



# **EVALUACIÓN EN GRAN ESCALA Y EVALUACIÓN EN AULA: ¿OPUESTAS O COMPLEMENTARIAS?**

**Felipe Martínez Rizo**

**La evaluación educativa: un continuo desafío  
hacia una perspectiva transformadora**

**IIDE-UABC, Ensenada, noviembre 5, 2012**

# Contenido

Planteamiento de la cuestión.

- 1) Las evaluaciones en gran escala: lo que pueden dar y lo que no está a su alcance.
- 2) Las evaluaciones que hacen los maestros en el aula y su potencial.
- 3) La necesidad de un sistema equilibrado.

Conclusión.

# Planteamiento de la cuestión

Las formas en que se evalúa el aprendizaje en las escuelas mexicanas están cambiando mucho.

Hay aspectos positivos, pero también tendencias que preocupan.

Propondré una reflexión al respecto considerando la posible relación entre la evaluación que utiliza pruebas estandarizadas que se aplican en gran escala y la que realizan los maestros en el aula.

Comenzaré revisando el proceso que ha llevado a la situación actual a lo largo del tiempo.

# Siglos XVI a XVIII

Escuelas pequeñas y sin grados. Dos estrategias:

- ✓ Alumnos recitaban en coro textos a memorizar; el maestro hacía preguntas que requerían respuesta corta. (Catecismo)
- ✓ El maestro indicaba textos a repasar, cada alumno leía y releía lo indicado; maestro "tomaba lección" que era recitación de lo aprendido.

Resultado desfavorable: recitar o repasar los mismos textos una y otra vez, hasta aprenderlos.

Resultado favorable: nuevos textos-nuevos pasajes.

# Masificación del siglo XIX

Enseñanza monitorial y escuelas por grados.

**En ambos casos la enseñanza tipo catecismo pasó a las aulas laicas**

Los contenidos se ampliaron: aritmética-redacción-ciencias-historia); ***forma de interactuar persiste***

- ✓ Alumno: leía texto o hacía ejercicios aritmética.
- ✓ Profesor: preguntaba para verificar si recordaba lo leído o respuestas a ejercicios eran correctas.

Según resultado, repasar o pasar a tema siguiente.

# Principios del siglo XX

Con generalización de la primaria prevalecía bajo nivel en las escuelas; la evaluación hecha por los maestros tenía serias deficiencias.

**Se desarrollaron pruebas que permitieran comparar niveles de distintas escuelas. Se volvieron forma de evaluación preferida.**

La evaluación de los maestros se subordinó a las pruebas, que debían imitarse metodológicamente. Así lo muestran los textos de evaluación utilizados en las instituciones formadoras de maestros.

# Hacia 1960

Clara influencia de nuevas formas de evaluación: se pide a maestros elaborar pruebas de opción múltiple y *calificar por curva* sin preparación para ello; las pruebas hechas por ellos son deficientes.

No sólo en América Latina:

*La práctica no se limitaba a un tiempo o lugar; esta autora confiesa que hacía lo mismo como maestra recién calificada en Inglaterra, en la década de 1970, creyendo que así era científica.*

# Desde fines del siglo XX

- ❖ Las pruebas se extienden a nivel internacional y suscitan grandes expectativas.
- ❖ Reacciones negativas ante riesgo de reducción del currículo, preparación de los alumnos para la pruebas y uso para evaluar escuelas y maestros sin sustento sólido.
- ❖ Otra reacción: fortalecer la evaluación que hacen los maestros, en especial formativa, para apoyar el aprendizaje.

**¿Qué relación debería haber entre pruebas en gran escala y evaluación en aula?**



# 1. La evaluación en gran escala

## **Perspectiva a principios del siglo XX**

Los pioneros de pruebas buscaban instrumentos que permitieran comparar las escuelas entre sí, dada *la escandalosa falta de confiabilidad de los exámenes que aplican los maestros...* (Thorndike)

## **Ya había conciencia de límites de pruebas:**

- *Las pruebas miden hechos aislados y piezas de información, en vez de capacidad de razonamiento, habilidad organizadora, etc.* (Word)
- *Necesario verlas no como proceso separado de la enseñanza, sino parte integral de ésta.* (Tyler)

## **Se impuso visión de Thorndike**

# Las pruebas en gran escala hoy

- Avance de psicometría (IRT, GT...) permite medir desempeños complejos en forma que hace un siglo impensable: pruebas de ejecución, adaptativas, matriciales, de respuesta construida...
- Ni las mejores pruebas miden expresión oral o aspectos como actitudes y valores.
- Muchas de las pruebas que se aplican realmente no son de la mejor calidad: preguntas de OM que miden memoria o mecanización, insensibles a instrucción, con sesgos fuertes y otros límites.
- No se pueden aplicar con la frecuencia que sería necesaria para dar retroalimentación oportuna.

