

## Evaluaciones Externas Internacionales del Sistema Educativo



### Bloque 4. Análisis de resultados: conceptos estadísticos y resultados

#### Autores

David Cervera Olivares  
Paloma González Chasco  
José M<sup>a</sup> Sánchez Echave  
INEE

## Índice

Análisis de resultados: conceptos estadísticos y resultados .....	2
Descargad estos tres informes .....	2
Conceptos estadísticos.....	3
Desviación típica (o desviación estándar) .....	3
Para saber más.....	4
Error típico (o error estándar).....	4
Para saber más.....	6
Resultados.....	6
Resultados globales (PISA) .....	6
Nota .....	8
Resultados globales (EECL) .....	9
Notas.....	11
Resultados por niveles de rendimiento .....	11
Notas.....	13
Resultados por dimensiones.....	13
Ejercicio: verdadero o falso.....	<b>¡Error! Marcador no definido.</b>
Pregunta Verdadero-Falso .....	<b>¡Error! Marcador no definido.</b>

## Análisis de resultados: conceptos estadísticos y resultados

Llegados a este punto tenemos que aprender a leer e **interpretar los informes** que sobre estos estudios internacionales se elaboran. Hemos seleccionado los realizados por el INEE, pero gráficos parecidos se pueden localizar en los informes internacionales de la OCDE, IEA, UE y otros. En este bloque vamos a explicar los más usuales que se utilizan para describir los resultados obtenidos:

- rendimientos **globales**
- rendimientos por **dimensiones**
- **niveles** de rendimiento
- resultados teniendo en cuenta el **contexto social, económico y cultural de los estudiantes**
- **otros factores** sobre los que ampliar la investigación

Se recogen los últimos informes de PISA, de EECL y de TIMSS, por poner los más actuales. Habitualmente en el segundo capítulo de dichos informes se reflejan los tres primeros apartados de este bloque, y se desarrollan como ejemplo los de PISA 2012. En el tercer capítulo y posteriores se tratan los dos últimos apartados del bloque, y el ejemplo a seguir será el de PIRLS-TIMSS 2011.

Este primer conjunto de contenidos se centra en **resultados** mediante los **rendimientos globales, por dimensiones y niveles**. Pero antes de nada, conviene introducir la explicación de dos conceptos estadísticos que se utilizan muy a menudo en este tipo de estudios: **desviación estándar** y el **error estándar**.

### Descargad estos tres informes

Podéis descargar los tres informes propuestos desde cualquiera de los enlaces propuestos en cada caso:

- **PISA 2012. Programa para la Evaluación Internacional de los Alumnos. Informe español. Volumen I: Resultados y contexto.**

<http://www.mecd.gob.es/dctm/inee/internacional/pisa2012/pisa2012lineavolumeni.pdf?documentId=0901e72b81786310>

[PISA 2012](#)

Descargad también las [tablas con los datos](#) de cada capítulo.

- **PIRLS - TIMSS 2011. Estudio Internacional de progreso en comprensión lectora, matemáticas y ciencias. IEA. Volumen I: Informe español**

<http://www.mecd.gob.es/dctm/inee/internacional/pirlstimss2011vol1-1.pdf?documentId=0901e72b81710232>

## [PIRLS TIMSS 2011](#)

- **Estudio Europeo de Competencia Lingüística EECL. Volumen I. Informe español**

[http://www.mecd.gob.es/inee/publicaciones/estudios-internacionales.html#EECL\\_I](http://www.mecd.gob.es/inee/publicaciones/estudios-internacionales.html#EECL_I)

[EECL](#)

## Conceptos estadísticos

Tanto la desviación estándar como el error estándar son dos conceptos estadísticos que es importante entender para poder leer y comprender adecuadamente los informes de evaluaciones externas internacionales.

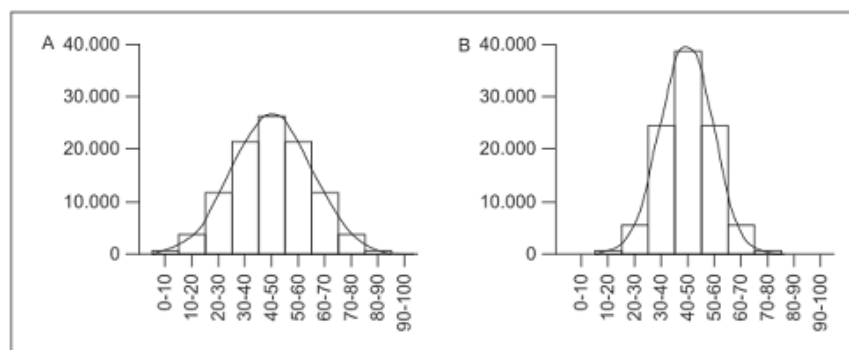
Hemos intentado huir de complicadas fórmulas matemáticas y por ello los vamos a explicar mediante ejemplos sencillos.

## Desviación típica (o desviación estándar)

La **desviación estándar** (en inglés “standard deviation”; SD) es una **medida de la dispersión de los datos**, cuanto mayor sea la dispersión mayor es la desviación estándar, si no hubiera ninguna variación en los datos, es decir, si fueran todos iguales, la desviación estándar sería cero. La desviación estándar cuantifica la dispersión alrededor de la media aritmética. Informa de la media de distancias que tienen los datos respecto de su media aritmética.

