

Capítulo 2

Resultados en matemáticas y ciencias

TIMSS 2023

Informe español

RESULTADOS EN MATEMÁTICAS Y CIENCIAS



MATEMÁTICAS

Rendimiento global



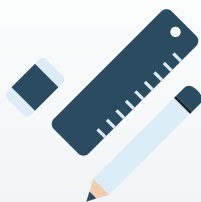
498



525



514



Rendimiento por dominios de contenido y cognitivos

Números

497

Medidas y geometría

497

Datos

502

Conocimiento

500

Aplicación

497

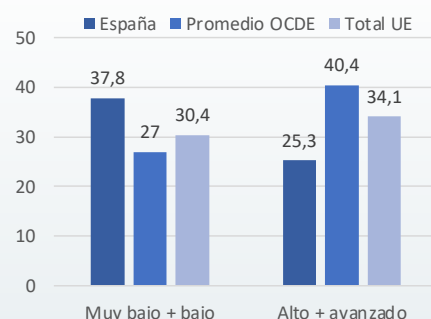
Razonamiento

500



Niveles de rendimiento

Porcentaje de estudiantes por nivel de rendimiento



CIENCIAS

Rendimiento global



504



526



513



Rendimiento por dominios de contenido y cognitivos

Ciencias de la vida

502

Ciencias físicas

505

Ciencias de la Tierra

505

Conocimiento

508

Aplicación

502

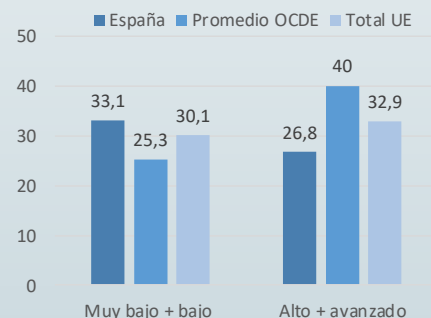
Razonamiento

498



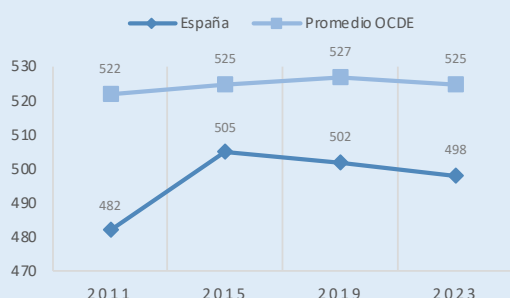
Niveles de rendimiento

Porcentaje de estudiantes por nivel de rendimiento

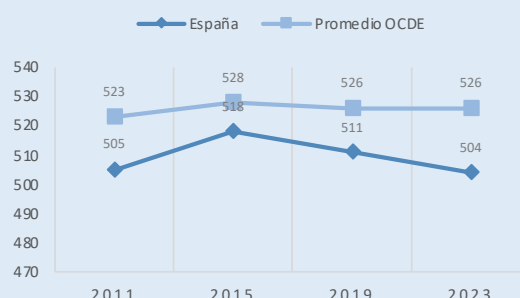


EVOLUCIÓN DEL RENDIMIENTO ENTRE CICLOS

MATEMÁTICAS



CIENCIAS



Capítulo 2

Resultados en matemáticas y ciencias

2.1. Introducción

Este capítulo presenta los resultados en el estudio TIMSS 2023, tanto en las competencias habituales del estudio (matemáticas y ciencias) como en el área de conciencia medioambiental. Como se mostró en el capítulo anterior, en TIMSS 2023 participaron 59 países. No obstante, en el presente informe se han seleccionado aquellos donde la comparación con los datos españoles es más relevante, países miembros de la OCDE y/o de la UE. De este modo en el desglose de los resultados del presente capítulo se considerarán las siguientes poblaciones:

- En primer lugar, dos **parámetros internacionales**. Uno de ellos, denominado **Promedio OCDE**, se expresa como la media aritmética de las puntuaciones de los estados miembros de la Organización para la Cooperación y el Desarrollo Económico (OCDE) participantes en TIMSS 2023. El segundo parámetro internacional, **Total UE**, se calculó como la media ponderada de los países de la Unión Europea (UE) según el tamaño de la población objetivo de cada socio comunitario. De este modo, los países que escolarizan a más estudiantes de 4.º curso contribuyen a los resultados del Total UE en mayor proporción que los países con menor población.
- Por otro lado, todas las comparativas incluirán los resultados de los **países** que cumplan al menos una de las siguientes condiciones: ser **miembros de la OCDE y de la UE**, organismos a los que España pertenece desde 1961 y 1986, respectivamente.
- Finalmente, en la presentación de los datos españoles, además del promedio de **España**, se incluirán los resultados de las **nueve comunidades autónomas que ampliaron su muestra** de estudiantes y que, por tanto, disponen de datos representativos y comparables: Andalucía, Principado de Asturias, Illes Balears, Canarias, Castilla y León, Cataluña, Galicia, Comunidad de Madrid y Comunidad Foral de Navarra.

2.1.1. ¿Cómo se expresan los resultados en TIMSS?

Una de las finalidades primarias de las evaluaciones educativas a gran escala es conocer y describir los conocimientos, destrezas y competencias del alumnado. TIMSS, como otras evaluaciones comparadas de sistemas educativos, expresa los logros del alumnado de dos maneras. Por un lado, como escalas de puntuaciones numéricas, del original en inglés *score scales* o *performance scales*. Por otro lado, los resultados también se expresan como niveles de rendimiento, del original en inglés, *proficiency scales* o *achievement levels*.

