

Las pruebas PISA: luces y sombras

Álvarez Morán, S.; Carleos Artime, C.; Corral Blanco, N.

Facultad de Ciencias
C/ Federico García Lorca, s/n.
33007-OVIEDO
www.unioviedo.es/ciencias

Octubre 2017



PISA es uno de los principales referentes en las evaluaciones educativas a gran escala.

- ▶ Aborda la pregunta «¿Qué es importante que los ciudadanos sepan y puedan hacer?»
- ▶ Evalúa competencias
- ▶ Relaciona los resultados de los estudiantes en las pruebas cognitivas con su contexto socio-económico y cultural, sus actitudes, características de centro, etc.
- ▶ Intenta obtener datos comparables a nivel internacional
- ▶ Permite analizar la evolución temporal de los sistemas educativos

A pesar de sus cualidades también recibe críticas

- ▶ La OCDE es una organización dedicada al desarrollo económico y potencia la visión económica de la educación
- ▶ La diferencia entre el enfoque educativo de la OCDE y las de otras agencias como la UNESCO y UNICEF
- ▶ Se centra en Ciencias, Lectura y Educación, olvidando otros aspectos básicos, como la salud, el desarrollo ético, la creatividad, etc.
- ▶ La metodología que utiliza no es la adecuada
- ▶ PISA evalúa la actividad escolar de manera homogénea, sin tener en cuenta la gran diversidad de los sistemas educativos
- ▶ Los rankings que establece entre los sistemas educativos de los países no son fiables

En la charla se van a mostrar algunos puntos débiles que pueden afectar a la fiabilidad de las conclusiones

1. Valores Plausibles (B1)
2. Fiabilidad y validez de los cuestionarios de contexto de los estudiantes (B2)
3. Eliminación de bloques de items y de índices complejos (B3)
4. Cambios en los items de cuestionarios de contexto (B4)
5. Informes PISA. En ocasiones, las conclusiones son de difícil justificación o se basan en evidencias muy débiles (B5)

B1. Valores Plausibles (1)

VALORES PLAUSIBLES: Rango de capacidades de un estudiante según sus respuestas a los items.

Se obtienen usando los siguientes procedimientos:

- ▶ El modelo Rasch de los Test de Respuesta al ITEM (TRI)
- ▶ Regresión lineal con variables latentes
- ▶ Estimación de la distribución a posteriori de la capacidad de cada estudiante

B1. Valores Plausibles. Modelo Rasch (2)

X_i representa la respuesta al ítem i

$X_i = 1 \Leftrightarrow$ Acierto en la respuesta al ítem i

Para un estudiante de capacidad θ y un ítem de dificultad δ_i

$$P(X_i = 1 | \theta, \delta_i) = \frac{\exp(\theta - \delta_i)}{1 + \exp(\theta - \delta_i)}$$

Capacidad	Dificultad	Probabilidad	
		Acuerdo	Fallo
-1	0	0.27	0.73
0	0	0.50	0.50
3	1	0.88	0.12
2	2	0.50	0.50

B1. Valores Plausibles. Modelo Rasch (3)

Sea $\mathbf{x} = (x_1, \dots, x_4)$ un test con cuatro items.

$$P(\mathbf{x}|\theta, \delta) = \prod_{i=1}^4 P(x_i|\theta, \delta)$$

Probabilidad de las respuestas a los items

	Dificultad items				$P(\mathbf{x} \theta, \delta)$		
	$\delta_1 = -1$	$\delta_2 = 0$	$\delta_3 = 0$	$\delta_4 = 1$	$\theta = -1$	$\theta = 0$	$\theta = 1$
\mathbf{x}_1	1	0	0	0	0.2354	0.1336	0.0318
\mathbf{x}_2	0	1	0	1	0.0117	0.0181	0.0117
\mathbf{x}_3	-	1	-	1	0.0321	0.1345	0.3655
\mathbf{x}_4	-	-	-	-	-	-	-

Estimaciones de la competencia de los estudiantes, según sus respuestas al test:

$$(\hat{\theta}_1 = -1,2; \hat{\theta}_2 = 0,0; \hat{\theta}_3 = 2,0)$$

B1. Valores Plausibles (4)

- ▶ A partir de la muestra se estiman los parámetros asociados a los ítems.

$$P(X|\delta) = \prod_{j=1}^N \left(\int_{\Theta} P(x_j|\theta, \delta) g(\theta) d\theta \right)$$