Supónganse dos **poblaciones** de 100.000 individuos en las que la edad se distribuye como se muestra en los histogramas A y B de la figura 1 (se tienen datos del 100% de los individuos en este ejemplo).

- La **edad media** (media aritmética) en ambas poblaciones es **44,5**.
- La **desviación estándar** es **14,9 en la población A y 10,0 en la B**.



**Figura 1.** Histogramas de frecuencias de la edad en unas poblaciones hipotéticas de 100.000 individuos. En ambas poblaciones la media es 44,5; las desviaciones estándar son 14,9 en A y 10,0 en B. Las líneas continuas muestran la distribuciones teóricas correspondientes a curvas gaussianas con las mismas medias y desviaciones estándar.

Si no ve correctamente la imagen, puede descargarse este [PDF](#).

En la población B las edades de los individuos están menos dispersas alrededor de la media, como corresponde a una desviación estándar menor.

Unas fórmulas muy difundidas establecen que en el intervalo comprendido entre la media menos UNA desviación estándar y la media más UNA desviación estándar están aproximadamente el **68% central de los datos**.

Además, en el intervalo comprendido entre la media menos 1,96 (aproximadamente 2) veces la desviación estándar y la media más 1,96 veces la desviación estándar están aproximadamente el **95% central de los datos**.

Según estas reglas, en la **población A** del ejemplo:

- aproximadamente los **68.000 individuos** (68%) con edades más cercanas a la media tendrían **edades comprendidas entre 29,6** ( $44,5-14,9$ ) **y 59,4** ( $44,5+14,9$ ) (realmente en ese intervalo hay 67.961 individuos).
- **95.000 individuos** (95%) estarían **entre 15,3** ( $44,5-1,96*14,9$ ) **y 73,7** años ( $44,5+1,96*14,9$ ) (realmente hay 95.006).
- de los 5.000 restantes (5%), la mitad tendrían menos de 15,3 (en los datos hay 2.528) y la otra mitad más de 73,7.

Estos intervalos para la **población B** son:

- entre 34,5 ( $44,5-10$ ) y 54,5 ( $44,5+10$ ) los 68.000 individuos (68%) con edades más cercanas a la media.
- entre 24,9 ( $44,5-1,96*10$ ) y 64,1 ( $44,5+1,96*10$ ) los 95.000 centrales (95%).

Hay que resaltar que esta regla solo es cierta si la variable se ajusta a una distribución normal o gaussiana; en el ejemplo la regla se cumple porque los datos ajustan perfectamente a la distribución normal como se observa en la figura.

Ejemplos extraídos de: <http://www.elsevier.es/es/revistas/semergen-medicina-general--familia-40/desviacion-estandar-error-estandar-13041428-notas-estadisticas-2002>

## Para saber más

Desviación estándar: [http://es.wikipedia.org/wiki/Desviaci%C3%B3n\\_est%C3%A1ndar](http://es.wikipedia.org/wiki/Desviaci%C3%B3n_est%C3%A1ndar)

## Error típico (o error estándar)

Las investigaciones rara vez se hacen sobre el conjunto de la población; lo habitual es realizarlas en un subconjunto (**muestra**) de ella. Esta práctica está justificada porque la teoría estadística establece que, si la muestra se selecciona aleatoriamente, sus características

(forma de la distribución, media, desviación estándar, etc.) son parecidas a las de la población y tanto más parecidas cuanto mayor sea la muestra.

En la figura 2 se representa el histograma de la **edad de una muestra aleatoria de 100 individuos** extraída de la población representada en la figura 1 A (que eran 100.000 individuos). Obsérvese que es parecido al de la población aunque, por ejemplo, su ajuste a la curva normal es peor.

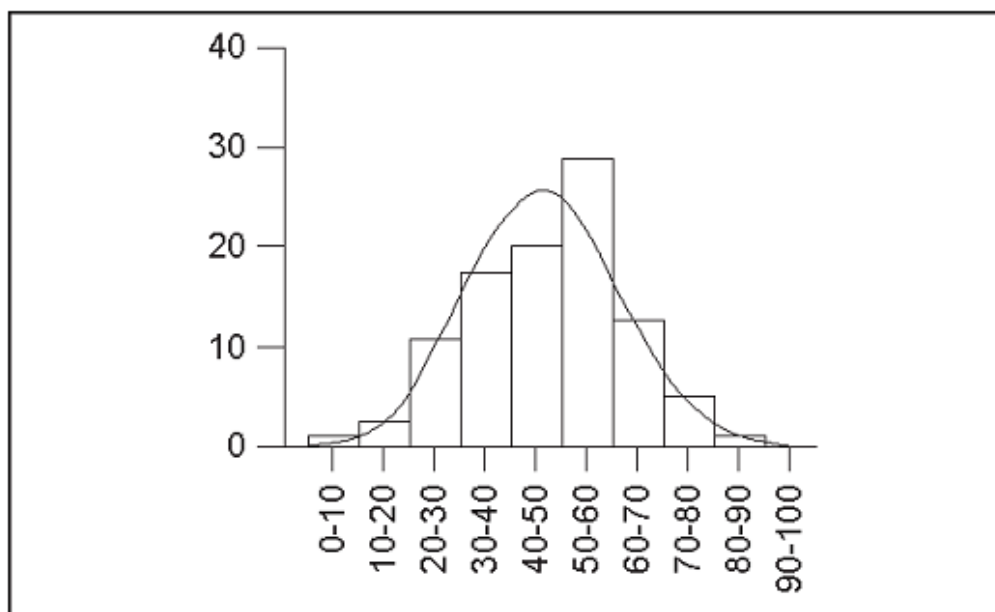


Figura 2. Histograma de frecuencias de la edad en una muestra aleatoria de 100 individuos de la población representada en la figura 1.A. En la muestra la media es 46,2 y la desviación estándar 14,7.

