

---

# **DIFICULTADES Y LIMITACIONES DE LOS MODELOS DE EVALUACIÓN DE RESULTADOS PARA LA MEJORA DE LOS CENTROS Y DEL SISTEMA EDUCATIVO**

**Juan Andrés Ligeró Lasa**

Profesor de la Universidad Carlos III de Madrid

Codirector del Máster en Evaluación de Programas y Políticas  
Públicas, UCM.

# DEFINICIONES

---

- Recoger y analizar sistemáticamente una información para poder determinar el valor o mérito de una intervención (*Joint Committee Standards Education*)
- Valoración sistemática de las operaciones o de los resultados de un programa o política comparada con un conjunto de estándares implícitos o explícitos como medio para contribuir a la mejora del programa o política (Weiss,1998: 4)
- *Evaluation is judging merit or worth of an entity (Alkin, 2011: 9)*

# Evaluación Profesional (Alkin)

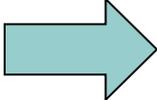
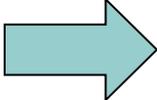
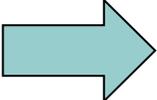
---

- Evaluación de productos
- Evaluación de personas
- Evaluación de programas (programas, proyectos, servicios e instituciones)

# Diferencias terminológicas

---

*We evaluate programs;  
we assess client knowledge;  
and we appraise staff. (Alkin 2011: 10)*

<i>Evaluation</i>		<i>Programs, centers</i>
<i>Assesment</i>		<i>Assessment test, assessing students (PISA)</i>
<i>Appraisal</i>		<i>Appraise staff</i>

# Posible adaptación de los términos

---

- **Evaluación** de programas, centros y sistemas educativos.
- **Exámenes** de alumnado
- **Valoración**, habilitación(?) de personal

# Propósitos (Greene, 2007)

---

- **Toma de decisiones y el rendimiento de cuentas** (responsabilización, accountability) **responsables políticos u otros actores con capacidad de decisión.**
- La **mejora del programa y el desarrollo organizacional** provee de una información valiosa para los **directivos u otros responsables** de las operaciones habituales del programa.
- La **comprensión** en profundidad y contextualizada del programa y sus prácticas normalmente atiende a las necesidades informativas e intereses del **personal del programa** y a veces también de los **participantes.**
- Promover más **justicia y equidad** en el programa y en el contexto en el que se evalúa, sirve usualmente a los intereses de los **participantes, sus familias y la comunidad.**
-

# Propósitos (Greene, 2007)

---

- Fundamentar la toma de decisiones y el rendimiento de cuentas (**responsabilización, *accountability***) puede dar respuesta a las necesidades informativas e intereses de los responsables políticos u otros actores con capacidad de decisión.
- La **mejora del programa** y el desarrollo organizacional provee de una información valiosa para los directivos u otros responsables de las operaciones habituales del programa.
- La **comprensión en profundidad y contextualizada del programa** y sus prácticas normalmente atiende a las necesidades informativas e intereses del personal del programa y a veces también de los participantes.
- **Promover más justicia y equidad** en el programa y en el contexto en el que se evalúa, sirve usualmente a los intereses de los participantes, sus familias y la comunidad.

# Propósitos y actores (Greene, 2007)

Propósitos evaluativos	Actores que los requieren
Toma de decisiones y rendimiento de cuentas (responsabilización, <i>accountability</i> )	Responsables políticos u otros actores con capacidad de decisión
Mejora del programa y desarrollo organizacional	Dirección u otros/as responsables de las operaciones habituales del programa
Comprensión del programa	Personal del programa y a veces también los/as participantes
Justicia y equidad social	Participantes, sus familias y la comunidad

# PRUEBAS Y EVALUACIONES (Rita López Panach)

---

NOMBRE Y FECHA DE LA PRUEBA	QUIÉN LA HACE
<b>LEA INFANTIL (5 AÑOS)</b>	
<b>LEA</b> (Lectura, Escritura, Aritmética) de 2º de Primaria: 5 de junio 2014.	
<b>Trinity 2º Primaria (a1)</b> . mayo-junio	
<b>Trinity 4º Primaria (a2)</b> (mayo-junio)	
<b>CDI</b> (Conocimiento y Destrezas Indispensables) 6º de Primaria. (7 de mayo); 3º ESO (29 de abril)	Viceconsejería, Comunidad de Madrid
<b>ED</b> (Evaluación de diagnóstico)	Dirección General
<b>British Council:</b> Placement Test, KET, PET (mayo-junio)	British (¿Comunidad de Madrid también?)
<b>Trinity 2º A1</b>	Trinity y Comunidad de Madrid (¿)
<b>Trinity 4º A2</b>	Trinity y Comunidad de Madrid (¿concertados y privados?) Título por valor de 2 años.
<b>Pruebas digitales 1º, 2º 3º y 4º de la ESO (mayo)</b>	
<b>PAU</b>	Dirección general (¿?)
<b>PISA</b> (Program for International Students Assesment)	OCDE (Organisation for Economic Co-operation and Development)
<b>Evaluación durante la etapa 3ª primaria</b>	LOMCE
<b>Evaluación final 6º primaria</b>	LOMCE
<b>Evaluación final ESO</b>	LOMCE

Mayoritariamente se han extendido las

Pruebas externas y estandarizadas  
de rendimiento del alumnado.

# Propósitos declarados

---

- **Toma de decisiones** padres (Fedea, Lomce, CM). Toma de decisiones admon. (CM)
- **Motivación y Mejora de rendimiento** (Fedea, Lomce)
- **Rendición de cuentas y transparencia** (Fedea, Lomce, CM)

# 1. Confusión *assesment* y evaluación

	<b>Evaluación</b>	<b><i>Assesment</i></b> <b>(exámen, prueba)</b>
<b>Objeto sobre el que emitir el juicio de valor</b>	Evaluación de programas, sistemas educativos, centros, políticas,...	Estudiantes, personas.
<b>Métodos</b>	CIPP: Context, Inputs, Process, Product. Responsive Evaluation,...	<i>Curriculum-Based External Exit Examination</i> (CBEEE); PISA (Program International Student Assesment)

