

Evaluaciones Internacionales a Gran Escala: Retos y Perspectivas

Eugenio Gonzalez
Septiembre 17, 2020



(c) The IEA/ETS Research Institute (www.IERInstitute.org)

Sobre las "Evaluaciones a Gran Escala"?

- Su nombre viene del Ingles "Large-scale Assessments"
- En los últimos 20 – 30 años se han convertido en un nombre muy familiar en el ámbito de la educación
- Son citados frecuentemente por políticos y educadores en respaldo a sus decisiones y recomendaciones
- Ha habido un incremento sustancial en cuanto a las opciones de estudios, y los países que participan en estas evaluaciones



(c) The IEA/ETS Research Institute (www.IERInstitute.org)

1

Sobre las "Evaluaciones a Gran Escala"?

- El desarrollo y expansión en el uso de tecnologías de comunicación y procesamiento de información han acelerado el crecimiento de estos estudios, y han facilitado enormemente la coordinación e implementación de estas evaluaciones
- Numerosos estudios han sido diseñados e implementados:
 - ✓ Alfabetismo lector, competencias y conocimiento matemático y científico, educación ciudadana, etc.
- Ha habido mucho interés intrínseco de parte de los países, así como de organizaciones que financian programas de reforma educativa en los países

Qué son las "Evaluaciones a Gran Escala"?

- Son encuestas o estudios diseñados a fin de obtener información acerca de lo que grupos de personas saben o pueden hacer en ciertas áreas de interés
- Se conocen también como evaluaciones de grupo, o "group-score assessment" en las que el foco y la optimización no es reportar resultados individuales, sino mas bien resultados a nivel de grupo
- Se consideran de "gran escala", no por la cantidad de participantes, o la longitud de los instrumentos que se utilizan, sino porque su fin es evaluar varias áreas, en combinación con varias variables adicionales (demográficas, contexto, factores asociados, etc.) que son utilizadas para describir y contextualizar los resultados

Sobre las "Evaluaciones a Gran Escala"?

- Están diseñadas para tener poca o ninguna consecuencia para los individuos que participan en ellas
- Por esto, la optimización se enfoca en reportar resultados a nivel grupal y no enfatiza la medición individual
 - ✓ Se basan en métodos y procedimientos que optimizan reporte de estadísticos a nivel de grupo
 - ✓ Utilizan diseños de muestreo de matrices múltiples, teoría de respuesta del ítem, y modelaje poblacional

Qué es lo que NO son?

- No son hechas para ser usadas para reportar resultados de individuos o grupos pequeños (escuelas, clases, etc.)
 - ✓ Carecen de este nivel de precisión
- No tienen el fin de rendición de cuentas de individuos o grupos pequeños
- No se limitan a la medición de poblaciones escolares
 - ✓ También las hay de adultos, maestros, jóvenes no-escolarizados, etc.
- No tienen la intención de dictar currículo o métodos de enseñanza, si no mas bien medir los resultados o efectos de estos

Áreas Evaluadas Frecuentemente

- Conocimientos o competencias matemáticas y científicas
- Lectura
- Educación cívica y ciudadana
- Alfabetismo computacional y manejo de tecnologías de la información
- Resolución de problemas
- Alfabetismo financiero
- Ciencias sociales

Evaluaciones vs. Pruebas a Gran Escala

- Difieren en sus metas en cuanto a la intención del uso de sus resultados, aunque ambas son grandes en su magnitud
 - ✓ Las evaluaciones se preocupan más por reportar resultados a nivel grupal
 - ✓ Las pruebas se preocupan más con evaluar el desempeño de los individuos con el fin de diagnosticar, seleccionar, certificar, etc. de manera que la unidad de reporte es el individuo, y no el grupo al que pertenece
- Distintas metas conllevan a distintos métodos
- Se ha comprobado que el agregar resultados individuales a nivel grupal resulta en sesgos, y el evaluar a nivel individual requiere mucha más información que la necesaria para evaluar a nivel grupal

Propósitos de las Evaluaciones a Gran Escala

- Comparación
 - ✓ Proveer indicadores del desempeño de grupos de estudiantes y practicas educativas en los distintos grupos evaluados
- Análisis
 - ✓ Sugerir o evaluar hipótesis acerca de las relaciones entre variables y rendimiento
 - ✓ Identificar áreas de fortaleza y debilidad de los distintos grupos
- Inventario
 - ✓ Recabar información acerca de lo que saben y pueden hacer los participantes de distintos grupos
 - ✓ Cómo de distribuyen las destrezas y conocimientos en los distintos grupos demográficos

Qué tipo de Información Recaudan?