Si no ve correctamente la imagen, puede descargarse este [PDF](#).

En esta **muestra**, la media y la desviación estándar son **46,2** y **14,7** respectivamente, también parecidas a las de la población. Si se tomara otra muestra aleatoria se obtendrían otros valores distintos, aunque probablemente también **parecidos a los de la población** (que son 44,5 de media y 14,9 desviación estándar).

El **error estándar** de la media (en inglés “standard error of the mean”; SEM o SE) es el valor que cuantifica cuánto se apartan los valores de la media de la población. Es decir, el error estándar de la media **cuantifica las oscilaciones de la media muestral** (media obtenida en base a los datos medidos en la muestra utilizada) **alrededor de la media poblacional** (verdadero valor de la media). Es una medida del error que se comete al tomar la media calculada en una muestra como estimación de la media de la población total.

A partir del error estándar se construye el **intervalo de confianza** de la medida correspondiente.

1) El **error estándar** de la media estimado en la muestra del ejemplo es **1,47**. Se calcula dividiendo la desviación estándar por la raíz cuadrada del tamaño muestral  $14,7/\sqrt{100}=14,7/10$

2) Calculado a partir de él, el **intervalo de confianza al 95%** para la media va desde **43,3** a **49,1**

- Límite inferior = media - 1,96 veces el error estándar =  $46,2 - 1,96 * 1,47 = 43,3$  (límite inferior)
- Límite superior = media + 1,96 veces el error estándar =  $46,2 + 1,96 * 1,47 = 49,1$  (límite superior)

Este es uno de los métodos estadísticos que exige normalidad de la población. Quiere decir que podemos afirmar, **con una confianza del 95%**, que **la media poblacional está incluida en dicho intervalo**. Compárese con el valor de la media poblacional (verdadero valor de la media) que, en este ejemplo y en contra de lo que ocurre en las investigaciones reales, es conocido. Vemos que el valor verdadero (44,5 años) está dentro del intervalo de confianza que habíamos calculado (43,3 – 49,1). Por tanto, hablar de una confianza del 95% significa que en el 95% de las posibles muestras que podríamos tomar de la población total, la edad media (44,5) estaría contenida en el intervalo construido.

Ejemplos extraídos de: <http://www.elsevier.es/es/revistas/semergen-medicina-general--familia-40/desviacion-estandar-error-estandar-13041428-notas-estadisticas-2002>

## Para saber más

Error estándar: [http://es.wikipedia.org/wiki/Error\\_est%C3%A1ndar](http://es.wikipedia.org/wiki/Error_est%C3%A1ndar)

## Resultados



En este apartado se van a explicar diferentes resultados por países que se obtienen de los análisis realizados por PISA sobre el rendimiento del alumnado. La combinación de estos resultados ofrece una muy valiosa información sobre el funcionamiento de los sistemas educativos.