## 2. Solo evalúo los resultados

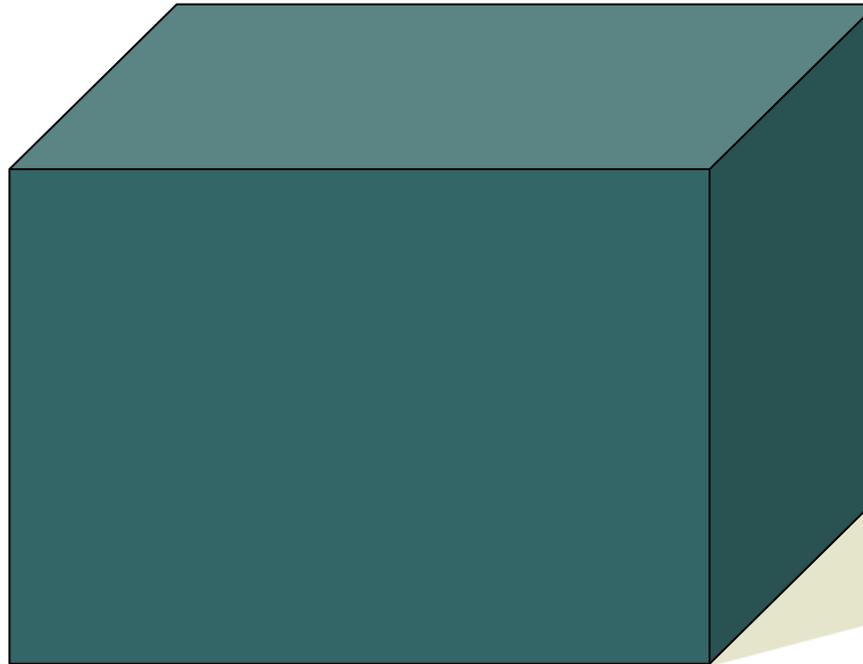
---

Emisión de un juicio de valor sobre la totalidad del centro (docencia, medios, competencias profesionales,...) solo apoyándose en una parte.

*Las críticas más comunes son que estos estudios conducen a una información final que no es oportuna ni pertinente para mejorar los procesos del programa, que la información a menudo es demasiado limitada para conformar una base suficiente para juzgar el mérito y el valor del objeto, que no contemplan los efectos imprevistos negativos y positivos y que pueden acreditar objetivos que no merecen la pena (Stufflebeam 2011:18)*

# CAJA NEGRA

---



*Sobre la base de observaciones cualitativas, Weiss y Rein (1970) critican las evaluaciones que no informan sobre lo que ocurre en la “caja negra”, refiriéndose con este término al proceso de implementación del programa (Ballart, 1996, pág 329).*

*Se desconocen los mecanismos que realmente provocan los resultados*

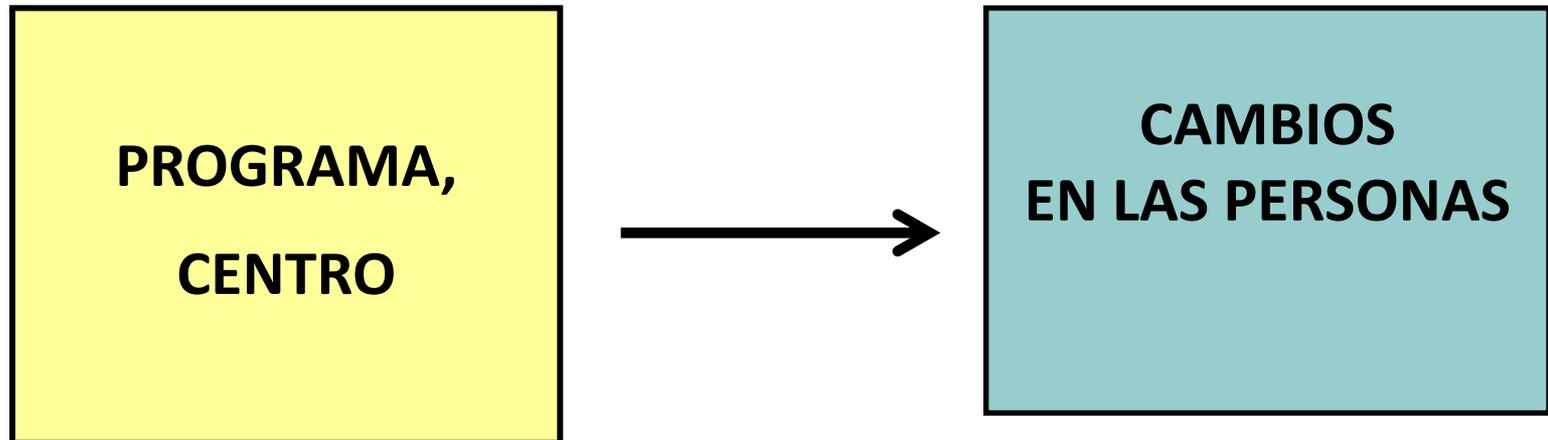
### 3. Resultados estandarizados

---

Se evalúan unas competencias/conocimientos definidas de forma estandarizada que pueden **coincidir o no con los objetivos de los proyectos educativos de centro.**

# Hipótesis principal de un centro educativo

**El centro produce cambios**



# Test estandarizados vs Evaluación objetivos

---

<b>Evaluación por objetivos (Tyler) 1942</b>	<b>Evaluación(<i>assessment</i>) test estandarizados (Pretyler)</b>
Se define los objetivos por centros (cambios, logros en los/as estudiantes) adecuados a su contexto y a su programa.	Medidas únicas y homogéneas para todo el sistema, al margen de los proyectos de centro.

# Contradicción Lomce

---

*Es necesario que cada centro tenga la capacidad de identificar cuáles son sus fortalezas y las necesidades de su entorno, para así poder tomar decisiones sobre cómo mejorar su oferta educativa y metodológica en ese ámbito, en relación directa, cuando corresponda por su naturaleza, con la estrategia de la administración educativa. (Lomce, VII)*

VS

*Por otro lado, estas pruebas normalizan los estándares de titulación en toda España, indicando de forma clara al conjunto de la comunidad educativa cuáles son los niveles de exigencia requeridos e introduciendo elementos de certeza, objetividad y comparabilidad de resultados. (lomce VIII)*

## 4. Medición conocimiento y no desarrollo personal

---

**Logros cognitivos:** lectura, matemáticas, conocimiento científico y otras “habilidades para la vida”. PISA

**Logros no cognitivos:** P. ej. Motivación para el aprendizaje. PISA

**Conocimientos:** matemáticas (ejercicios y problemas) y lengua (dictado, lectura, conocimiento general y preguntas relacionadas con un texto) Fedea.

Las evaluaciones externas estandarizadas suelen medir más conocimientos que procesos no cognitivos o de desarrollo personal.

De nuevo posible conflicto con el trabajo de los centros.

## 5. Resultados no atribuibles a los centros

---

Problemas serios de atribución de los resultados a los centros.