- Información demográfica a fin de clasificar a los participantes en los grupos demográficos
- Variables de contexto que describen las condiciones situaciones específicas de los participantes
- Variables que se cree están relacionadas con los resultados a fin de estudiar relaciones estadísticas
- Otras variable tales como opiniones, actitudes, comportamientos y prácticas
- Información sobre participación o exposición a distintos programas existentes

Evaluaciones Nacionales vs. Internacionales

- Nacionales
 - ✓ Las metas, marcos de referencia, métodos, procedimientos y todos los aspectos relacionados con la evaluación son establecidos y acordados por expertos nacionales utilizando un marco de referencia nacional
- Internacionales
 - ✓ Las metas, marcos de referencia, métodos, procedimientos y todos los aspectos relacionados con la evaluación son establecidos y acordados por un grupo internacional de expertos utilizando un marco de referencia internacional
 - ✓ Cubren un marco de referencia un poco mas amplio y diverso para atender la diversidad de los participantes
 - ✓ No necesariamente cubren aspectos específicos de algún país





Evaluaciones Nacionales vs. Internacionales

- Aunque hay diferencias claras de enfoque, comparten las siguientes características
 - ✓ El propósito de reportar resultados a nivel grupal
 - ✓ Estudiar relaciones entre variables
 - ✓ Proporcionar información a los responsables políticos
 - ✓ Comparten una cajita de herramientas con métodos y técnicas que específicamente optimizan la información para lograr estas metas

Métodos y Técnicas Utilizadas

- Diseños muestrales complejos
 - ✓ Multietapa, estratificados, y por conglomerados
- Diseños de pruebas complejos
 - ✓ Existe un grupo relativamente amplio de preguntas, de las cuales cada participante solo responde algunas
- Utilizan teoría de respuesta del ítem para reportar los resultados de las diferentes formas de la prueba, en una escala común
- Utilizan modelos estadísticos poblacionales para obtener estimadores óptimos de la distribución de los valores de rendimiento, y diferencias grupales
 - ✓ Modelos probabilísticos de acondicionamiento y valores plausible

Tantas opciones...

<p>Our long-haul fleet ↗</p>  <ul style="list-style-type: none">→ Airbus A380-800→ Boeing 747-8→ Boeing 747-400→ Airbus A340-600→ Airbus A340-300→ Airbus A330-300→ Boeing BBJ 737-800 IGW	<p>Our medium-haul fleet ↗</p>  <ul style="list-style-type: none">→ Airbus A321-100/200→ Airbus A320-200→ Airbus A319-100→ Boeing 737-500→ Boeing 737-300
<p>Our short-haul fleet ↗</p>  <ul style="list-style-type: none">→ Embraer ERJ-195→ Embraer ERJ-190→ Bombardier CRJ900→ Bombardier CRJ700→ ATR ATR72-500→ Bombardier DH8-Q400	<p>Our cargo fleet ↗</p>  <ul style="list-style-type: none">→ MD-11F

Como Decidir?

- Definir claramente el propósito y uso que se le va a dar a la información producto de la evaluación
- Establecer los riesgos y consecuencias de las alternativas posibles
- Hacer una lista con propósitos, ventajas y desventajas y ordenarlas en orden de prioridad
- No hay decisiones únicas o perfectas, pero si decisiones erradas
 - ✓ El proceso es una priorización de las distintas opciones basado en un análisis cuidadoso de los propósitos de la evaluación, y los recursos disponibles
- Es un proceso que responde a una serie de valores y prioridades
 - ✓ Importante hacer estos explícitos

Como Decidir?

- En resumen, el diseño y los elementos de una evaluación son el resultado de considerar y escoger cuidadosamente entre lo que se desea y lo que se puede hacer, dadas las circunstancias, recursos y priorización de opciones
 - ✓ Hay que comparar ventajas y desventajas de distintas opciones
 - ✓ Emitir un juicio global acerca de como proceder
 - ✓ No siempre podemos medir lo importante; y no siempre lo que medimos es importante

Evaluaciones a Gran Escala

- Por su extensión y complejidad, toman tiempo en planificar, diseñar e implementar
 - ✓ Año 1: Planificar, diseñar preparar instrumentos
 - ✓ Año 2: Aplicar y evaluar los instrumentos y procesos, preparar para la aplicación final
 - ✓ Año 3: Llevar a cabo la aplicación de campo reportar los resultados
 - ✓ Años 4+: Utilizar los datos con propósitos de investigación
- Por esto, generalmente solo se llevan a cabo cada varios años
- Las decisiones en base a los resultados también muestran su efecto por sobre muchos años

Evaluaciones Internacionales a Gran Escala

- Poblaciones Escolares
 - ✓ ERCE: Estudio Regional Comparativo y Explicativo
 - ✓ TIMSS: Trends in International Mathematics and Science Study
 - ✓ PIRLS: Progress in International Reading Literacy Study
 - ✓ ICCS: International Civic and Citizenship Education Study
 - ✓ PISA: Program for International Student Assessment
 - ✓ ICILS: International Computer and Information Literacy Study
- Poblaciones de adultos y juvenes
 - ✓ PIAAC: Programme for the International Assessment of Adult Competencies
 - ✓ PISA-D Strand C: Jóvenes de 14 – 16 años fuera de la escuela

Retos a Futuro

- Atención a asuntos locales
- Creación de diseños innovadores para mejoras a la medición
- Mejoras a la medición de individuos
- Transición de papel a computadora/tablet
- Adaptación y Expansión de los dominios existentes
- Introducción de nuevos dominios de interés

Retos a Futuro

- Utilización de los datos de proceso
 - ✓ Tiempo de respuesta, etc.
 - ✓ Secuencias de acción
- Validez y utilidad de la información a fin de tomar decisiones y hacer recomendaciones pedagógicas
 - ✓ Relevancia para los participantes de tan diversos contextos
- Entender las limitaciones en el uso de la información

Gracias!



(c) The IEA/ETS Research Institute (www.IERInstitute.org)

20