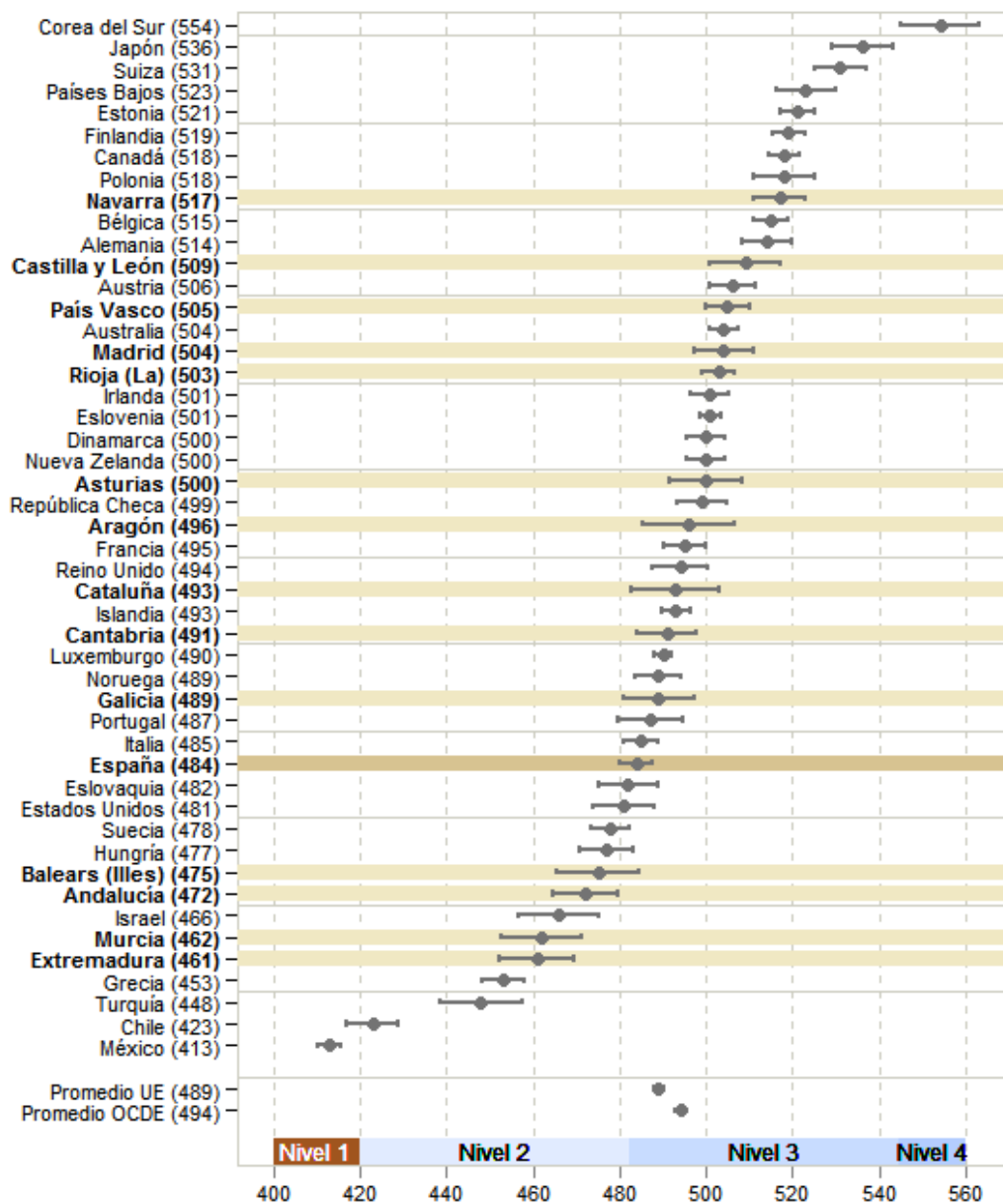
**Se han tomado datos del informe español PISA 2012.**

## Resultados globales (PISA)

Los resultados de **PISA** se presentan por medio de escalas con una **puntuación media de 500** y una **desviación típica de 100**, lo que significa que el 68% de los alumnos de los países

de la OCDE obtuvieron entre 400 y 600 puntos. Estas puntuaciones representan distintos grados de competencia en el área de conocimiento.

En el año 2012 la competencia principal fue matemáticas, dentro del marco establecido en PISA 2003 (primer año en que las matemáticas fueron área principal de evaluación) que se ha tomado como referencia para los resultados posteriores y para valorar su evolución, y en el que se fijaba que los resultados se refieren a una desviación típica de 100 puntos. Los **resultados globales** se presentan gráficamente mediante las **puntuaciones medias obtenidas por los alumnos de los distintos países** de la muestra seleccionada. Estos resultados se refieren a una escala continua en la que se hace equivaler a 500 puntos el promedio de las puntuaciones medias obtenidas por los alumnos de los países OCDE en PISA 2003 (área principal matemáticas) y en la que la desviación típica es 100 puntos.

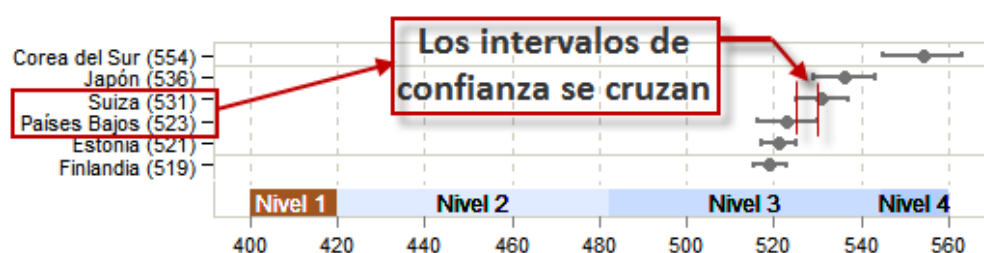




En este promedio, los resultados de todos los países OCDE han sido ponderados por igual como si aportaran todos ellos el mismo número de alumnos. Este **Promedio OCDE** que tenemos marcado en la figura anterior es, por tanto, la **media aritmética de las puntuaciones medias de los países OCDE**. Se utiliza el Promedio OCDE para comparar el rendimiento medio de los alumnos de un país con el rendimiento medio de los países de la OCDE. El Promedio UE tiene el mismo significado, pero teniendo en cuenta a los 27 países miembros de la Unión Europea (todos excepto Malta).

Los **resultados promedio de cada país** se representan junto con los correspondientes **intervalos de confianza del 95%**, estimados a partir de sus errores típicos, en los que se puede afirmar, con una confianza del 95%, que se encuentra su promedio verdadero. Estos datos deben interpretarse del modo siguiente: hay una probabilidad del 95% de que el verdadero valor promedio de un país esté en el intervalo de confianza señalado en la figura.