- El rendimiento de los jóvenes puede ser provocado por varios factores: su centro educativo, su bagaje económico y social, sus características personales, situaciones del contexto,....
- Se atribuye al colegio los resultados que, por ejemplo, provienen de estatus socio-económico.

# 5. Resultados no atribuibles a los centros

---

Como ya se ha mencionado, son múltiples los factores que pueden condicionar el rendimiento de los alumnos. Las características de los centros educativos pueden causar diferencias en los resultados. Por ejemplo los factores geográficos o socioeconómicos y culturales de la población en la que se encuentra el centro o los recursos humanos, materiales o financieros puestos a su disposición. Además, la variabilidad de los resultados académicos puede ser atribuible a procesos internos del centro, más complejos de cuantificar, como el grado de autonomía, la eficacia o eficiencia en la gestión de procesos organizativos y educativos, el funcionamiento del equipo docente o el trabajo de aula, etc. Por supuesto, las diferencias también pueden deberse a las características propias de los alumnos y producirse dentro de un mismo centro: cambios en su trayectoria escolar, la confianza que en ellos se deposita, etc. Un alumno puede mostrar capacidades, esfuerzo, actitudes o niveles de motivación e interés diferentes a los de sus compañeros de clase, aunque sean homogéneos desde el punto de vista socioeconómico.