2.1.1.1. Resultados como puntuaciones numéricas

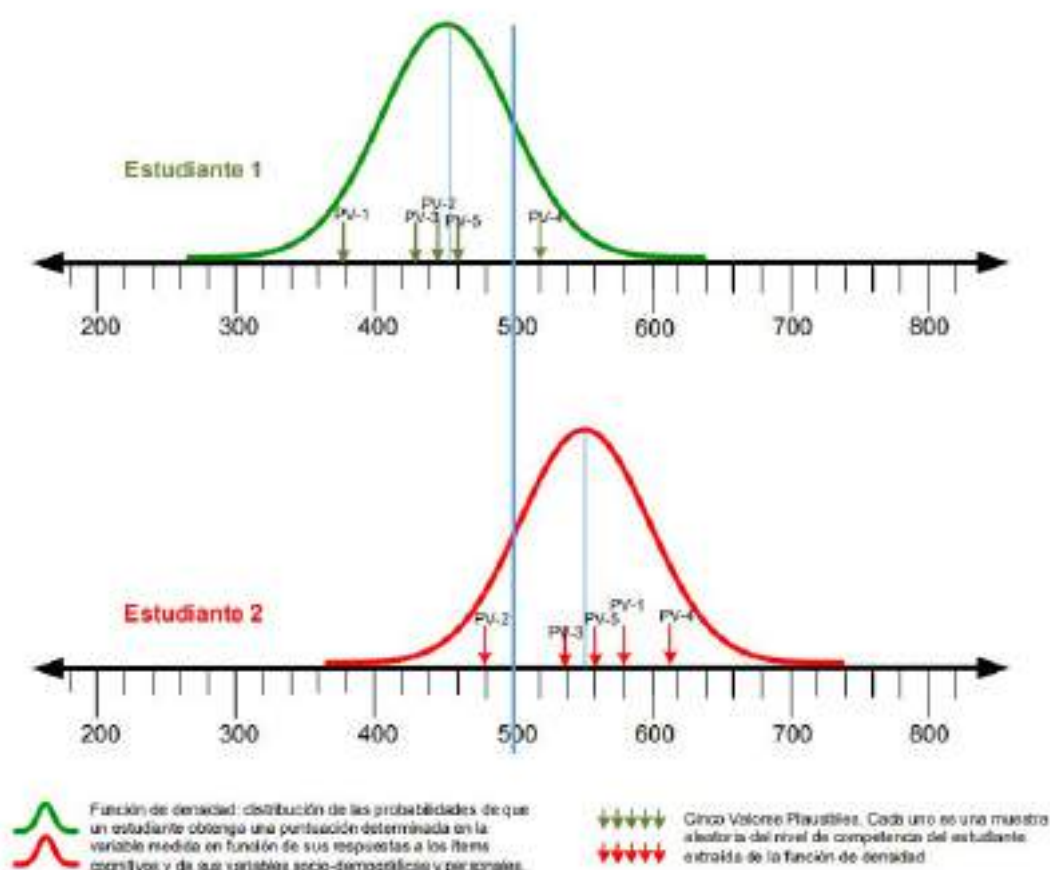
TIMSS expresa sus puntuaciones numéricas como el promedio agregado a nivel poblacional (países), de estrato (regiones o tipos de centro) u otras variables de interés (género, condición de emigrante ...). Estas puntuaciones se dan en una escala continua y normalmente distribuida que, por convención, tiene como punto de referencia central 500 puntos y una desviación típica de 100 puntos *N(500,100)*. Estos parámetros de referencia fueron fijados en la primera edición del estudio (TIMSS 1995, Martin y Kelly 1997) y permanecen constantes en los diferentes ciclos. No obstante, puesto que en cada edición participa un número diferente de países, la media internacional, que se obtiene como la media aritmética de todos los países participantes, puede variar de un ciclo a otro.

La forma en que TIMSS y todas las comparaciones a gran escala de sistemas educativos estiman estas puntuaciones difiere de los procedimientos clásicos para el cálculo de los resultados individuales en la

evaluación educativa y de los aprendizajes. Tradicionalmente la evaluación educativa expresa sus resultados mediante estimadores puntuales, es decir, cada estudiante recibe una puntuación única que refleja su desempeño en una prueba o ejercicio (Muñiz, 2018). Sin embargo, en la evaluación de los sistemas educativos las puntuaciones individuales no tienen interés y, por ello, no se emplean los clásicos estimadores puntuales de rendimiento.

Mazzeo (2018) denomina a la evaluación comparada de sistemas educativos como “evaluación de resultados grupales” (*group-score assesement*) para resaltar que su objetivo es estimar y comparar parámetros poblacionales (promedios por país) y no evaluar desempeños individuales como ocurre en la mayoría de la evaluación educativa y de los aprendizajes. Además, está demostrado que los estimadores puntuales tradicionales presentan sesgos al recuperar los parámetros poblacionales (Von Davier, Gonzalez y Mislevy, 2009). Por ello, la evaluación de sistemas educativos ha desarrollado un procedimiento singular y específico para reportar los resultados cognitivos: los valores plausibles (VP).

La lógica de los VP es la siguiente: a partir de las respuestas del alumnado se calcula una distribución probable de la destreza de cada estudiante en las áreas y dominios correspondientes. Entonces, de dichas distribuciones se extraen al azar cinco VP para cada área y dominio evaluado. Por tanto, a la hora de interpretar los resultados en TIMSS no es posible afirmar que un estudiante concreto ha obtenido una puntuación puntual o determinada en matemáticas o ciencias.



Fuente: Fernández-Alonso y Muñiz (2019)

El cálculo de resultados mediante valores plausibles (VP) emplea un algoritmo que tiene en cuenta los siguientes parámetros: el vector con las respuestas de los estudiantes a un conjunto de ítems; la dificultad y otros parámetros de dichos ítems; el peso muestral de cada estudiante; y las características sociodemográficas del alumnado y otras variables personales y de interés, que funcionan como covariables en la estimación. Con esta información se estima una función de densidad multivariada a posteriori que describe la distribución de puntuaciones probables de cada estudiante conocidos los parámetros señalados. De dicha distribución en TIMSS se extraen aleatoriamente cinco puntuaciones, que se conocen como VP.

La figura ejemplifica las funciones de densidad de dos estudiantes de características sociodemográficas y personales similares que respondieron a los mismos ítems. El estudiante 2 respondió correctamente a más ítems que el estudiante 1 y, por ello, su función de densidad se ubica más a la derecha en la escala $N(500,100)$. No obstante, los valores probables de cada estudiante (representados por la curva normal) presentan un rango muy amplio. Nótese que, al extraer aleatoriamente los 5 VP de cada estudiante, en general, los valores probables del estudiante 2 son más altos que los del estudiante 1. Sin embargo, el VP-2 del estudiante 2 (en torno a 480 puntos) es inferior al VP-4 del estudiante 1 (en torno a 520). Es por esta razón que los PV, al contrario que los estimadores puntuales, no pueden emplearse para reportar resultados individuales y solo se usan en la evaluación de sistemas educativos.