# Se extienden malos usos

*...aspectos importantes del currículo se están haciendo a un lado, porque no son medidos por las pruebas... los niños están siendo entrenados sin descanso para que dominen el contenido de esas pruebas de alto impacto y, en consecuencia, están comenzando a odiar la escuela... en muchos casos los maestros se dedican a preparar a sus alumnos para las pruebas, lo que se parece mucho a hacer trampa, porque están inflando las puntuaciones de los alumnos sin elevar su competencia...*

*(Popham, 2001)*

# Usar resultados de alumnos para evaluar maestros y escuelas

Difusión de pruebas masivas hizo pensar en forma diferente de evaluar desempeño de maestros.

La idea básica es sencilla: el trabajo docente busca que los alumnos aprendan; por tanto, *si hay datos comparables y fiables de rendimiento de alumnos* no es necesario observar el trabajo de un maestro para valorar su desempeño; basta inferirlo a partir de los resultados de sus estudiantes.

**Son mejores los maestros cuyos alumnos obtienen resultados más altos en las pruebas.**

# Modelos de Valor Agregado

**Objeción:** algunos alumnos comienzan el curso en posición mejor que otros; los resultados al fin del año no reflejan sólo el desempeño del docente.

**Respuesta:** no considerar simplemente resultado de cada alumno al fin del curso, sino la diferencia entre lo que sabía al inicio y lo que sabe al fin del mismo, **controlando otras posibles influencias.**

Esa diferencia es el *valor agregado* (*value added*), que se supone atribuible al trabajo del docente, y en lo que deberán basarse las evaluaciones.

# Problemas de los VAM

- ✓ Límites de las pruebas, que no miden aspectos importantes, v. gr. competencias complejas.
- ✓ Imprecisión de las mediciones e inestabilidad en el tiempo; el margen de error de estimaciones del valor agregado por el maestro es demasiado grande para servir para decisiones individuales.
- ✓ Dificultad práctica para tener bases de datos completas, en especial para seguir en forma individual a alumnos que cambian de escuela.
- ✓ Dificultad básica: atribuir avance de alumnos en un grado a un maestro, ya que ambos pueden cambiar, puede haber más de un maestro, y en los resultados influyen maestros previos...

## 2. La evaluación en aula

Aún si es de buena calidad, la evaluación hecha por el maestro no permite comparar resultados de diferentes escuelas, porque cada docente ajusta los criterios de evaluación según las condiciones de sus alumnos y no puede tener en cuenta al sistema.

Pero por la misma pequeña escala en que se hace, la evaluación en aula tiene el potencial de ser más fina y completa que la de gran escala.

# Potencial de evaluación en aula

El trabajo del maestro es insustituible para que los alumnos aprendan y para evaluar si ocurre. Pero muchas veces la evaluación del maestro se reduce a lo más simple y superficial. Es necesario que mejore de manera que pueda cubrir bien:

- ◆ Todas las áreas del currículo;
- ◆ Todos los niveles cognitivos;
- ◆ El avance a lo largo del ciclo escolar.
- ◆ Y que se use para dar a cada alumno información que le sirva para avanzar mejor.

Eso es lo que pretende la **Evaluación Formativa**.



# Desarrollo de la EF

- El término lo introdujo Scriven en 1967.
- Bloom lo aplicó al aprendizaje de los alumnos con el sistema Mastery Learning, desde 1968.
- Sanders precisó el concepto en 1989 señalando que la EF debe precisar el punto de llegada (el aprendizaje esperado); el de partida (el estado en que se encuentra el alumno); y sobre todo la forma de pasar de uno a otro, orientando al alumno para que alcance el aprendizaje.
- En las dos últimas décadas se ha desarrollado en el marco de las corrientes cognitivistas, e incluyendo lo relativo al impacto afectivo.

# La evaluación formativa

- *Mastery Learning*: neo-conductista: enseñanza-prueba-realimentación o corrección.
- Después: EF en el marco de visión cognitivista de e/a: constructivismo o socio-constructivismo, comunidades de práctica, aprendizaje situado...
- En última instancia, EF busca conducir al alumno a desarrollar habilidad de "aprender a aprender", "estrategias metacognitivas", "autorregulación".
- Un elemento clave de la EF es la información que proporcione al alumno para orientar su esfuerzo por mejorar: retroalimentación o devolución.