PISA

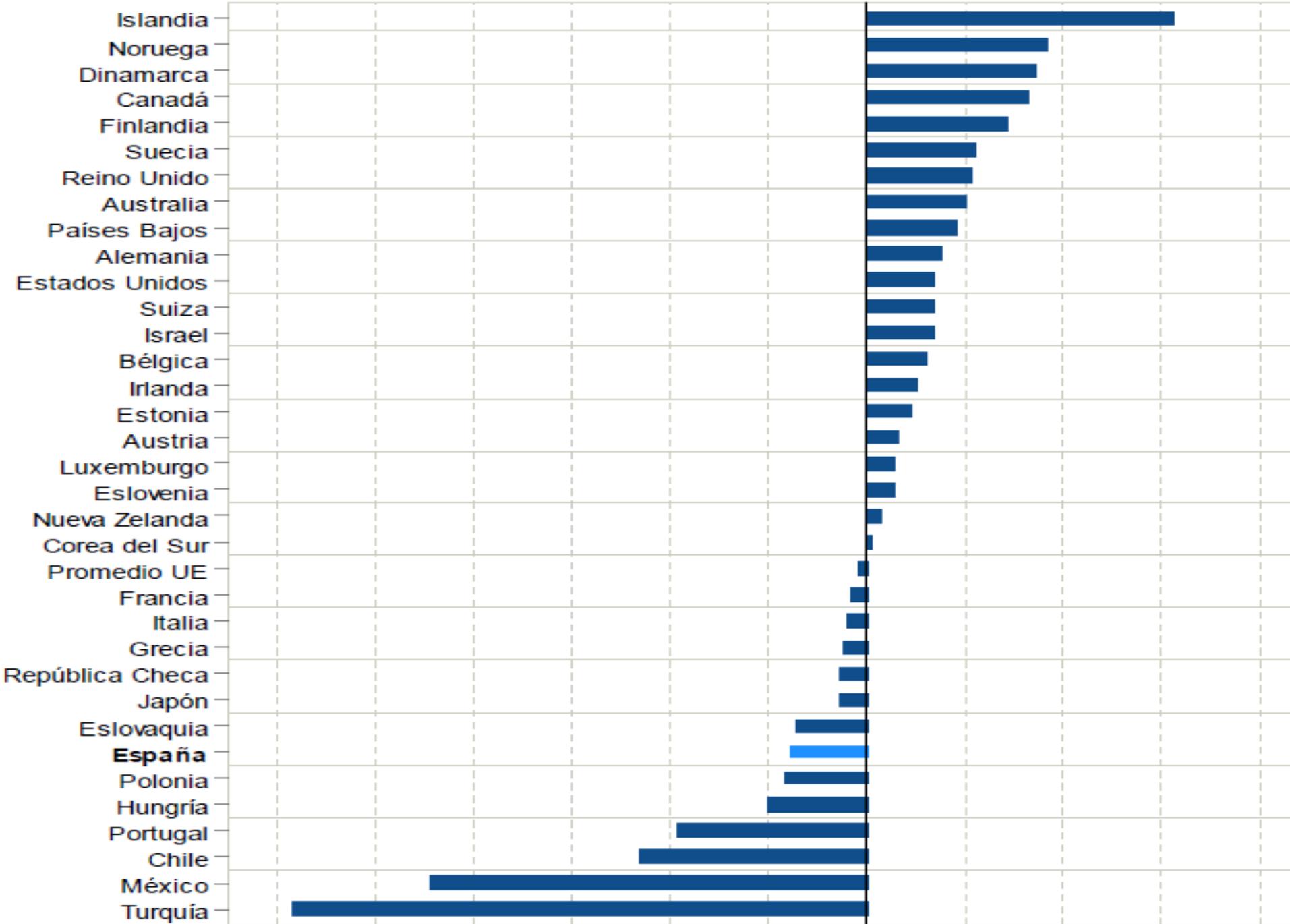


Figura 3.2. Valor promedio del ESCS de las CCAA participantes en PISA 2012

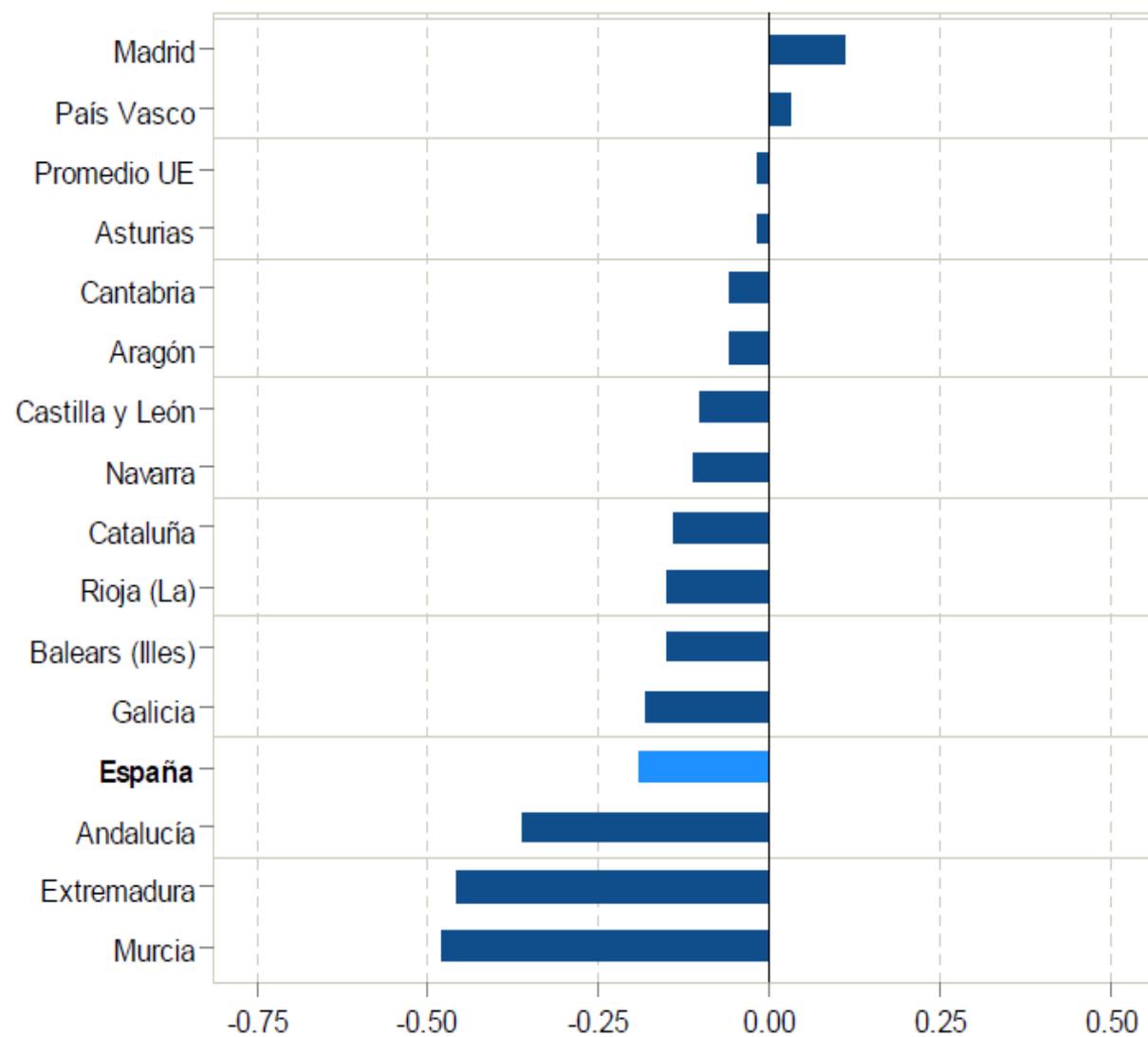
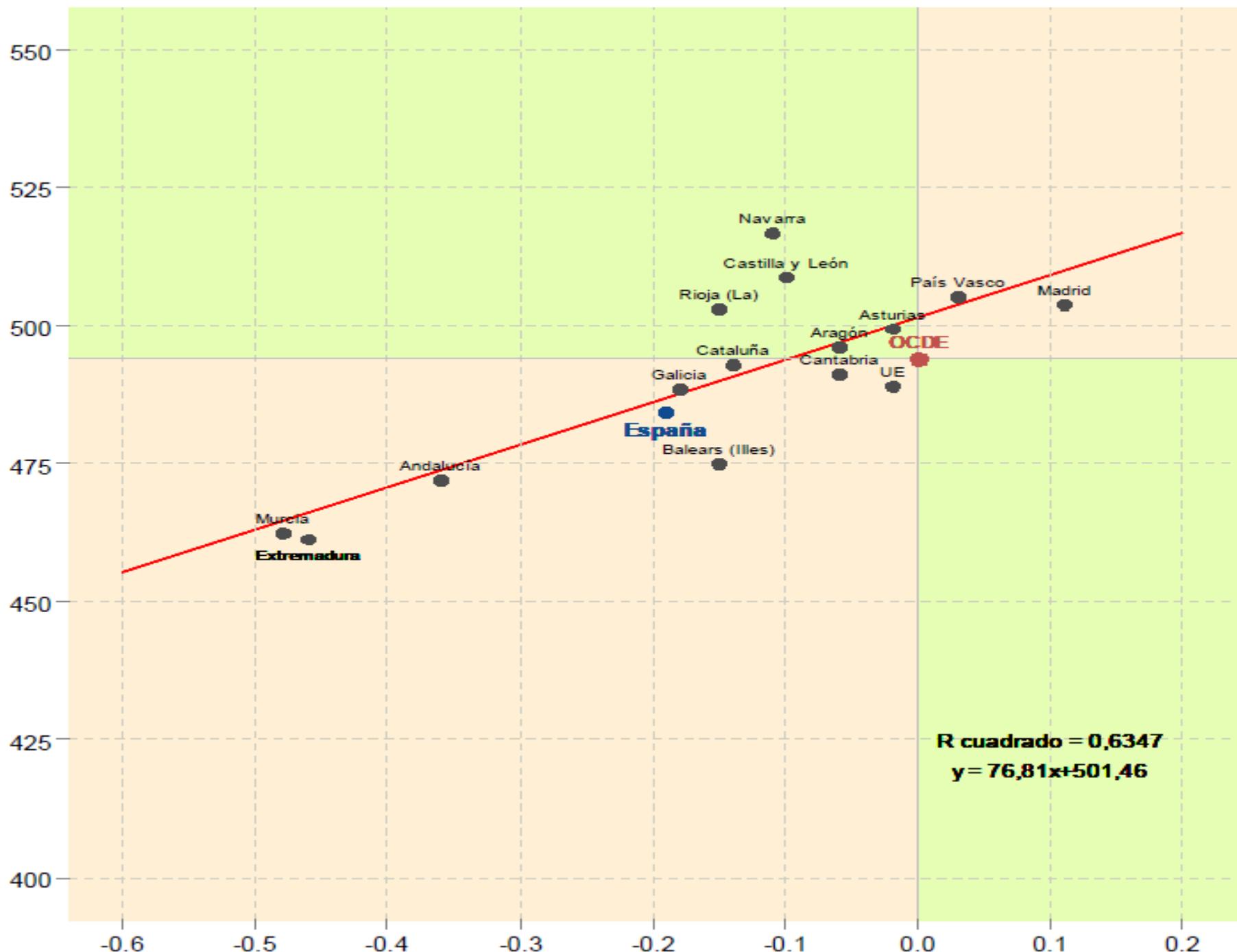
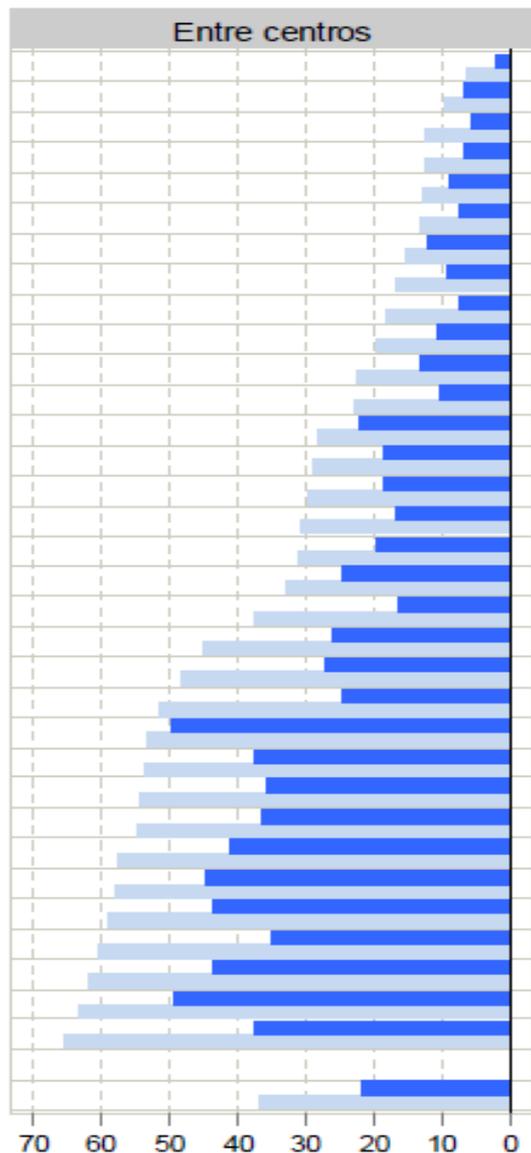