Por otro lado, debido al diseño de las pruebas, tampoco es posible asignar puntuaciones estimadas a los centros escolares. Las evaluaciones a gran escala como TIMSS emplean un gran número de ítems o preguntas de evaluación que se organizan en diferentes modelos de pruebas (o cuadernillos de evaluación) de acuerdo con algún arreglo experimental y, el día de la prueba, cada estudiante responde a un modelo de cuadernillo (Frey *et al.*, 2009). Sin embargo, el número de estudiantes que se seleccionan (por lo general un grupo-aula de 4.º de Educación Primaria) es muy inferior al elevado número de cuadernillos que conforman el juego completo de preguntas de la evaluación. Ello impide, por tanto, obtener una puntuación media de centro con suficientes garantías de fiabilidad.

2.1.1.2. Resultados por niveles de rendimiento

Las puntuaciones numéricas que se acaban de describir son sintéticas, permiten la comparación de resultados poblacionales y suelen tener impacto mediático. Sin embargo, presentan una limitación importante ya que no describen las competencias medidas. En otras palabras, no responden a preguntas del tipo: ¿qué sabe hacer un estudiante que obtiene 520 puntos en matemáticas?, ¿qué puede hacer un estudiante que obtenga 550 puntos frente a otro con solo 500 puntos?

Las escalas o niveles de rendimiento superan esa limitación ya que traducen los resultados numéricos a estándares de desempeño que describen detalladamente los conocimientos y competencias de las poblaciones de estudiantes.

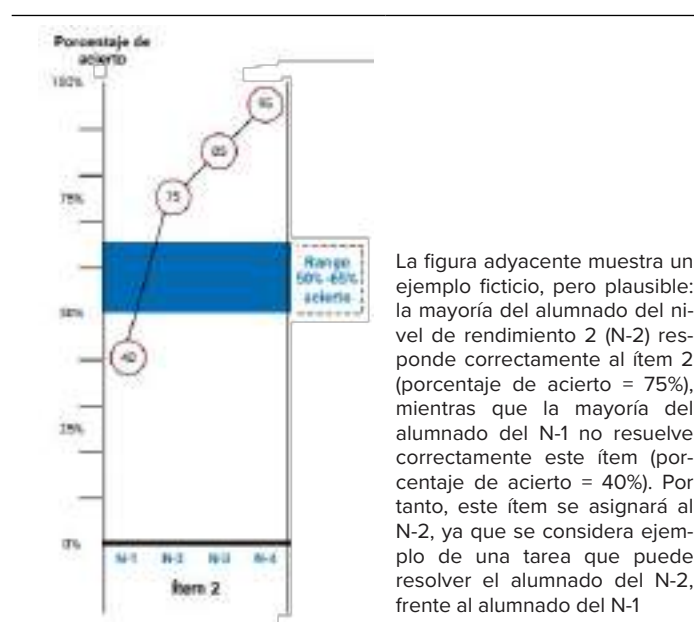
La construcción de escalas y niveles de rendimiento es un proceso que supone tres etapas o tareas: determinar puntos de corte en la escala numérica para establecer los niveles de rendimiento; asignar los ítems o preguntas de la prueba a un nivel de rendimiento; y elaborar descripciones que resuman las competencias del alumnado en cada uno de los niveles de desempeño.

Establecer puntos de corte para dividir la escala de resultados en diferentes niveles de rendimiento es un procedimiento no exento de arbitrariedad pero que, al tiempo, es práctico y eficiente. Su lógica es muy similar al uso que la industria textil hace de las medidas antropométricas. Los rasgos físicos de las personas, al igual que las puntuaciones en la escala $N(500,100)$ de TIMSS, son muy variables y también se expresan en escalas numéricas y continuas. Por ejemplo, el ancho de la cadera de los hombres oscila normalmente entre 65 y 150 centímetros, es decir, en un rango de 85 centímetros. Sin embargo, a efectos prácticos, la industria textil colapsa o agrupa este rango en unas pocas categorías: talla S (entre 78 y 85 cm.); talla M (entre 86 y 94 cm.); talla L (entre 95 y 99 cm.) y así sucesivamente. Salvando las distancias, determinar puntos de corte sigue la misma idea: arbitrar unos límites o intervalos en una escala continua para agrupar las puntuaciones en unos pocos niveles de desempeño. Las evaluaciones a gran escala han desarrollado diferentes procedimientos para establecer puntos de corte. En concreto, TIMSS señala cuatro hitos o puntos de corte en la escala de puntuaciones continuas: 400, 475, 550 y 625 puntos (Mullis *et al.*, 2016). De este modo se crean cinco grupos o niveles de rendimiento (Cuadro 2.1).

Cuadro 2.1. Niveles de rendimiento y puntos de corte de la escala

Niveles de rendimiento	Escala
Muy Bajo	Menos de 400 puntos
Bajo	De 400 a menos de 475 puntos
Intermedio	De 475 a menos de 550 puntos
Alto	De 550 a menos de 625 puntos
Avanzado	625 puntos o más

Una vez que cada estudiante ha sido clasificado en uno de los cinco grupos o niveles de rendimiento, la segunda tarea consiste en asignar cada ítem o pregunta de la evaluación a uno de los niveles de rendimiento. Ello permite identificar los ítems que acierta la mayoría del alumnado de un determinado nivel y que, al mismo tiempo, falla la mayoría del alumnado del nivel inmediatamente inferior. Existen dos procedimientos según se comparen porcentajes o probabilidades de acierto, si bien TIMSS se decanta por el primero. Para ello se calcula el porcentaje de acierto de los ítems en cada nivel de rendimiento y, a continuación, los ítems se asignan al nivel de rendimiento correspondiente atendiendo a diferentes criterios. Mullis *et al.*, (2016) describen con detalle los criterios de asignación, aunque con carácter general un ítem se asigna a determinado nivel cuando el alumnado de dicho nivel presenta un porcentaje de acierto superior al 65 % y el alumnado de los niveles inferiores presenta un porcentaje de acierto inferior al 50 %.