Esto significa que entre todos los países con puntuaciones medias próximas (con diferencias tales que sus intervalos de confianza coinciden en parte) nada se puede decir sobre el “orden” que unos tienen con respecto a otros. Por ejemplo, **no hay diferencias significativas** entre Suiza con 531 puntos y Países Bajos con 523 puntos. Dicho de otro modo, nada se puede afirmar sobre cuál de estos dos países tiene un “resultado” mejor que el otro, pues sus **intervalos de confianza comparten un tramo común** (observa la siguiente figura).



**España** obtuvo una puntuación **media de 484** puntos, siendo el intervalo de confianza desde 480,60 a 488,04 puntos, extremos entre los que se encuentra el resultado real español con una confianza del 95%.

En la presentación gráfica de los resultados promedio de rendimiento se ha optado por resaltar en la parte inferior del gráfico unas franjas delimitadas por los valores correspondientes a los extremos inferior y superior de los **niveles de rendimiento** (se explican en el siguiente apartado). Todos los países situados en estas franjas tienen resultados con diferencias inferiores a un nivel. Por ejemplo, encontramos a la mayoría de países de la OCDE y los europeos participantes en el nivel 3 de rendimiento.

### Nota

La OCDE ha constatado que **un año académico** más incrementa **10 puntos** de media el rendimiento del alumno.

## Resultados globales (EECL)

Acabamos de ver cómo se representan los resultados globales en PISA. En otros estudios como PIRLS o TIMSS se hace de forma muy parecida. Sin embargo, en el **Estudio Europeo de Competencia Lingüística (EECL)** no es posible asignar puntuaciones medias, dado que no existe una población objetivo uniforme. Entre otras discrepancias podemos observar las siguientes:

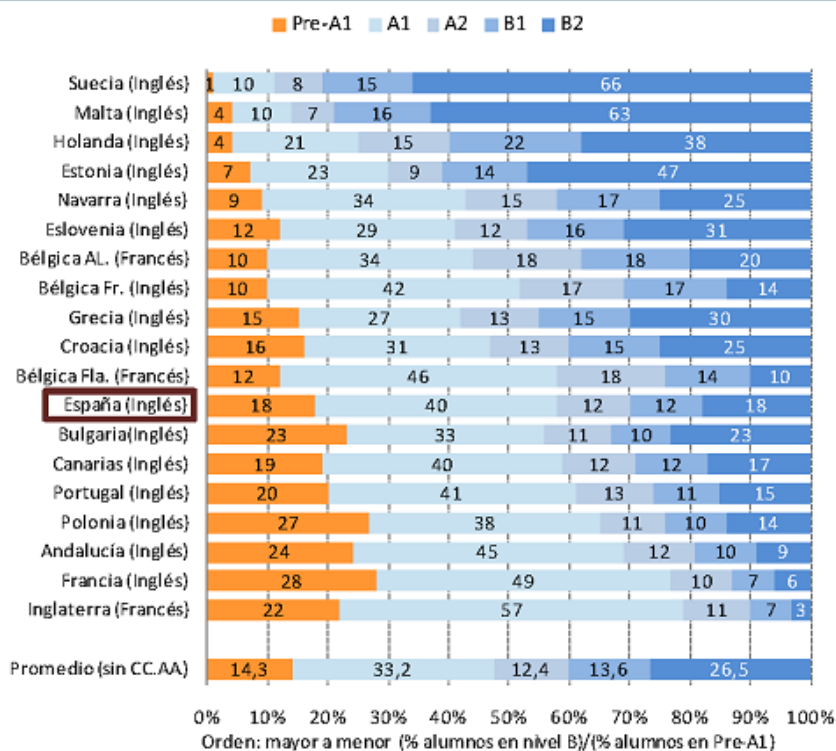
- cada entidad eligió dos lenguas diferentes con dos muestras de alumnos distintas;
- el curso en el que se realizó la prueba no fue el mismo en todas las entidades;
- los alumnos tenían distintas edades;
- existen diferencias en cuanto a la obligatoriedad, edad de comienzo de la enseñanza de las lenguas extranjeras.

Por ello los **resultados** de este estudio son estimaciones de la **distribución de los alumnos** de cada población **según** los niveles descritos por el **MCERL**.

La primera lengua evaluada por la mayoría de las entidades fue el Inglés. Y las comunidades alemana y flamenca de Bélgica e Inglaterra evaluaron el Francés. Resulta complejo establecer un **criterio de comparación** de los resultados y por lo tanto de ordenación de las entidades. Lo deseable es tener el mayor porcentaje de alumnos en el nivel B (suma de B1 y B2) y el menor porcentaje posible en el nivel Pre-A1, ya que estos estudiantes son aquellos que no alcanzan un mínimo nivel de competencia como el descrito en el nivel A1. Por todo ello, el criterio empleado en estos gráficos para ordenar las entidades es el cociente entre los porcentajes de alumnos en los niveles B y los situados en Pre-A1, interpretándose esta razón como el número de alumnos B por cada Pre-A1. Una entidad con mayor valor para ese cociente tiene, según este criterio, mejores resultados.