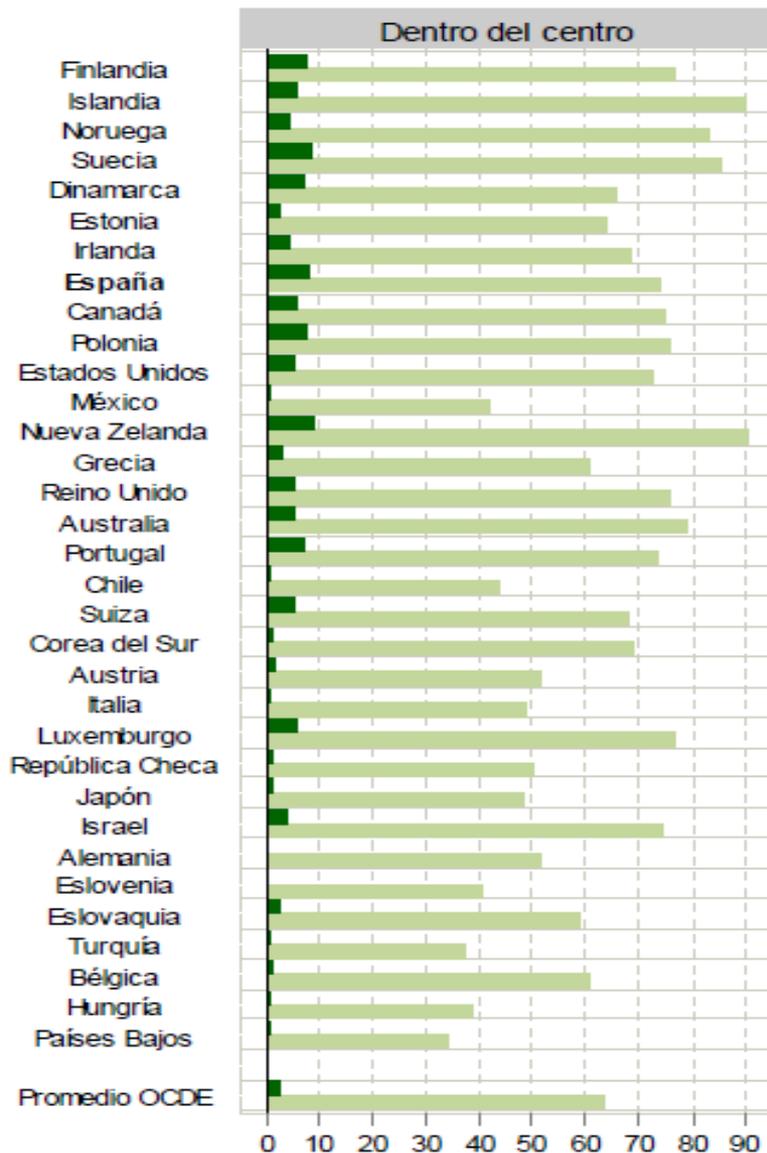
Figura 6.10. Relación entre la puntuación en matemáticas en OCDE y el PIB per cápita







■ Varianza total entre centros  
■ Varianza entre centros explicada por medio del ESCS de los alumnos y de los centros