Distribuidos los ítems a su nivel de rendimiento correspondiente, la última etapa consiste en un análisis curricular de los ítems. Un panel de expertos recibe los ítems ordenados por el nivel de desempeño para desarrollar el siguiente encargo de tareas: (1) elaborar pequeñas descripciones de lo que supone responder correctamente a cada uno de los ítems; (2) redactar una descripción general que resuma y caracterice cada nivel de desempeño; y (3) seleccionar algunos ítems que ejemplifiquen las competencias propias de cada nivel de rendimiento.

2.1.1.3. Los errores de media

Toda medida lleva asociada un error, una incertidumbre. Cuando se afirma que el promedio de un país es de, por ejemplo, 500 puntos, debe entenderse que dicha estimación está sujeta a cierta incertidumbre. Es decir, no debería descartarse la posibilidad de que 499 o 501 puntos fueran puntuaciones posibles para dicho país. Todos los estimadores que se recogen en las tablas de este informe (medias, porcentajes, coeficientes de regresión...) llevan aparejados su error de medida, que se reporta entre paréntesis. Este error señala el rango o intervalo de puntuaciones probables para un país, además de la primariamente estimada y reportada en las tablas.

Tradicionalmente en estadística se calculan los errores típicos de los estimadores asumiendo que los datos provienen de un muestreo aleatorio simple. Sin embargo, en TIMSS, como en todas las evaluaciones educativas a gran escala, no es posible mantener esta asunción ya que los muestreos empleados tienen al menos dos etapas (primero se seleccionan centros y luego estudiantes o aulas completas).

Cuando, en las evaluaciones a gran escala, se calcula el error típico de los estimadores asumiendo que los datos provienen de un muestreo aleatorio simple, dicho error tiende a infraestimarse. La consecuencia de tener errores más pequeños es el aumento de falsos positivos, es decir, de encontrar diferencias estadísticamente significativas entre los países y grupos comparados, cuando realmente dichas diferencias no existen.

Para evitar el riesgo de falsos positivos, el cálculo del error de los estimadores en las evaluaciones a gran escala debe tenerse en cuenta, además de la variabilidad de las estimaciones propiamente dichas, otras dos fuentes potenciales de variabilidad de los datos: las diferencias debidas al muestreo y la variabilidad proveniente del hecho de que cada estudiante recibe 5 VP y no una estimación puntual. Para el cálculo de ese error se emplean procedimientos de remuestreo que consisten en replicar, siguiendo un plan sistemático, múltiples muestras a partir de la muestra original y calcular el parámetro de interés (media, porcentaje...) para cada una de las réplicas. La variabilidad entre todas las replicaciones resultantes con respecto al parámetro calculado con la muestra original es el error estándar o típico de la estimación del parámetro correspondiente (OECD, 2009).

2.1.2. Estructura del capítulo

El capítulo se organiza en los siguientes apartados. Los apartados 2.2 y 2.3 están dedicados a matemáticas. El primero de ellos presenta los resultados como puntuaciones numéricas, tanto generales (subapartado 2.2.1) como por dominios de contenido y cognitivos (subapartado 2.2.2). Por otra parte, el apartado 2.3 comienza describiendo las escalas o niveles de rendimiento en matemáticas (subapartado 2.3.1) y, a continuación, compara la distribución del alumnado en dichos niveles en España, las nueve comunidades autónomas que ampliaron muestra, los países seleccionados y los parámetros internacionales.

Los apartados 2.4 y 2.5 están dedicados a ciencias y replican el mismo contenido de los apartados dedicados a matemáticas.

El apartado 2.6 se centra en analizar la relación entre los resultados en matemáticas y ciencias en las poblaciones ya mencionadas: España, comunidades autónomas con muestras representativas, países seleccionados y parámetros internacionales.

El apartado 2.7 analizará la evolución del rendimiento de España en matemáticas y ciencias en las últimas cuatro ediciones de TIMSS (2011, 2015, 2019 y 2023) en las cuales nuestro país ha concurrido con muestras representativas y comparables de 4.º de Educación Primaria. Ello permite trazar una tendencia de resultados que abarca algo más de una década y compararla internacionalmente con la tendencia de la OCDE durante el mismo periodo temporal.

Finalmente, el último apartado de este capítulo estará dedicado a presentar los resultados en la materia conciencia medioambiental. Este apartado se subdivide en dos puntos, uno de ellos dedicado a comparar el rendimiento promedio de las poblaciones ya mencionadas y el otro a presentar las escalas y niveles de rendimiento en conciencia medioambiental.

2.2. Rendimiento en matemáticas

Como se acaba de mencionar, la presentación de los resultados en matemáticas se organiza en dos apartados. En el primero se comparará el rendimiento promedio de los países seleccionados y de las comunidades autónomas con datos representativos con los parámetros internacionales (Promedio OCDE y Total UE). En el segundo se mostrarán los resultados por los dominios de contenido y cognitivos que conforman la especificación del rendimiento matemático según TIMSS.

2.2.1. Rendimiento promedio en matemáticas

La Figura 2.1 representa la distribución de las puntuaciones en matemáticas de los parámetros internacionales (Promedio OCDE y Total UE), los países seleccionados y las comunidades autónomas con muestras ampliadas. La figura contiene los siguientes datos: puntuación media de matemáticas, junto con el intervalo de confianza al 95 % para la media poblacional, y las puntuaciones correspondientes con los cuartiles (25, 50 y 75) y los percentiles 5 y 95. En el eje inferior están representados los niveles de rendimiento en matemáticas (desde muy bajo hasta avanzado), que serán tratados con mayor detalle en el apartado 2.3. Los países, por un lado, y las comunidades autónomas, por otro, se ordenan descendientemente por la puntuación media estimada en la competencia matemática.