# Verdadera y falsa EF

No cualquier retroalimentación es formativa; no basta hacer saber al alumno que respondió mal.

- ✓ Pegar al niño retroalimenta, pero... ¿le ayuda?
- ✓ No basta decir que necesita esforzarse.
- ✓ No basta repetir más despacio y más fuerte lo que los alumnos no entendieron.

A fines del s. XX se extiende interés por EF pero la desvirtúan aplicaciones frecuentes que interrumpen la enseñanza con poca retroalimentación útil.

- ✓ ¿Aplicar pruebas frecuentemente y comunicar rápidamente el resultado ayuda a mejorar?

# El riesgo de trivializar la EF

**En gran escala:** pruebas que son sólo mediciones periódicas para valorar avance a medio ciclo. Esas pruebas de medir avance, monitoreo o intermedias se ofrecen como formativas pero si no dan detalle necesario para corregir el rumbo e interrumpen enseñanza sin retroalimentación útil **no lo son.**

**En aula:** unos maestros creen que usar portafolios es EF. Si no hay retroalimentación que capacite a los estudiantes para modificar y mejorar su trabajo a medida que avanzan **tampoco lo es.**

# Cómo se desvirtúa la EF

EF no igual a dar calificaciones frecuentemente.

Es necesaria retroalimentación que diga al alumno por qué su respuesta no es satisfactoria, qué falta, en dónde están las deficiencias, en qué consisten y qué debe hacer para superarlas.

*Lo distintivo de la EF no es el tipo de información o las circunstancias en que se genera, sino el efecto positivo que tiene en el aprendiz.*

***La diferencia se notará cuando la pregunta clave de los alumnos no sea ¿Vendrá esto en la prueba?, sino ¿Cómo puedo aprender más?***

# EF y EPA

En vez de **Evaluación Formativa** unos prefieren decir **Evaluación Para el Aprendizaje** (EF-EPA).

La buena *EF-EPA* no es distinta de la enseñanza de buena calidad: es un aspecto de ella.

No es una técnica más y no es fácil hacerla bien, pero no implica algo distinto de lo que se requiere para hacer buena enseñanza.

Un maestro capaz de desarrollar prácticas de enseñanza acordes con los principios cognitivistas y constructivistas, no superficialmente, sino de manera profunda, será capaz al mismo tiempo de hacer realmente **Evaluación Formativa**.

# 3. Hacia un sistema equilibrado

- ◆ Las pruebas no pueden evaluar el aprendizaje de cada alumno con la finura que puede alcanzar la evaluación a cargo del maestro.
- ◆ Por ello ninguna decisión importante sobre un alumno debería basarse sólo en pruebas.
- ◆ Pero las pruebas pueden dar algo que el maestro no puede: información comparable entre escuelas.

**Por eso las pruebas son indispensables para evaluar el sistema como tal y pueden complementar la evaluación de alumnos en lo individual que debe hacer el maestro.**

# Dos formas de ver las pruebas

- ❖ Como herramienta clave para evaluar aprendizaje de alumnos, mejor que la evaluación que realizan los maestros en el aula, que nunca sería confiable.
- ❖ Como medio útil pero insuficiente, a complementar con evaluaciones que llevan a cabo los maestros.

## **Pero:**

- ❖ La evaluación que hace el maestro puede ser de buena o mala calidad; es necesario apoyarlo para que no trate de imitar las pruebas en gran escala, sino que haga buena evaluación en aula.
- ❖ Ninguna prueba, por bien hecha que esté, puede remplazar el trabajo de un buen maestro.



# Conclusión

Un visión equilibrada de la evaluación que se debe hacer en un sistema educativo supone dos cosas:

- ❖ Un uso parsimonioso de las pruebas en gran escala, que considere sus alcances y límites reales y evite hacer mal uso de sus resultados.
- ❖ Una evaluación en aula a cargo de los maestros fortalecida, para que adopte verdaderamente un enfoque formativo y por ello apoye el avance de cada uno de los alumnos.

Los dos puntos son importantes y ambos enfrentan dificultades para volverse realidad.

# Los malos usos de las pruebas

En contexto de preocupación por calidad educativa con fundamento, pero manejada en forma simplista.

- Absurda competencia de los medios por ver quién hace el comentario más alarmista sobre la situación educativa “catastrófica” que revelarían las pruebas ignorando o minimizando datos positivos.
- Uso de pruebas para hacer ordenamientos simples con poco sustento que sirven para dar recompensas económicas o al menos “nombrar y avergonzar”.
- Juicios que ignoran la complejidad de los sistemas educativos y llevan a plantear estrategias simplistas.