En el siguiente gráfico se muestra la **distribución por niveles** de los alumnos en las distintas entidades y comunidades autónomas que ampliaron muestra, en concreto para la primera lengua evaluada (inglés en el caso de España) y para la destreza llamada comprensión lectora. También se muestran en la parte inferior del gráfico los porcentajes globales medios para poder establecer comparaciones.

**Gráfico 3.2b. Distribución por niveles del MCERL en la primera lengua evaluada. Comprensión lectora**



Si no ves correctamente la imagen anterior, puedes descargar este [PDF](#).

Algunas de las conclusiones que se extrajeron en el informe español son las siguientes:

- Malta, Suecia, Holanda, y Estonia consiguen un elevado nivel en el dominio de la primera lengua, con más de un 50% de alumnos en los niveles B y Eslovenia, muy cercana a este porcentaje. Dejando aparte Malta, cuyo pasado como colonia británica explica claramente estos resultados, los otros cuatro países tienen ciertas características comunes como son tener poblaciones pequeñas e idiomas minoritarios, lo que hace que sea muy importante para su población poder comunicarse en otra lengua diferente a la materna para cualquier tipo de relación exterior.
- El porcentaje de estudiantes en Pre-A1 y A1 es superior o cercano al 50% en España, Francia, Inglaterra, Polonia, Portugal y Bulgaria. Se observa que la lengua materna en las entidades que evaluaron Inglés es una lengua romance o eslava, todas ellas con gran distancia lingüística respecto al Inglés.
- España obtiene sus mejores resultados relativos en comprensión lectora y los peores en comprensión oral (descarga este [PDF](#) si quieres ver los resultados en comprensión oral). Diferentes razones pueden explicar estos resultados; entre otras, la metodología de enseñanza empleada durante años, que ha insistido en contenidos gramaticales y de comprensión lectora y escrita, la escasa exposición al uso de esta lengua en el contexto social español.

## Notas

- En el EECL, cuando se habla de entidades se hace referencia a los distintos países o comunidades lingüísticas participantes en el estudio. Por ejemplo Bélgica participó con sus tres comunidades lingüísticas de forma independiente (alemana, flamenca y francesa).
- En España ampliaron muestra Andalucía, Canarias y Comunidad Foral de Navarra.

## Resultados por niveles de rendimiento

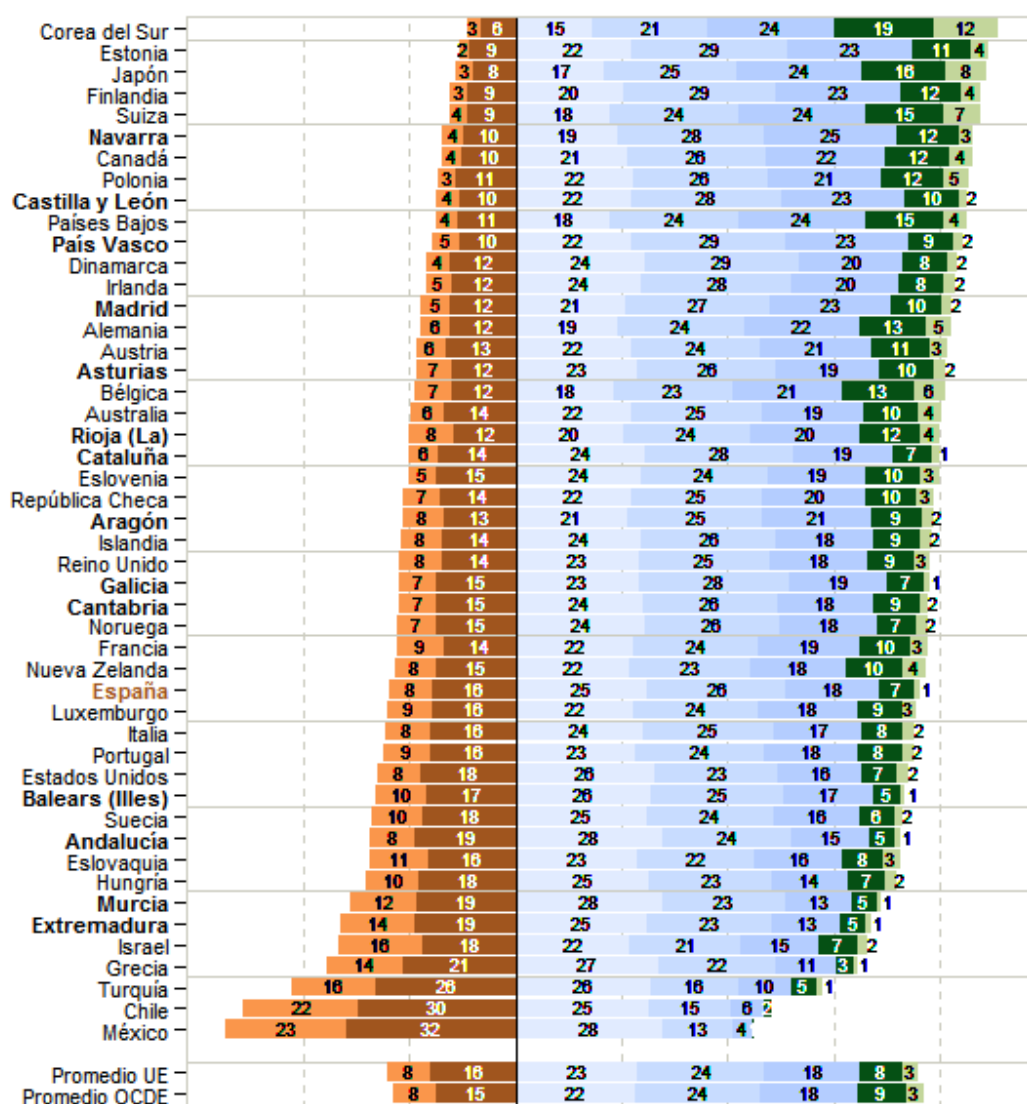
En PISA 2012 se establecieron distintos **niveles de rendimiento**, que se corresponden a distintos grados de dificultad en las tareas requeridas, en cada una de las tres competencias básicas evaluadas. Estos niveles se describen de acuerdo con los ítems que se encuentran en cada uno de ellos. Las puntuaciones de los participantes se organizaron en los niveles de rendimiento establecidos para cada competencia.