Los promedios de puntuación más altos corresponden a Corea (594 puntos) y Japón (591 puntos). Las evidencias aportadas por TIMSS permiten afirmar que el rendimiento de los dos países asiáticos es mejor que el del resto de países seleccionados. A continuación, se encuentran Lituania (561), Turquía (553) y Reino Unido (Inglaterra) (552 puntos). En este punto debe notarse que la muestra de alumnado turco estaba en el quinto año de escolarización obligatoria y no en el cuarto como ocurre con el resto de los países. En todo caso, la media de estos tres países supera los 550 puntos y, por tanto, coloca su promedio en el nivel de rendimiento alto en matemáticas. Dentro de la Unión Europea, además de la mencionada Lituania, los mejores rendimientos corresponden a Polonia e Irlanda (ambos con 546) y Rumanía (542).

En el extremo contrario se encuentran cinco países cuya puntuación está por debajo de 500 puntos. Tres de ellos pertenecen a la UE: España, Bélgica (Fr.) y Francia, siendo los otros dos Nueva Zelanda y Chile. Este último es el único país seleccionado cuyo promedio de puntuación (444 puntos) cae sobre la franja que corresponde con el nivel de rendimiento bajo en matemáticas.

El resultado de España (498 puntos) está ligeramente por debajo de los 500 puntos y es inferior al obtenido cuatro años atrás en TIMSS 2019 (502 puntos). En términos estadísticos el promedio español es significativamente más bajo que el Promedio OCDE (525 puntos) y el Total UE (514 puntos) y similar al obtenido por Canadá (504) y Nueva Zelanda (490).

Dentro de España las comunidades autónomas con muestras ampliadas presentan variaciones. Entre las comunidades autónomas participantes con muestra ampliada, Castilla y León, Principado de Asturias, Comunidad de Madrid y Comunidad Foral de Navarra son las regiones con mejores puntuaciones (entre 522 y 517 puntos), todas ellas por encima del Total UE (514 puntos) y próximas a la media de OCDE (525 puntos). En todo caso, desde el punto de vista estadístico no existen diferencias significativas entre el desempeño de las cuatro comunidades autónomas mencionadas y los dos promedios internacionales. En el extremo contrario se encuentran Cataluña, Canarias e Illes Balears, todas ellas con puntuaciones medias inferiores a la media de España. En el caso de Illes Balears su promedio se sitúa en la frontera entre los niveles de rendimiento bajo e intermedio.

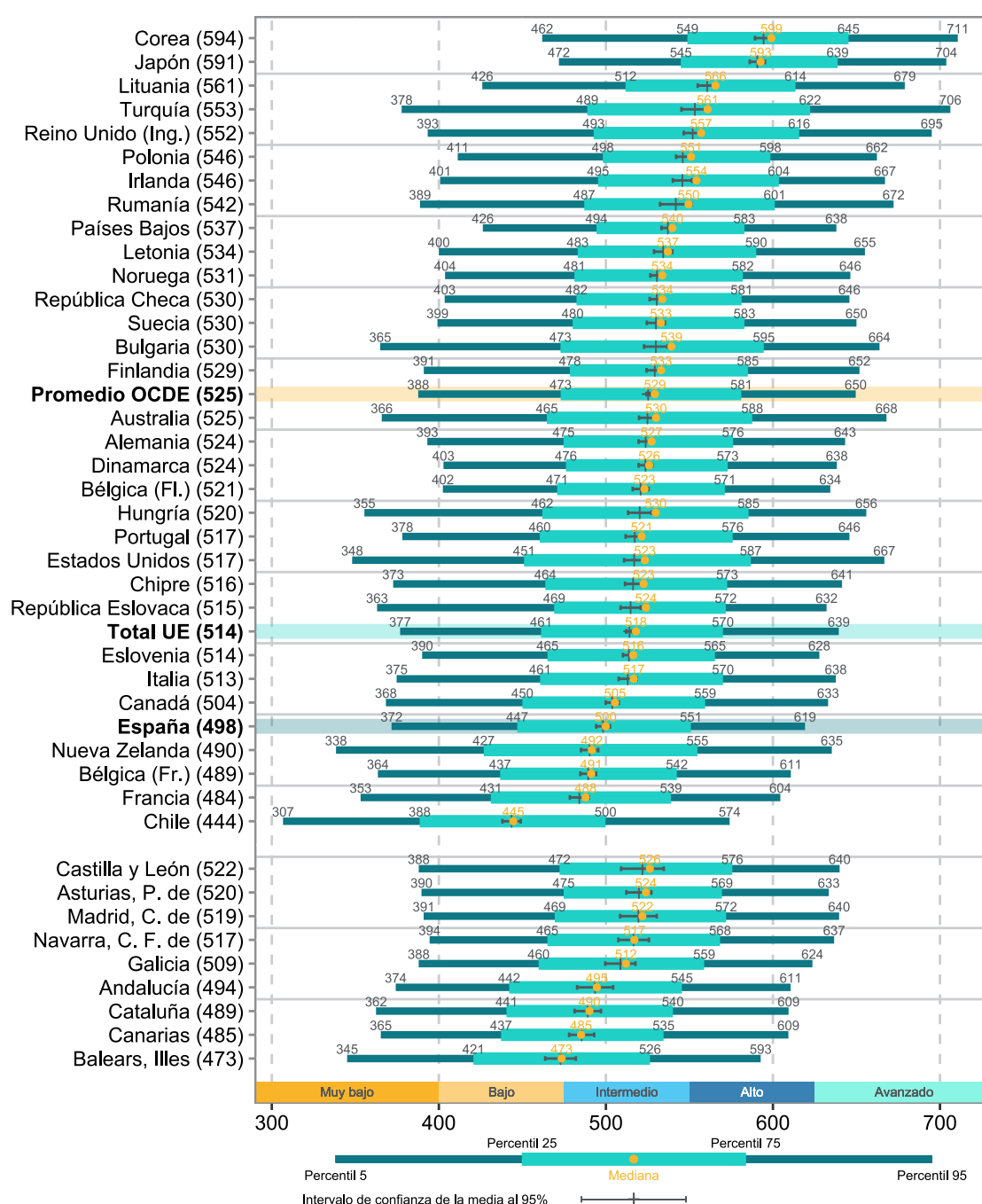
Como se mencionó anteriormente, la Figura 2.1 también muestra los cuartiles de la distribución de las puntuaciones de matemáticas y los percentiles del 5 % y el 95 %. Estos últimos proporcionan información sobre la variabilidad de las puntuaciones, una vez descartado el 5 % de las puntuaciones más bajas y más altas de la distribución. Dicho de otro modo, el tamaño del rango de puntuaciones comprendidas entre los dos percentiles extremos se interpreta como una evidencia del grado de homogeneidad de del desempeño del alumnado.