- ▶ Regresión con las componentes principales, *y*, del cuestionario de contexto que explican el 80 % de variabilidad total

$$P(\theta|y) \equiv N(y|\Gamma, \Sigma); \text{ se estiman } (\Gamma, \Sigma)$$

- ▶ Se calcula la distribución a posteriori de la capacidad de cada estudiante y se generan los valores plausibles

$$P(\theta|x, y) \propto P(x|\theta) P(\theta|y)$$

B1. Valores Plausibles. Conclusiones (5)

- ▶ Algunas investigaciones indican que no se verifican las condiciones para aplicar el modelo Rasch
- ▶ En las bases de datos no aparecen las estimaciones directas de las competencias en Ciencias, Lectura y Matemáticas
- ▶ No se sabe exactamente cómo se calculan los valores plausibles y tampoco 'está abierto' el código utilizado. (The software DGROUP. Rogers, Tang, Lin, & Kandathil, 2006)
- ▶ Una proporción muy alta de estudiantes no pasan ninguna prueba de Lectura (58 %) o Matemáticas (57 %).
Se les imputan los valores plausibles

B1. Valores plausibles. Conclusiones (6)

- ▶ Estudiantes con respuestas idénticas en las pruebas cognitivas podrían tener valores plausibles muy distintos
Sean A y B dos estudiantes con un 9 (sobre 10). Sus padres tienen estudios 'primarios' y 'superiores', respectivamente
El estudiante A tendría valores plausibles más bajos que el B
- ▶ ¿Las competencias que miden las pruebas cognitivas de PISA son las adecuadas para nuestros estudiantes?

B2. Fiabilidad y Validez

One of the major challenges of an international study such as PISA is the cross-cultural validity and applicability of all instruments.

Se observaron situaciones que afectan a este objetivo. Por ejemplo:

- ▶ Algunos constructos del cuestionario de contexto se asocian de manera inesperada con el rendimiento.
- ▶ Las diferencias interculturales en los estilos de respuesta pueden producir un sesgo importante con las escalas Likert

PISA plantea diferentes alternativas para eliminar esos problemas

1. Preguntas trampa ('overclaiming')
2. Preguntas de anclaje ('anchoring vignettes')
3.

Technical report. PISA 2012. Capítulo III

B2. Preguntas trampa

Preguntas trampa 'overclaiming' (Cuestionario Estudiantes 2012)

En relación con los siguientes conceptos matemáticos, ¿en qué medida te resultan familiares los términos?

- ▶ Número genuino
- ▶ Escalas subjuntivas
- ▶ Funciones declarativas

	Nunca	Una vez	Varias	A menudo	Entiendo
Número genuino	66,0	16,2	8,5	5,1	4,2
Escalas subjuntivas	67,8	15,5	8,7	4,5	3,5
Funciones declarativas	68,9	12,4	8,8	4,8	5,1

B2. Viñetas de anclaje de control del aula (1)

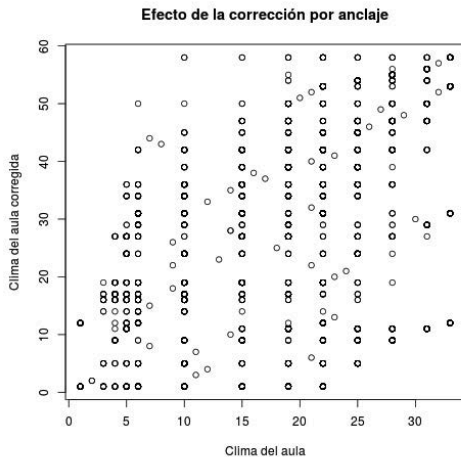
Lee la descripción de estos tres profesores y di tu grado de acuerdo con la afirmación final. (Cuestionario Estudiantes 2012)

- ▶ Los estudiantes de la Sra. 'A' la interrumpen con frecuencia en sus clases. La profesora siempre llega con cinco minutos de antelación. **La Sra. 'A' controla su clase.**
- ▶ Los estudiantes de la Sra. 'B' están tranquilos y ordenados en su clase. La profesora llega en punto a su clase. **La Sra. 'B' controla su clase.**
- ▶ Los estudiantes del Sr. 'C' le interrumpen con frecuencia en sus clases. Por eso a menudo llega cinco minutos tarde a su clase. **El Sr. 'C' controla su clase.**

Muy de acuerdo; De acuerdo; En desacuerdo; Muy en desacuerdo

Con estas viñetas se corrige la puntuación de Clima del aula

B2. Viñetas de anclaje (2)