Para matemáticas, podemos ver descargando este [PDF](#) una tabla donde se presenta un resumen con la **descripción** de lo que los **alumnos son capaces** de realizar en cada uno de los **niveles** establecidos (no es necesario leerla de forma exhaustiva). Esta descripción ha sido elaborada por los especialistas correspondientes.

La utilidad de la definición y valoración de los niveles de rendimiento es doble:

- Permite cuantificar el **porcentaje de alumnos** que se encuentra **en cada nivel en cada** uno de los **países**.
- La correspondencia entre las preguntas planteadas y el nivel resultante permite a los expertos describir qué caracteriza el grado de adquisición de la competencia por parte de los alumnos que en él se encuentran, es decir, **qué saben hacer, cómo aplican** sus conocimientos y con **qué grado de complejidad**.

En la figura vemos, para la competencia matemática, los **porcentajes de alumnos de cada país cuyo resultado se encuentra dentro del intervalo de puntuaciones de cada nivel de rendimiento**. A los niveles establecidos se añadió un nivel inferior al primero (nivel menor que 1), que es aquel en el que se encuentran los alumnos cuyo rendimiento es tan bajo que PISA no es capaz de describir.



■ Nivel <1 
 ■ Nivel 1 
 ■ Nivel 2 
 ■ Nivel 3 
 ■ Nivel 4 
 ■ Nivel 5 
 ■ Nivel 6

El **nivel de rendimiento 1 y <1** corresponde al grado más modesto de adquisición de la competencia. PISA considera que los alumnos que se encuentran en este nivel tienen riesgo de no poder afrontar con suficiente garantía de éxito sus retos formativos, laborales y ciudadanos posteriores a la educación obligatoria. España e Italia obtuvieron un 24% de alumnos en esos niveles, cifra igual a la UE y similar al promedio OCDE (23%).

Como señala el informe internacional, puede considerarse que los estudiantes que se encuentran en el **nivel de rendimiento 2** en matemáticas demuestran el tipo de competencia mínimo requerido para el aprendizaje posterior y la vida social y laboral.

En España, el porcentaje de alumnos en los **niveles 5 y 6**, que corresponden a los rendimientos más elevados, es del 8%, frente al 12% del Promedio OCDE.

## Notas

Observa que, según las puntuaciones marcadas para la separación entre los niveles de rendimiento en matemáticas, cada nivel supone 62,3 puntos:

- 2 - 1:  $420,1 - 357,7 = 62,4$  puntos.
- 3 - 2:  $482,4 - 420,1 = 62,3$  puntos.
- 4 - 3:  $544,7 - 482,4 = 62,3$  puntos.
- 5 - 4:  $607 - 544,7 = 62,3$  puntos.
- 6 - 5:  $669,3 - 607 = 62,3$  puntos.

Recordemos que el alumno que está, por ejemplo, en el nivel 3 es capaz de desarrollar las tareas también de los niveles 2 y 1.

## Resultados por dimensiones

PISA 2012 ofrece una valoración del rendimiento de los alumnos en las **cuatro sub-áreas de contenido** matemático (Cantidad; Cambio y relaciones; Incertidumbre y datos; Espacio y forma) y en relación con los **tres procesos** matemáticos (Formular, Emplear e Interpretar).

Los resultados permiten estimar las **diferencias entre el promedio global** en matemáticas y el **obtenido para cada sub-área o proceso** valorados.