Los parámetros internacionales indican que la diferencia entre el alumnado con mejores y peores resultados en matemáticas está en torno a 260 puntos (Promedio OCDE, 262 y Total UE 262). Es decir, una distancia

superior a 2,5 puntos en unidades de desviación típica. Turquía, Estados Unidos, Australia, Reino Unido (Ing.) y Hungría son los países con mayores diferencias en el desempeño en matemáticas. En todos estos casos el rango está por encima de los 300 puntos, indicando importantes diferencias en el rendimiento en matemáticas. De otra parte, las distribuciones con menos variabilidad corresponden a Países Bajos (212 puntos), Japón (232), Bélgica (Fl., 232) y Eslovenia (238).

España, con una diferencia entre los percentiles 5 y 95 de 247 puntos, presenta una variabilidad en la distribución de puntuaciones de matemáticas inferior tanto a la del Promedio OCDE como a la del Total UE (ambas con 262 puntos).

Figura 2.1. Distribución de los resultados de matemáticas (percentiles) y puntuación media estimada con intervalo de confianza al 95 % para la media poblacional



El rango de la diferencia entre los percentiles 5 y 95 de las comunidades autónomas con muestra propia oscila entre 237 puntos de Andalucía y los 252 de Castilla y León. Se trata en todos los casos de diferencias más pequeñas que los promedios internacionales, lo que confirma que, en comparación con los países seleccionados la distribución de las puntuaciones en España es relativamente más homogénea.

2.2.2. Dominios de contenido y cognitivos en matemáticas

Como se ha comentado en la introducción de este capítulo, la evaluación del área de matemáticas comprende dominios de contenido y dominios cognitivos. En el Capítulo 1 se puede ver el peso que tienen en la evaluación del área de matemáticas cada uno de los dominios de las dos dimensiones, de contenido y cognitiva. A su vez, cada dimensión se estructura en tres dominios. Así, la dimensión cognitiva distingue entre los dominios: números, medidas y geometría y datos, y la dimensión cognitiva se organiza sobre los dominios: conocimiento, aplicación y razonamiento (Cuadro 2.2). En este apartado se mostrarán los resultados segregados por los dominios que se acaban de mencionar. En todas las figuras que se mostrarán este apartado las poblaciones comparadas se ordenan descendientemente por el promedio en matemáticas.

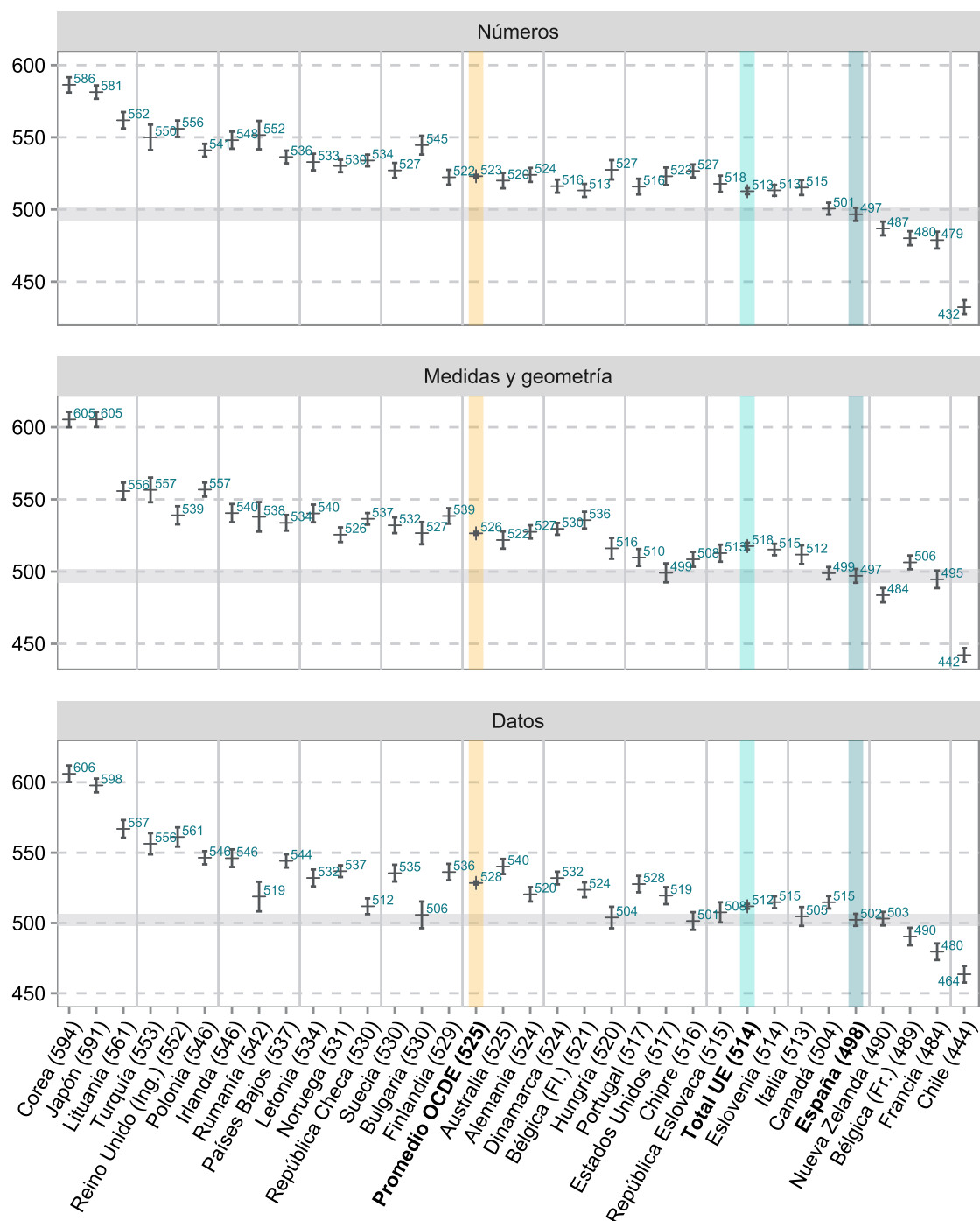
Cuadro 2.2. Distribución de porcentajes de los dominios de contenido y cognitivos en la evaluación de matemáticas