# Estrategias simplistas

- Muchas personas ignoran dificultad que implica dar buenos resultados educativos con grupos de alumnos de medio social desfavorable.
- Es frecuente que dirigentes empresariales vean con simpatía estrategias simplistas, pensando que fallas de la escuela pública se corregirían fácilmente con escuelas privadas, ignorando que menos del 10% de los niños asisten a ellas.
- Por eso es frecuente opinión de que bastará con aplicar pruebas masivamente y tomar medidas correctivas simples para que la calidad mejore.

# Pero...

Las estrategias simplistas de mejora parten de un supuesto falso: que hacer buena educación en cualquier contexto es fácil:

*Los sistemas de rendición de cuentas basados en pruebas se basan en creencia de que la educación pública puede mejorar gracias a una estrategia sencilla: haga que todos los alumnos presenten pruebas estandarizadas de rendimiento, y asocie consecuencias fuertes a las pruebas, en la forma de premios cuando resultados suben y sanciones cuando no ocurra así.*

# Fortalecer EF tampoco es fácil

Las práctica docentes, en especial de evaluación, conforman esquemas de conducta profundamente arraigados, que reciben influencia de un complejo entramado de factores, muchos de los cuales no dependen del control de los maestros.

Por ello transformar en profundidad esas prácticas implica mucho más que las actividades habituales de actualización, por lo general de corta duración y muchas veces pobremente implementadas.

# Factores que inciden en prácticas

- Variables remotas: características de maestros (personales, experiencia escolar temprana, formación inicial y en servicio, experiencia).
- Var. intermedias: conocimientos, concepciones y percepciones de los maestros (sobre sí mismos, los alumnos, las materias, la enseñanza, la evaluación misma).
- Variables intervinientes de los alumnos (número y características...).
- Var. intervinientes de escuela y aula (incluyendo el currículo y las pruebas externas).

# Lo deseable

- Hacer evaluación formativa no es fácil; sin ella la utilidad de la evaluación para la mejora de la calidad educativa será reducida.
- El papel preponderante de la evaluación en gran escala no es correcto. Sin desconocer aspectos positivos, ha llegado el momento de advertirlo.
- Alternativa: un sistema que equilibre evaluación en gran escala parsimoniosa y consistente con un rico trabajo de evaluación formativa en aula a cargo de los maestros.

# El Manifiesto de Stiggins

*Hoy entendemos mucho mejor que antes cómo usar la evaluación productivamente. Debemos sustituir los pasados sistemas, muy desequilibrados, por otros que satisfagan las necesidades de todos los usuarios...*

*Para conseguir ese equilibrio tan necesario y ausente, debemos implementar prácticas de evaluación en aula que mantengan la fe de los alumnos en sí mismos...*

*...es tiempo de sustituir la intimidación de la rendición de cuentas como principal motivación, por la promesa del éxito académico para todos los aprendices, como esa fuerza motivacional.*



*El miedo a veces funciona, pero sólo motiva a los que tienen esperanza de éxito. Si se trata de aprender la desesperanza siempre pesa más que la intimidación.*

*Los empresarios han creído que tendremos mejores escuelas si las comparamos según sus resultados en las pruebas anuales y las premiamos o castigamos con base en ello; no entienden el impacto negativo sobre las escuelas que siempre pierden...*

*El político piensa que si poca intimidación no funciona mucha lo hará y las pruebas han sido la forma de subir la presión; tampoco entiende sus repercusiones sobre escuelas y alumnos de bajos resultados.*

En México es necesario contar con un sistema que combine equilibradamente la evaluación en gran escala y la evaluación en aula.

Según Stiggins:

*El inmenso potencial de la evaluación para apoyar el aprendizaje se ha desaprovechado; ha pasado desapercibido en los niveles más altos de decisión. No tiene por qué ser así. Está a nuestro alcance una nueva visión de una evaluación de excelencia, que libere la fuente de confianza, motivación y potencial de aprendizaje que hay en cada alumno.*

# Para terminar

La evaluación por sí sola no mejora la calidad; lo irremplazable para ello es un trabajo intenso y bien orientado; una buena evaluación puede ayudar.

Necesitamos buena evaluación formativa hecha por maestros profesionales, y pruebas que no busquen sustituir a los docentes, sino apoyarlos.

Lo que NO hace falta es un uso de los resultados que desorienta a los tomadores de decisiones y la sociedad, lleva a unos a jactarse de logros dudosos y a muchas personas serias y comprometidas a perder la poca esperanza que aún tenían.

**¡Ojalá tomemos el camino adecuado!**



**MUCHAS GRACIAS  
POR SU ATENCIÓN**