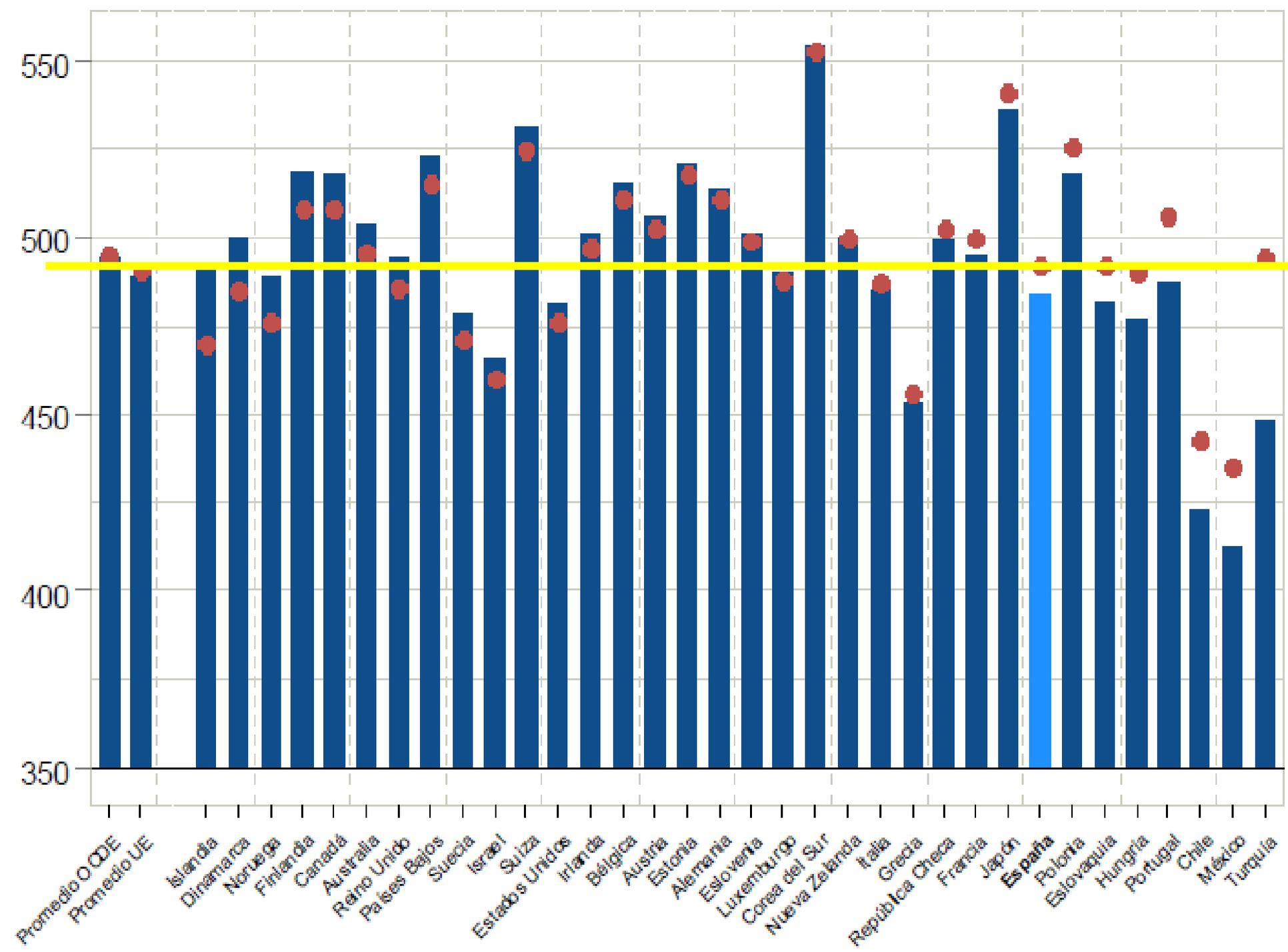


■ Varianza dentro de un centro explicada por medio del ESCS de los alumnos y de los centros  
■ Varianza total dentro del centro

---

joan subirats @subirats9

#salvados Tenemos un problema con el ejemplo de Finlandia: llegamos tarde. En 1900 en Finlandia 5% de analfabetismo, en España más del 50%



## Qué es lo que quiero evaluar

---

¿Diagnóstica una situación social dada o la capacidad de cambio (la eficacia) del sistema educativo?

# Resumen críticas Evaluaciones Ext. Est.

1. Confusión *assesment* y evaluación
2. Solo se evalúa los resultados (ni los procesos ni los elementos estructurales)
3. Resultados estandarizados que pueden no coincidir con los objetivos del proyecto educativo o las necesidades sociales.
4. Tendencia a medir los conocimientos y no otros desarrollos personales.
5. Resultados no atribuibles al sistema educativo

# Usos Evaluaciones Externas Estandarizadas

---

- Toma de decisiones:
  - Padres y madres para escoger colegio, centro
  - Dan una información a los decisores políticos
- Rendimiento de cuentas:
  - Conocimiento público, medios de información, sociedad de las puntuaciones.
- Mejora
  - Centros mejor puntuados refuerzo.
  - Presunción de que restos de centros sea un acicate para la mejora.
  - Mejora del rendimiento (en cuestión)

# ¿Relación pruebas externas y resultados?

---

Sin embargo, algunos estudios han encontrado la misma asociación positiva entre exámenes centrales y los logros de estudiantes dentro de países donde algunas regiones cuentan con sistemas de exámenes externos y otras donde no los tienen.

Ver, e.g. Bishop (1997) para provincias canadienses, Jürges et al. (2005) y Woessman (2010) para Alemania, y Bishop et al. (2001), para EE.UU.

Veinte países de la OCDE realizan a sus alumnos y alumnas pruebas de esta naturaleza y las evidencias indican que su implantación tiene un impacto de al menos dieciséis puntos de mejora de acuerdo con los criterios de PISA (Lomce Preamble VIII)

# Consecuencias I

---

- Orientar la actividad docente de todo el sistema a los ítems estandarizados que se van a medir:
  - Desvinculación con las necesidades sociales del contexto.
  - Menor atención a la diversidad individual.
  - Menos foco en el desarrollo personal y más en los conocimientos.
- A causa de lo no atribución, pueden darse juicios injustos que desemboquen en acciones como “premiar a los que más tienen y menos consiguen”
- Desconocimiento de las causas reales de por qué se producen los resultados. Pérdida del *Know how* pedagógico y de gestión de los centros.

# Consecuencias II

---

- Un *accountability* sin contextualización puede provocar una polarización entre los centros y una segregación de facto de los peor valorados.
- Con relación a la mejora:
  - La evaluación no provee de información válida para la mejora, solo otorga una posición.
  - En centros peor valorados desmotivación de los profesionales y de la comunidad educativa para la mejora.
  - Sin acuerdo y comprensión de la evaluación por los educadores no hay mejora (Huber)

## CRÍTICAS. Greene, J. (1999) The inequality of Performance Measurements.

---

- Se centran en los resultados y no en los procesos.
- La calidad de un programa no puede capturarse suficientemente con las mediciones de impacto. *Bob Stake: la calidad de una intervención es intrínsecamente compleja y no puede reducirse a un simple punto final*
- Los resultados requieren de una contextualización.
- El consenso falla cuando intenta reconstruir la diversidad de una sociedad. No se puede reducir la diversidad a un estándar universal.
- La calidad nos pertenece más a nosotros que al objeto evaluación.

20 años atrás, Lee Cronbach observó que la demanda de *accountability* es un signo de patología social en el sistema político.

Jennifer Greene: *Yo también encontré el constructo de accountability regresivo y generalmente irrelevante para mi trabajo como evaluadora (160)*

---

¿Qué es lo que se quiere  
con la evaluación?