MATEMÁTICAS	DOMINIOS DE CONTENIDO	%	DOMINIOS COGNITIVOS	%
	Números	50 %	Conocimiento	40 %
	Medidas y geometría	30 %	Aplicación	40 %
	Datos	20 %	Razonamiento	20 %

2.2.2.1. Resultados en los dominios de contenido de matemáticas

La Figura 2.2.a muestra las distribuciones de los resultados de los dominios de números, medidas y geometría y datos en los países seleccionados, el Promedio OCDE y el Total UE. La banda gris que recorre horizontalmente el gráfico señala el intervalo de confianza al 95 % para la media poblacional de ciencias de España en cada dominio de contenido. Los países están ordenados por su promedio en la escala general de matemáticas. Este tipo de análisis permite detectar puntos fuertes y débiles de cada país. Por ejemplo, el promedio de Japón en el dominio geometría y medidas (605 puntos) es relativamente mejor que el promedio en el dominio números (581). Sin embargo, en Bulgaria los resultados en números (545 puntos) son claramente superiores a los obtenidos en el dominio de datos (506 puntos).

Figura 2.2.a. Puntuación media estimada con intervalo de confianza al 95 % para la media poblacional de los dominios de contenido en matemáticas. Países, Total UE y Promedio OCDE



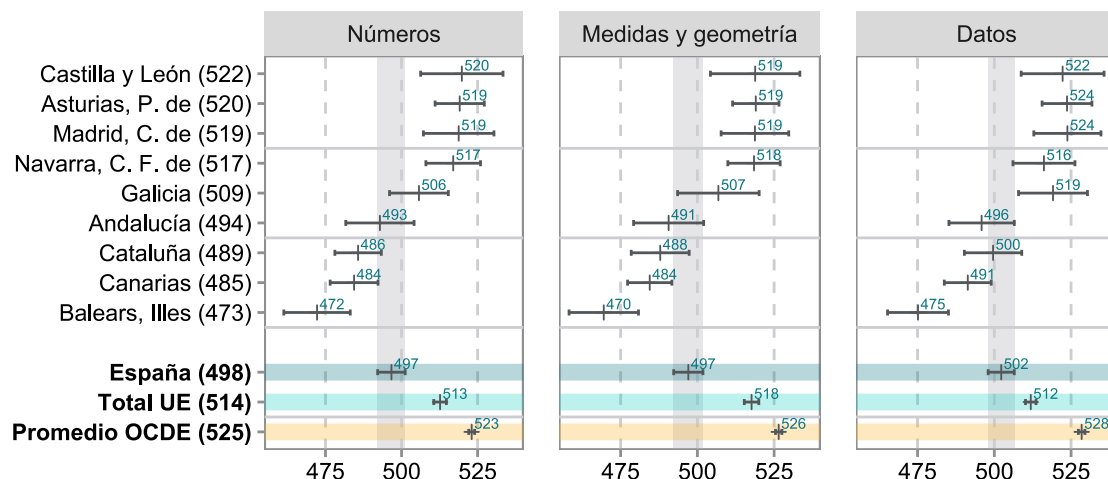
En clara relación con los resultados generales, el alumnado de Corea y Japón es el que obtiene la mayor puntuación en todos los dominios y, por el contrario, Chile el país que peor puntuación muestra en los tres dominios de contenido.

España presenta resultados similares en los tres dominios: 497 puntos en números y medidas y geometría y 502 puntos en datos, por lo que no se observan grandes desequilibrios en la comparación por dominios. En los tres dominios de contenido el promedio español es inferior a los parámetros internacionales y en todas las comparaciones las diferencias son estadísticamente significativas. Con respecto al promedio de la OCDE, España se encuentra a 26 puntos en números y datos, y a 29 puntos en medidas y geometría. En relación con el Total UE, el promedio de España está a 21 puntos en medidas y geometría, a 16 en numeración y a 10 en datos.

En todo caso, la posición relativa de España varía de un dominio de contenido a otro debido a los desequilibrios que presentan algunos países en los promedios de los tres dominios. En el dominio de los números es donde España presenta una posición relativa menos ventajosa en la comparación con el resto de los países. En este caso, el promedio de España (497 puntos) es similar al de Canadá (501 puntos). En medidas y geometría, además de Canadá, el promedio de España es estadísticamente igual al de Estados Unidos y Francia. La mejor puntuación relativa de España se encuentra en el dominio de datos. El promedio nacional (502 puntos) es estadísticamente similar al de países como Bulgaria, Hungría, Chipre, República Eslovaca, Italia y Nueva Zelanda.

La Figura 2.2.b recoge los resultados de los dominios de contenido, pero centrado en las comunidades autónomas con muestra representativa. En este caso, la banda gris que recorre verticalmente el gráfico señala el intervalo de confianza al 95 % para la media poblacional de España en cada dominio de contenido en matemáticas. Las regiones están ordenadas por su promedio en la escala general de matemáticas. En términos generales no se advierten grandes desequilibrios en las puntuaciones por dominios dentro de las comunidades autónomas. La mayor diferencia se observa en Cataluña, donde hay 14 puntos de diferencia entre el promedio del dominio números (486 puntos) y el dominio de datos (500). En el caso de Galicia se observan 13 puntos de diferencia entre el promedio de números (506 puntos) y el de datos (519). En todo caso, ninguna de estas diferencias alcanza el nivel de significación estadística, no pudiendo confirmarse, por tanto, la existencia de desequilibrios en los promedios por dominios de contenido dentro de las regiones.

