de contexto*: valorar la pertinencia de las preguntas, su extensión, su continuidad en el tiempo y su utilidad.
- *Distribución de materiales*: Asegurar que un operador traslade y colecte los materiales evaluativos de la escuela; evitar instancias intermedias.

USOS Y CONSECUENCIAS (1)

- *Uso de pruebas:* Las pruebas deben utilizarse para los propósitos que se diseñan. El uso diverso de los resultados debe especificarse y jerarquizarse. ENLACE no se alinea a los usos más visibles.
- *Grupos de usuarios:* Hay que definir grupos de usuarios (maestros, comunicadores, tomadores de decisión) y asegurar que los resultados lleguen en tiempo y forma.
- Hay que considerar la necesidad de capacitar en el uso e interpretación de resultados, teniendo en cuenta la precisión de los puntajes.

USOS Y CONSECUENCIAS (2)

- *Coordinación entre organismos evaluadores y entidades.* No se aprecia una estrategia coordinada en ENLACE. Se aprecia gran diferencia en capacidad técnica entre estados. Excale perdió visibilidad por la atención que generó ENLACE.
- *Uso de bases de datos:* Mecanismos insuficientes para acceder a BD y para desarrollar capacidades en investigadores y estudiantes de grado.
- *Estudios longitudinales:* No hay bases de datos para realizar este tipo de estudios.

USOS Y CONSECUENCIAS (3)

- *Impacto de programas*: Considerar mecanismos para evaluar programas educativos, tales como acceso a ítems liberados y asistencia para construir pruebas apropiadas para analizar el impacto de programas específicos.
- *Usos y consecuencias imprevistas*: No hay mecanismos de reacción cuando se detectan usos injustificados y perniciosos.
- *Consecuencias negativas*: Inexistencia de estudios sobre efectos negativos de las pruebas, como la inflación de resultados, el estrechamiento curricular y enseñar-para-la-prueba.

Las pruebas que necesitamos



Propósitos acotados y diseño congruente.

- *Evaluaciones con varios propósitos*: Una prueba no puede atender bien todas las necesidades. En su caso, se deben satisfacer las exigencias específicas de cada uno de los usos que se le quieran dar.
- *Múltiples pruebas*: atender una gama amplia de propósitos con varias pruebas implica una excesiva carga de trabajo para escuelas y alumnos y un efecto negativo en el aprendizaje. Hay que priorizar necesidades.

Referentes alineados con puntos centrales y estables del currículo

- *Necesidad de referentes:* El currículo es el referente nacional, pero cambia con frecuencia. Si se quiere medir el cambio en el aprendizaje no hay que modificar la medida.
- *Tendencias de aprendizaje:* Para que las pruebas informen sobre los cambios en el tiempo, es necesario que el currículo tenga un núcleo de propósitos centrales comunes y estables.

Diseño y desarrollo consistentes con los avances psicométricos

- *Estándares psicométricos*: No es razonable exigir que toda prueba atienda los últimos avances de la Psicometría, pero sí los aspectos básicos y fundamentales.
- *Manuales Técnicos*: Cada prueba debe estar bien documentada, donde se precisen los criterios utilizados para su diseño, desarrollo y validación.

Cuidar diversidad socioeconómica y cultural

- *Atención a la diversidad*: la ausencia de sesgo cultural debe atenderse en todas las etapas del desarrollo, aplicación y entrega de resultados.
- Los mejores sistemas de evaluación tienen fallas, pero en un país multicultural, como México, es importante comenzar a utilizar procedimientos que consideren su diversidad.
- Hacer entrevistas cognitivo-culturales y estudios de generalizabilidad, así como análisis de sesgo de los resultados.