# ¿Qué necesitas saber...para comprar ?

---



# ¿Qué necesitas saber...para arreglar?

---



# Apuntes apoyados en la experiencia

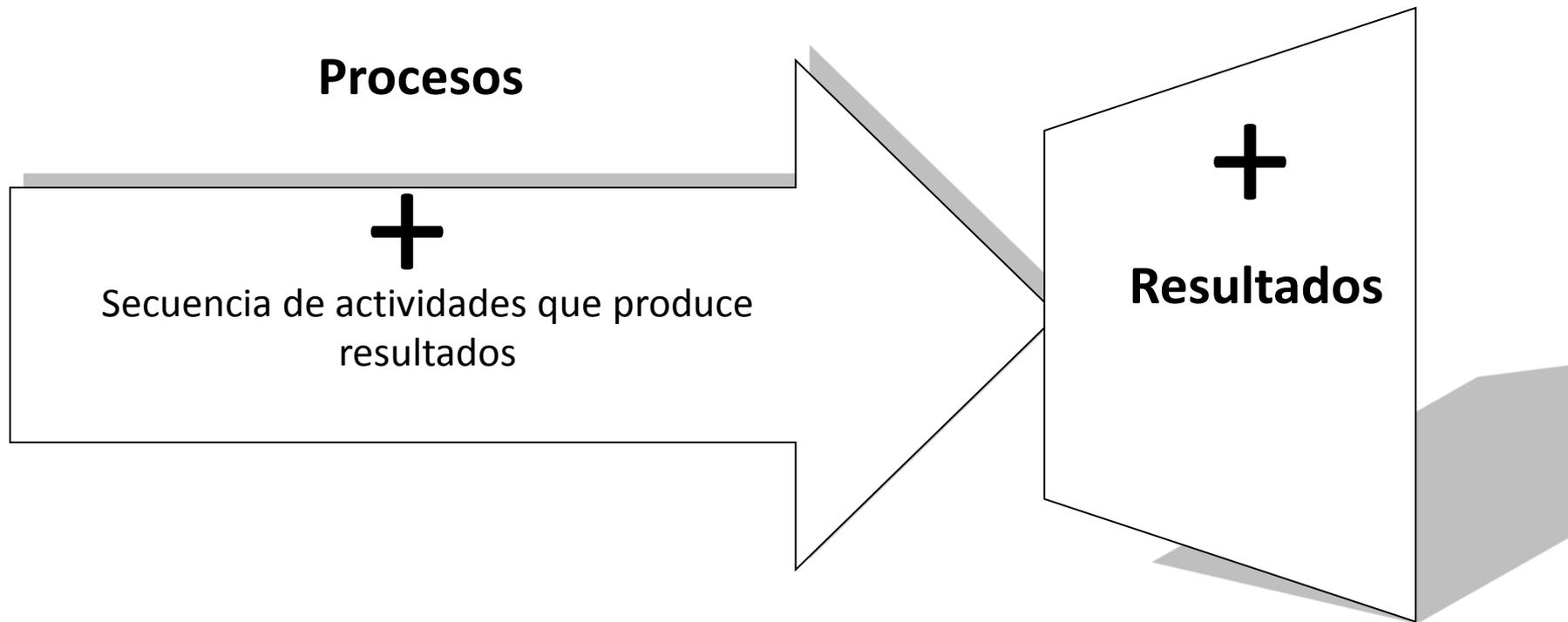
---

## Métodos facilitadores según los propósitos

Informar a los que toman decisiones	<i>Accountability</i> y	Comprensión	Mejora	Democracia, justicia social
Juicios finales, sumativos/ Orientada por la teoría	Juicios finales, sumativos/ Orientada por la teoría	Orientada por la teoría	Orientada por la teoría	Métodos bajo paradigma transformado r.

# LOS RESULTADOS ESTÁN PROVOCADOS POR LOS PROCESOS

---



# LOS PROCESOS PROVOCAN RESULTADOS.

---

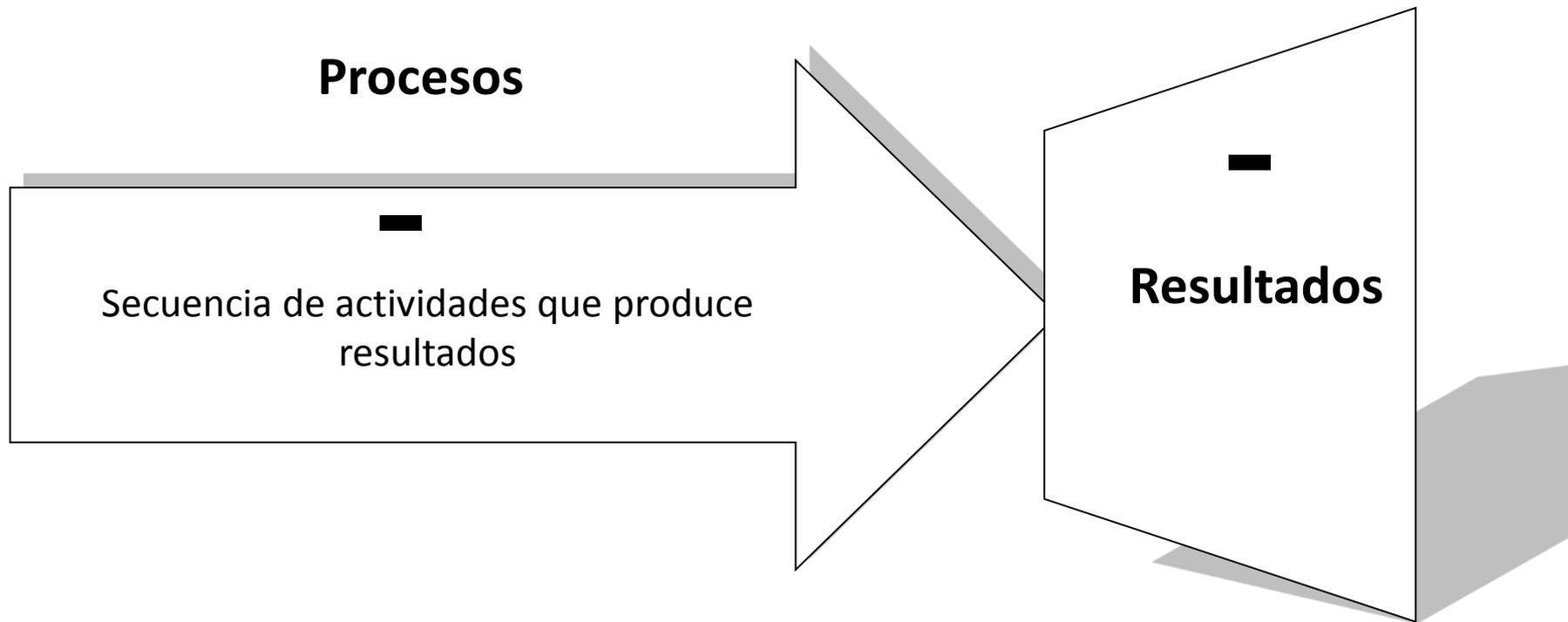


Table 2: Measures to be included in the core context assessment for PISA

(Measures in italics must be adapted to the respective major domain, e.g. science in PISA 2015)