Correlación de Pearson = 0.55

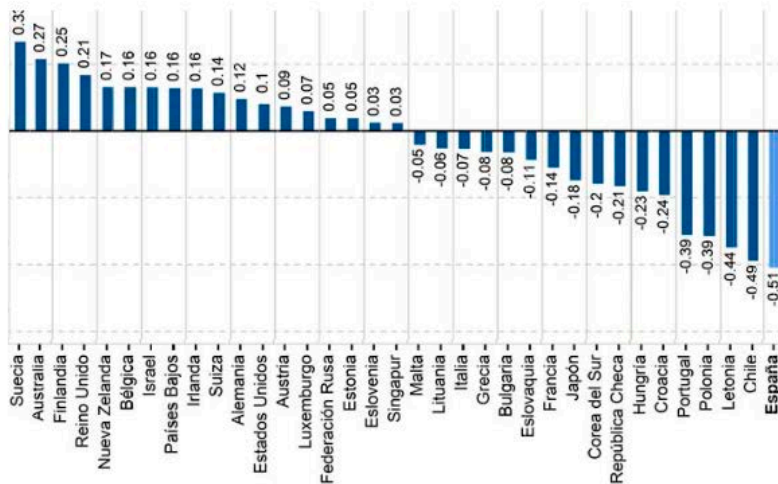
Comentarios del informe PISA de resultados. Volumen I. 2012

- ▶ Los resultados de las pruebas de campo y de la encuesta principal mostraron que las viñetas conseguían resultados más claros en el estudio de los comportamiento dentro de las aulas.
- ▶ Para usar las viñetas de anclaje se deben cumplir la hipótesis de 'equivalencia de la viñetas' y la de 'respuestas consistentes'
- ▶ Es recomendable ser prudentes con la interpretación de los índices ajustados y hacerlo usando, a la vez, los índices sin ajustar

PISA-2012-results-volume-I.pdf. Annex A6. PISA 2012 Technical Report

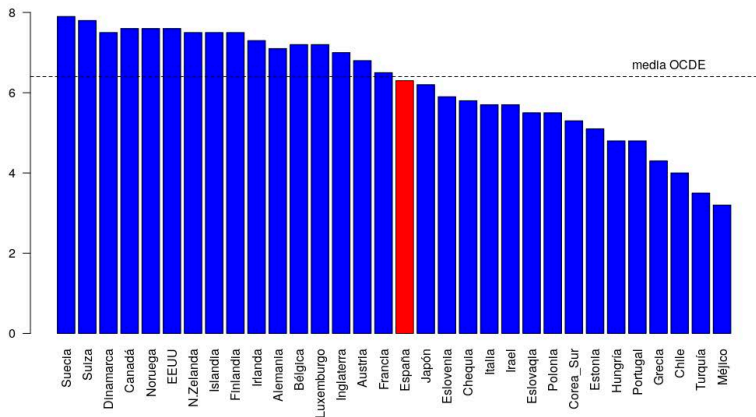
B2. Índice Social económico y cultural (ISEC) (1)

PISA 2015. Informe español. Figura 3.12 (Parte central de la gráfica)



B2. Índice Social económico y cultural. Calidad de Vida; OCDE 2016. (2)

¿Teniendo en cuenta la gráfica anterior, es verosímil la posición en el ISEC de algunos países?



B2. Validez y fiabilidad. Conclusiones

- ▶ Ninguno de los procedimientos planteados para mejorar la fiabilidad y validez de los cuestionarios de contexto se utilizó explícitamente en los informes de PISA 2012.
- ▶ En 2015 las preguntas trampa, las variables con puntuaciones corregidas por viñetas de anclaje, etc. no aparecen en las bases de datos
- ▶ En el capítulo III del Informe Técnico de PISA 2015, no aparece el problema de validez de las comparaciones con grandes diferencias interculturales ni las alternativas tratadas en el 2012

PISA-2012-results-volume-I.pdf. Annex A6. PISA 2012 Technical Report. Capítulo III
Technical report. PISA 2015. Capítulo III

B3. Eliminación de bloques de items

Ejemplo de bloque sobre la forma de estudiar para un examen de matemáticas. (Cuestionario Estudiantes 2012)

En cada grupo de tres items, elige la alternativa que mejor describa tu enfoque de las matemáticas.