Sistemas eficientes de aplicación y de procesamiento de resultados

- *Calidad de las aplicaciones:* Implica personal profesional; sistematizar procesos en manuales; controles de calidad; uso de TIC para bases de datos de alumnos, maestros y escuelas; manejo y control de procesos que posibiliten acciones correctivas.
- *Uniformidad en las aplicaciones:* En México es necesario hacer acuerdos con los estados para alcanzar los mismos niveles de calidad en todo el país.

Reconocer el impacto de las evaluaciones

- *Expectativas falsas*: La difusión de los resultados no hace que el aprendizaje mejore; tampoco hace que los docentes cambien sus prácticas pedagógicas, ni que los directores modifiquen su gestión escolar.
- *Factores que impactan el aprendizaje*: El logro de los estudiantes se debe a varios factores; si no cambian las condiciones en que ocurre el proceso de enseñanza-aprendizaje, no mejorará el rendimiento académico de los alumnos.

Monitoreo sistemático de usos y consecuencias, pretendidos o no

- *Usos inapropiados*: Su impacto negativo puede ser muy grande (ej.: corrupción de la medida e inflación de resultados).
- *Calidad técnica*: Para lograrla es necesario el escrutinio de especialistas externos y de personas interesadas.
- *Transparencia*: Es esencial la participación democrática basada en la deliberación informada.

Apertura a escrutinio externo, documentación completa y accesible

- *Documentación accesible*: Todas las etapas del diseño, desarrollo, aplicación, análisis y difusión de resultados de una prueba deben estar documentados en: manuales, especificaciones, ítems liberados, estudios de validez, nivel de precisión y soporte de recomendaciones sobre usos apropiados de resultados.

Una visión al futuro
inmediato



Investigación asociada a la evaluación

Hay que fomentar:

- Trabajos sobre bases de datos para estudiar relaciones entre variables y factores asociados al aprendizaje.
- Estudios longitudinales que exploren patrones del comportamiento de variables y resultados a través del tiempo.
- Estudios de innovación para fortalecer prácticas de evaluación en el aula.
- Alianzas y mecanismos de apoyo entre distintas instituciones.

Promover formación de especialistas

- Falta de posgrados en evaluación. Urge programa estratégico de formación de especialistas, con alianzas nacionales-internacionales.
- Además de psicometría, se debe formar en: gestión de bancos de ítems y bases de datos; administración de pruebas; validez lingüística y cultural; difusión de resultados; uso para evaluar programas-políticas, tec.
- Se deben formar a cientos de personas de niveles altos a operativos.

Cuidar tiempos-resistir presión política

- Desarrollar pruebas implica años de trabajo. Fallas por la premura por razones políticas poco ajustadas a la realidad educativa y técnica.
- El interés de autoridades es importante, pero por el desconocimiento de implicaciones técnicas lleva a exigencias incompatibles con la calidad.
- Los recursos no bastan para acortar plazos.

Apoyar el trabajo de los docentes

- Las pruebas estandarizadas pueden dar al maestro referentes útiles sobre el aprendizaje, pero no la información con detalle y oportunidad de la evaluación en el aula.
- Los alumnos no aprovechan resultados de la evaluación externa o de aula si el maestro no los apoya y los guía.
- Los docentes son clave para que los resultados de las pruebas se interpreten y se usen bien, y para hacer evaluaciones en aula más efectivas.

Construir la cultura de la evaluación

- La evaluación es necesaria pero no suficiente para la mejora del alumnos. Hay que desarrollar una cultura de la evaluación:
 1. Las autoridades necesitan evaluaciones de calidad y capacidad de interpretarlas correctamente.
 2. Se debe desarrollar en escuelas un visión rica de las prácticas de enseñanza y evaluación.
 3. Los padres y la ciudadanía deben entender los resultados para apoyar a sus hijos y orientar demandas a escuelas.
 4. Los medios de comunicación pueden apoyar a difundir la información correctamente.



José María Velasco 101,
Col. San José Insurgentes
Deleg. Benito Juárez, CP. 03900
México, D.F. Tel. 5482 0900

www.inee.edu.mx