	<b>Student and School Background</b>	<b>Processes</b>	<b>Non-cognitive Outcomes</b>
<b>System Level</b>	()	<b>Governance:</b> Decision making, horizontal and vertical differentiation	(aggregated student data)
<b>School Level</b>	School location, type and size of school, amount and source of resources (incl. ICT) Social/ethnic/academic composition  Class size, teacher qualification	<b>School policies:</b> Programmes offered, admission and grouping policies, allocated learning time, additional learning time and study support, <i>extra-curricular activities</i> , professional development, leadership, parental involvement, assessment/evaluation/accountability policies, school climate (teacher and student behaviour)  <b>Teaching and learning:</b> Disciplinary climate, teacher support, <i>cognitive challenge</i>	(aggregated student data) drop-out rate
<b>Student Level</b>	Gender, SES (parental education and occupation, home possessions, number of books at home), language and migration background, grade level, pre-primary education, age at school entry	Grade repetition, programme attended, learning time at school (mandatory lessons and additional instruction), <i>out-of-school learning</i>	Domain-general non-cognitive outcomes (e.g. achievement motivation, well-being in school); Domain-specific non-cognitive outcomes ( <i>motivation, domain-related beliefs and strategies, self-related beliefs, domain-related behaviour</i> )

## Mira lo que le dicen que mire

---

De un mismo programa se pueden evaluar muchas cosas y dejar otras sin analizar por la sencilla razón de que la realidad de las intervenciones es diversa y amplia. Es difícil, o al menos costoso, abarcar y comprender toda esa realidad en un único ejercicio de valoración. (...). La evaluación valora lo que le dicen que valore y mira lo que le dicen que mire. Las propuestas y métodos escogidos suelen ser un fiel reflejo de los valores del contexto en donde surge y se encarga la evaluación (Ligero y otros, 2014)

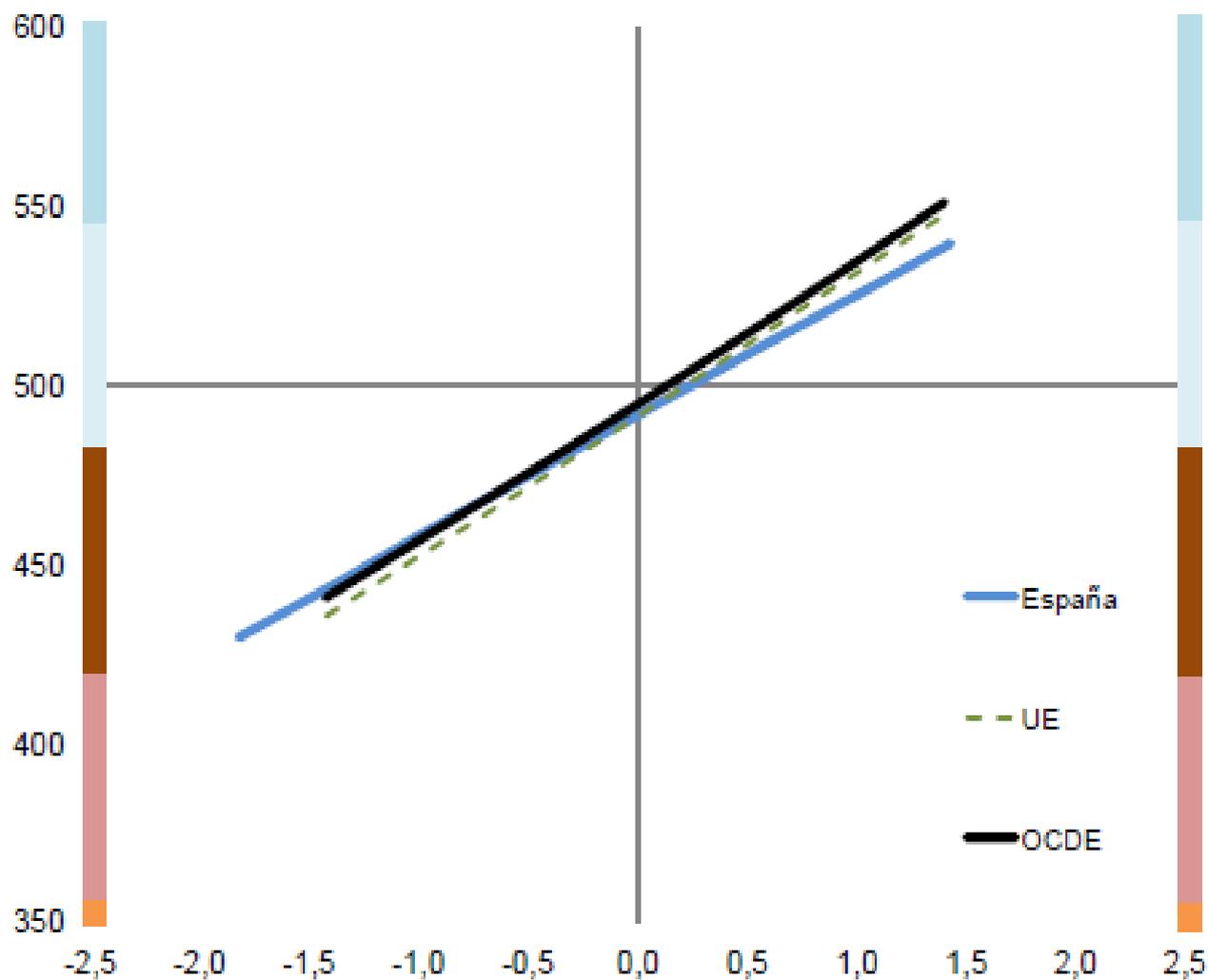
---

Muchas gracias por la atención,  
a vuestra disposición

[Jligero@polsoc.uc3m.es](mailto:Jligero@polsoc.uc3m.es)

Juan Andrés Ligeró Lasa

Figura 3.13. Líneas de regresión de los resultados de matemáticas del alumnado de España, OCDE y UE



---

Como se ve en la Figura 3.13, los alumnos españoles con ESCS bajo alcanzan puntuaciones ligeramente superiores a las puntuaciones medias de la OCDE y de la UE para su mismo nivel de ESCS. Por el contrario, en los tramos medio y alto del ESCS las puntuaciones de España son inferiores a las de la media de dichos países. Este hecho pone de manifiesto que el sistema educativo español contribuye en mayor medida a mejorar el rendimiento académico de los alumnos procedentes de entornos sociales, económicos y culturales menos favorecidos que en el conjunto de los países de la OCDE y UE. Sin embargo, los alumnos con más ventajas socioeconómicas tienen resultados más bajos que los de la OCDE y UE. (103)