- ▶ Cuando estudio para un examen de matemáticas intento centrarme en las partes que considero más importantes
- ▶ Cuando estudio para un examen de matemáticas intento entender los nuevos conceptos relacionándolos con las cosas que ya conozco.
- ▶ Cuando estudio para un examen de matemáticas, intento aprender todo lo que puedo de memoria.

B3. Eliminación de bloques de items

Bloque de resolución de problemas de la vida cotidiana
(Cuestionario Estudiantes 2012)

Supón que has estado enviando mensaje por tu teléfono móvil durante varias semanas. Sin embargo, hoy no lo puedes hacer
Para intenta resolver el problema

- ▶ Pulso todos los botones tratando de descubrir qué va mal
- ▶ Pienso qué puede haber provocado el problema y lo intento resolver
- ▶ Leo el manual
- ▶ Le pregunto a un amigo

1. Lo hago seguro;

3. Probablemente no lo hago;

2. Probablemente lo hago

4. Seguro que no lo hago

B3. Eliminación de bloques de items y de índice complejos

Ejemplos de 'constructos' que desaparecieron entre 2012 y 2015.

- ▶ Perseverancia (PISA 2012 Results: Ready to Learn. Vol. III)
- ▶ Métodos de estudio para preparar un examen de matemáticas
- ▶ Métodos de resolución de problemas de la vida cotidiana

Aunque los temas tienen bastante interés se eliminaron sin ninguna explicación

B4. Cambios en los cuestionarios

ENSEÑANZA DIRIGIDA POR EL PROFESOR

¿Con qué frecuencia ocurren estas cosas en tus clases de la asignatura 'xxx'? (Cuestionario estudiantes)

- ▶ El profesor establece unos objetivos claros para nuestro aprendizaje
- ▶ El profesor nos anima a presentar nuestras ideas y razonamientos en profundidad
- ▶ El profesor nos hace preguntas para comprobar si hemos entendido lo que explica
- ▶ Al inicio de una clase el profesor hace un pequeño resumen de la clase previa
- ▶ El profesor nos informa de lo que tenemos que aprender

En todas las clases; En la mayoría; En algunas; Nunca o casi nunca

ENSEÑANZA DIRIGIDA POR EL PROFESOR

¿Con qué frecuencia ocurren estas cosas en tus clases de la asignatura 'xxx'? (Cuestionario estudiantes)

- ▶ El profesor explica conceptos científicos.
- ▶ Se lleva a cabo un debate entre toda la clase y el profesor.
- ▶ El profesor explica nuestras preguntas.
- ▶ El profesor demuestra un concepto.

En todas las clases; En la mayoría; En algunas; Nunca o casi nunca

B4. ENSEÑANZA DIRIGIDA POR EL PROFESOR

TEACHER DIRECTED INSTRUCTIONS (2012. TCHBEHTD)

- ▶ Establecer objetivos claros
- ▶ Animarnos a pensar y razonar
- ▶ Comprobar nuestro grado de aprendizaje
- ▶ Resumir las lecciones anteriores
- ▶ Informarnos sobre nuestros objetivos de aprendizaje

TEACHER DIRECTED INSTRUCTIONS (2015. TDTEACH)

- ▶ El profesor explica conceptos científicos.
- ▶ Se lleva a cabo un debate entre toda la clase y el profesor.
- ▶ El profesor explica nuestras preguntas.
- ▶ El profesor demuestra un concepto.

B5. METODOLOGÍAS DE ENSEÑANZA (1)

How science is taught at school can make a big difference for students

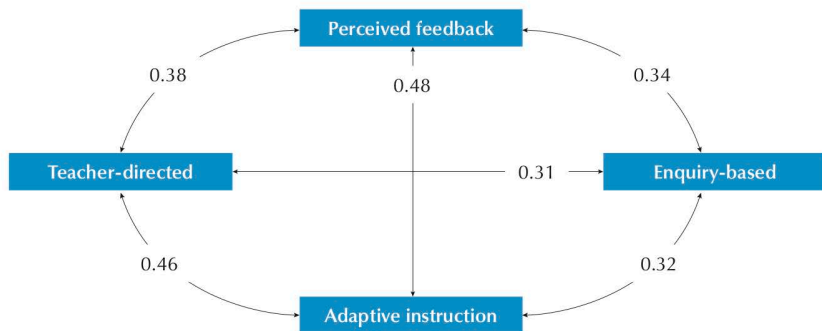
En PISA 2015 se consideran cuatro tipos de metodologías:

- ▶ **Dirigida por el profesor (TDTEACH)**: El profesor explica la materia de forma clara y bien estructurada y responde a las preguntas de sus estudiantes
- ▶ **Adaptativa (ADAINST)**: Los profesores adaptan las clases a las necesidades y conocimientos de todos los estudiantes del aula
- ▶ **Retroalimentación (PERFEED)**: Los profesores informan a los estudiantes sobre la tarea o evaluación realizada.
- ▶ **Basada en proyectos (IBTEACH)**: Los estudiantes tienen que realizar experimentos y extraer conclusiones para profundizar en sus conocimientos

PISA 2015 Results. POLICIES AND PRACTICES FOR SUCCESSFUL SCHOOLS. VOLUME II

B5. METODOLOGÍAS DE ENSEÑANZA (2)

Figure II.2.12 ■ **Relationships among instructional practices in science**
Correlations at the student-level based on students' reports, OECD average



¿Cómo interpretar que todas las correlaciones sean positivas?

B4. ENSEÑANZA DIRIGIDA POR EL PROFESOR (1)

TCHBEHTD y TDTEACH (componentes principales)

TEACHER DIRECTED INSTRUCTIONS (2012. TCHBEHTD)

- ▶ Establecer objetivos claros
- ▶ Animarnos a pensar y razonar
- ▶ Comprobar nuestro grado de aprendizaje
- ▶ Resumir las lecciones anteriores
- ▶ Informarnos sobre nuestros objetivos de aprendizaje

TEACHER DIRECTED INSTRUCTIONS (2015. TDTEACH)

- ▶ El profesor explica conceptos científicos.
- ▶ Se lleva a cabo un debate entre toda la clase y el profesor.
- ▶ El profesor explica nuestras preguntas.
- ▶ El profesor demuestra un concepto.

B5. ENSEÑANZA DIRIGIDA POR EL PROFESOR (2)

Relaciones con MATEMÁTICAS, LECTURA y CIENCIAS

	TCHBEHATD			TDTEACH		
País	PISA 2012 (-)			PISA 2015 (+)		
39 países	MATE	LECT	CIEN	MATE	LECT	CIEN
España	-0.06	-0.05	-0.04	0.15	0.12	0.14
Francia	-0.15	-0.17	-0.13	0.10	0.10	0.12
Rusia	-0.11	-0.14	-0.12	0.13	0.08	0.11
Brasil	-0.18	-0.19	-0.17	0.14	0.12	0.15

País	PISA 2012 (+)			PISA 2015 (+)		
10 países	MATE	LECT	CIEN	MATE	LECT	CIEN
Korea	0.06	0.08	0.05	0.02	0.01	-0.01
Finland	0.04	-0.03	0.02	0.14	0.20	0.20

Australia y Japón mantienen correlaciones positivas estables

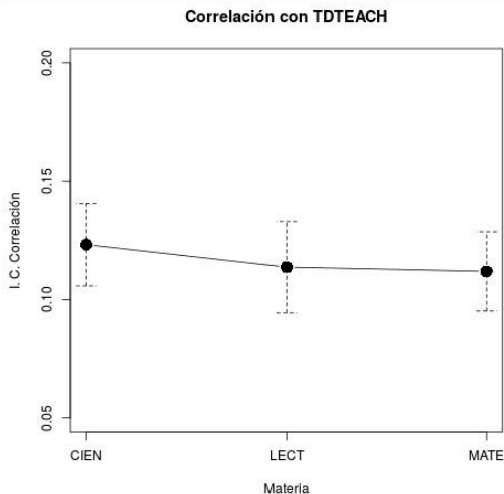
B5. ENSEÑANZA DIRIGIDA POR EL PROFESOR (3)

En 2015 la relación de **Enseñanza dirigida por el profesor (TDTEACH)** con Ciencias resultó positiva en 67 países (95%).

El uso de la **TDTEACH** solo se midió en asignaturas de Ciencias, ¿qué relación puede tener con Matemáticas y Lectura?

- ▶ Similar a la de Ciencias
- ▶ Bastante más baja (casi nula)

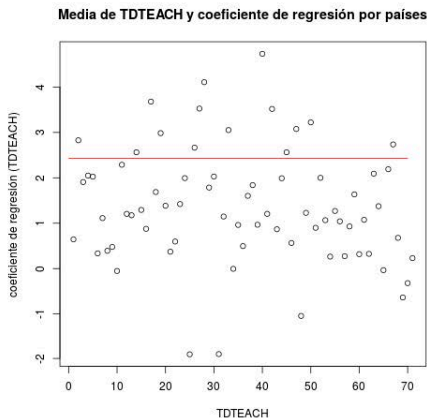
B5. ENSEÑANZA DIRIGIDA POR EL PROFESOR (4)



¿TDTEACH mide realmente lo que se quiere?