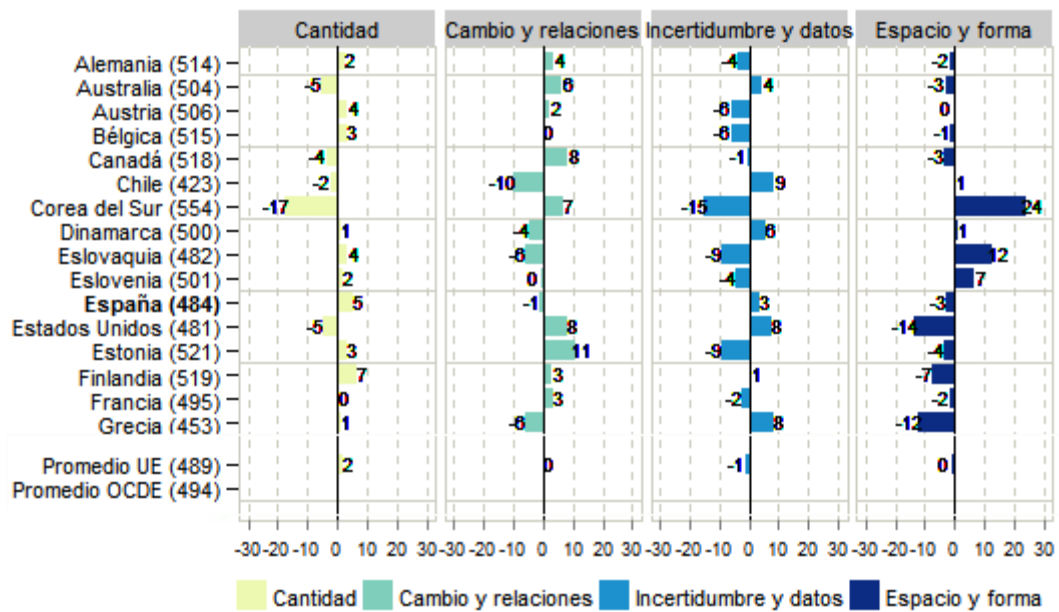
Tomando como referencia el promedio general y el de cada sub-área o proceso en la OCDE, las siguientes figuras muestran la diferencia entre dos distancias  $d_1$  y  $d_2$  (en el gráfico se representa  $d = d_2 - d_1$ ):

- **$d_1$** : Puntuación media del país menos promedio de la OCDE.
- **$d_2$** : Puntuación media en la sub-área o proceso del país menos promedio de la OCDE en dicha sub-área o proceso.

Un resultado positivo en una sub-área o proceso significa que el alumnado tiene mejor rendimiento relativo en el mismo que en el conjunto de las matemáticas.

Si tomamos, por ejemplo, para España la sub-área **Cantidad**, el dato representado (5) está calculado de la siguiente manera:

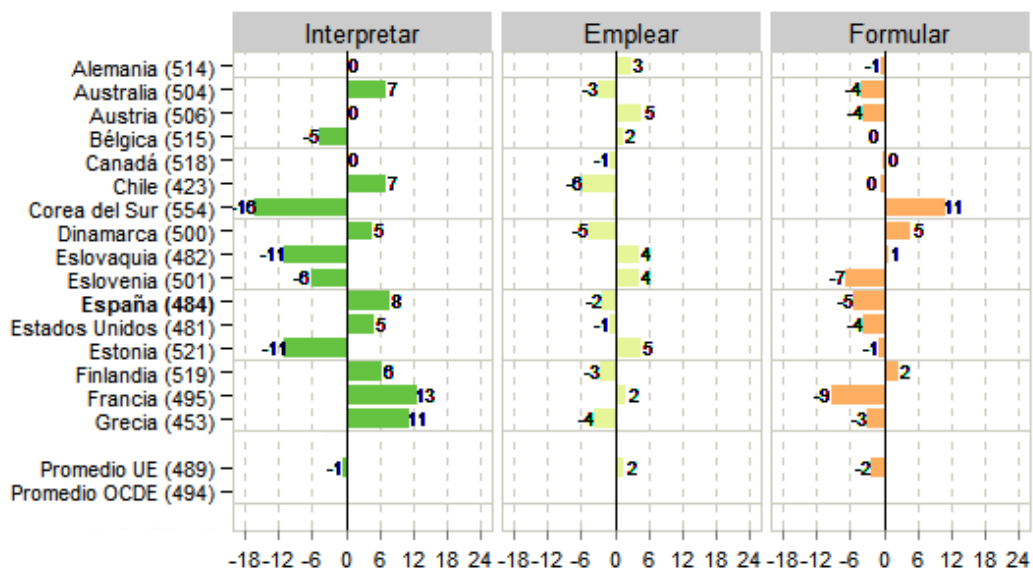
$d = (490,84 - 495,15) - (484,32 - 494,05) = (-4,30) - (-9,73) = 5,42$  que, redondeado, se representa como un 5 en la gráfica.



España presenta mejor rendimiento relativo en las sub-áreas de cantidad e incertidumbre y datos que en cambio y relaciones y espacio y forma. Por tanto, el rendimiento relativo es peor en las dos últimas que en las dos primeras.

En España, el proceso **Interpretar** tiene un resultado positivo (8) y está calculado como sigue:

$d = (495,08 - 497,01) - (484,32 - 494,05) = (-1,93) - (-9,73) = 7,8$  que, redondeado, se representa con un 8 en la gráfica.



En general, los alumnos de 15 años españoles tienen más dificultades a la hora de formular los conceptos matemáticos que de interpretar los mismos. Es decir, les resulta más complicado identificar oportunidades para aplicar (formular) las matemáticas que reflexionar sobre las soluciones o los resultados matemáticos e

interpretarlos en el contexto de un problema o desafío. En las tareas relacionadas con el empleo de conceptos, datos, procedimientos y razonamientos matemáticos no se observa un patrón definido en cuanto al rendimiento de los alumnos.

---



Formación en Red del INTEF

Licencia [Creative Commons Atribución-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-sa/4.0/).