B5. ENSEÑANZA DIRIGIDA POR EL PROFESOR (5)

$$CIENCIAS = \beta_0 + \beta_1 ESCS + \beta_2 RGeneral + \beta_3 TDTEACH + \epsilon$$



Coeficientes de regresión de TDTEACH por países en PISA 2015

¿TDTEACH influye en el rendimiento de Ciencias?

B5. ENSEÑANZA BASADA EN PROYECTOS (1)

Comentarios del informe PISA 2015. Vol II

- ▶ La Enseñanza basada en proyectos (IBTEACH) es muy importante en la enseñanza de las ciencias físicas y de la vida
- ▶ Los estudios previos indican que IBTEACH puede mejorar el aprendizaje de los estudiantes y la adquisición de destrezas como el razonamiento crítico.
- ▶ Teniendo en cuenta el nivel socio educativo y económico, la relación de IBTEACH con el rendimiento en Ciencias, resultó negativo en 56 países. Estas relaciones deben interpretarse con precaución.

¿Cómo explicar estos resultados de apariencia contradictoria?

Comentarios del informe PISA 2015. Vol II

- ▶ El laboratorio sólo mejora el aprendizaje si los estudiantes realizan actividades manuales y manejan conceptos
- ▶ El tiempo dedicado a los laboratorios se emplea de manera ineficiente
- ▶ En ocasiones, las 'actividades manuales' se utilizan para hacer más atractiva la ciencia a los estudiantes con problemas
- ▶ Los estudiantes solo trabajan realmente cuando hay buen material y el profesor está bien formado

¿IBTEACH mide 'bien' la enseñanza basada en proyectos?

Problemas en el diseño del cuestionario de contexto

- ▶ Basada en Proyectos (IBTEACH) se mide conjuntamente sobre todas las asignaturas de Ciencias.
- ▶ Dirigida por el Profesor (TDTEACH) se mide sobre una asignatura concreta elegida por cada estudiante
- ▶ No se conoce la variabilidad intra-aula de la percepción de los estudiantes sobre la metodología del profesor

1. Valores plausibles

- ▶ Información insuficiente sobre su cálculo.
- ▶ No se cumplen las condiciones del modelo Rasch para todos los items
- ▶ ¿Qué relación existe entre los conocimientos que mide PISA y los que se piden en las escuelas?

2. Dificultades para hacer estudios longitudinales

- ▶ Eliminación de bloques de preguntas interesantes (Métodos de preparación de los exámenes)
- ▶ Cambios en los cuestionarios de contexto (Enseñanza dirigida por el profesor)
- ▶ Desaparecen algunos procedimientos para afrontar las diferencia interculturales, como las viñetas de anclaje

3. Resultados dudosos

- ▶ Correlaciones positivas entre las metodologías
- ▶ Influencia de TDTEACH en los resultados de Ciencias
- ▶ Falta de justificación ante resultados sorprendentes (IBTEACH)
- ▶ Dudas razonables sobre el significado real de algunos constructos o su precisión (TDTEACH, IBTEACH, ISEC)

Los datos e informes de PISA proporcionan una información muy valiosa, pero es necesario mejorar sus instrumentos de medida y ser muy prudentes con las interpretaciones

- ▶ Carabaña, J. La inutilidad del informe PISA. (2015). Editorial La Catarata
- ▶ Fernandez-Cano, A. Una crítica metodológica de las evaluaciones PISA. (2016) RELIEVE, Vol 22. nº 1
DOI: <http://dx.doi.org/10.7203/relieve.22.1.8806>
- ▶ Kreiner S., Christensen, K. L. Analyses of Model Fit and Robustness, A new look at the PISA Scaling plausible values in Large-scale surveys. Psychometrika. Vol 79. Issue 2, pp 210-231
- ▶ PISA Results 2012. Vol I, II, III. OECD
- ▶ PISA Results 2015. Vol I, II, III. OECD
- ▶ Wu, M. The role of plausible values in large-scale surveys 31. (2005), 114-128 - Studies in Educational Evaluation.

Gracias por la atención