Figura 2.2.b. Puntuación media estimada con intervalo de confianza al 95% para la media poblacional de los dominios de contenido en matemáticas. Comunidades Autónomas, España, Total UE y Promedio OCDE



En el dominio de números Castilla y León (520 puntos), Principado de Asturias (519), Comunidad de Madrid (519) y Comunidad Foral de Navarra (517) son las que muestran las puntuaciones medias más altas, similares desde el punto de vista estadístico al Promedio OCDE (523 puntos) y al Total UE (513 puntos), y significativamente mejores que la media de España (497 puntos). Galicia (506 puntos) ocupa una posición intermedia: su puntuación es estadísticamente más baja que el Promedio OCDE, pero similar al Total UE y al promedio de España. En el extremo contrario, Illes Balears se encuentra significativamente por debajo del promedio de España con un rango de puntuaciones a caballo entre los niveles de rendimiento bajo e intermedio.

En el dominio medidas y geometría la lectura es similar: hay un grupo de cuatro regiones (Castilla y León, Principado de Asturias, Comunidad de Madrid y Comunidad Foral de Navarra) sin diferencias estadísticas con respecto a la OCDE y la UE; otras tres (Galicia, Andalucía y Cataluña) por debajo de los promedios internacionales y atravesando el intervalo de confianza de la media de España; y, finalmente, Canarias e Illes Balears, cuyo promedio se ubica por debajo de la media nacional.

En el dominio de datos, Galicia (519 puntos) presenta su mejor resultado y supera, junto con las cuatro regiones con mejores puntuaciones en matemáticas, la media de España. La posición relativa del resto de comunidades autónomas con respecto a la franja que señala el intervalo de confianza de la media española es similar a la comentada en los otros dos dominios de contenido.

2.2.2.2. Resultados en los dominios cognitivos de matemáticas

La Figura 2.3.a muestra las distribuciones de los resultados de los dominios cognitivos de matemáticas en los países seleccionados. La banda gris que recorre horizontalmente el gráfico señala el intervalo de confianza al 95 % para la media poblacional de matemáticas de España en cada dominio cognitivo. Los países están ordenados por su promedio en la escala general de matemáticas. Como se señaló en el apartado anterior, este tipo de análisis permite detectar puntos fuertes y débiles de cada país.

En el dominio cognitivo de conocimiento, los estudiantes de España alcanzan una puntuación media estimada de 500 puntos, similar a la de Canadá (502 puntos), y solo por encima de Chile (433 puntos), Francia (484), Nueva Zelanda (488 puntos) y Bélgica (Fr.) (496 puntos). Dicha puntuación es significativamente inferior a la del Promedio OCDE (526) y a la del Total UE (515), y se encuentra a 100 puntos de la de Corea y a 91 puntos de la de Japón (países con las mayores puntuaciones). Asimismo, España se encuentra a más de 56 puntos de Lituania, que muestra la puntuación media más alta entre los países participantes de la UE (Figura 2.3.a).

En las comunidades autónomas españolas participantes, se mantiene la misma tendencia que en los dominios de contenido: Castilla y León (523 puntos), Comunidad de Madrid (522 puntos) y Principado de Asturias (521 puntos) muestran los mejores resultados en el dominio cognitivo de conocimiento, mientras que Cataluña (488 puntos), Canarias (486 puntos) e Illes Balears (469 puntos) obtienen las puntuaciones medias estimadas más bajas entre las comunidades (Figura 2.3.b).

Como se ha podido ver en los dominios de contenido, en los cognitivos la variabilidad sigue también un patrón similar. En el dominio cognitivo de aplicación de matemáticas, los estudiantes de España (497 puntos) no superan la barrera de los 500 puntos. Al igual que ocurrió en el dominio de conocimiento, España obtiene una de las puntuaciones más bajas entre los países seleccionados, a 16 puntos del Total UE, a más de 25 puntos del Promedio OCDE. Los alumnos españoles solo superan a los de Nueva Zelanda (489 puntos), de Bélgica (Fr.) (489 puntos), de Francia (484 puntos) y, especialmente, a los de Chile (444 puntos). Además, la puntuación obtenida por España en aplicación está a 100 puntos de Japón y a 96 puntos de Corea, de nuevo los países con las puntuaciones más altas (Figura 2.3.a).

Entre las comunidades autónomas se mantienen los mismos resultados que los observados para el dominio de conocimiento: los estudiantes de Castilla y León, Principado de Asturias y Comunidad de Madrid obtienen las puntuaciones medias más altas en el dominio de aplicación, más de 20 puntos más altas que la media de España (Figura 2.3.b). Del mismo modo, los alumnos de Illes Balears son los que peor rendimiento obtienen (471 puntos), a 54 puntos del Promedio OCDE y a 42 puntos del Total UE.

La tendencia observada en los dominios precedentes se mantiene en el dominio cognitivo de razonamiento: 500 puntos (puntuación semejante a la obtenida por Canadá). Asimismo, se mantienen las diferencias con el Promedio OCDE (24 puntos) y con el Total UE (14 puntos). En comparación con Japón y Corea, países que, como en los anteriores dominios, presentan los mejores resultados entre los seleccionados, la diferencia está por encima de los 75 puntos. Igualmente, por debajo de la puntuación de España solo están la de Nueva Zelanda (495 puntos), Bélgica (Fr.) (484 puntos), Francia (482 puntos) y Chile (453 puntos).

En las comunidades autónomas participantes, también se repite el orden establecido en los dominios anteriores, tanto en la tendencia como en la magnitud de las puntuaciones medias estimadas (Figura 2.3.b).

En resumen, el alumnado de 4.º de Educación Primaria de España obtiene en el dominio cognitivo de aplicación de matemáticas resultados significativamente peores que en los otros dos dominios cognitivos. Además, la puntuación media estimada en aplicación (497) es semejante a la media general de España (498) en matemáticas, mientras que la puntuación estimada de razonamiento y conocimiento (500 puntos) queda ligeramente por encima de la media general.

Figura 2.3.a. Puntuación media estimada con intervalo de confianza al 95 % para la media poblacional de los dominios cognitivos en matemáticas. Países, Total UE y Promedio OCDE

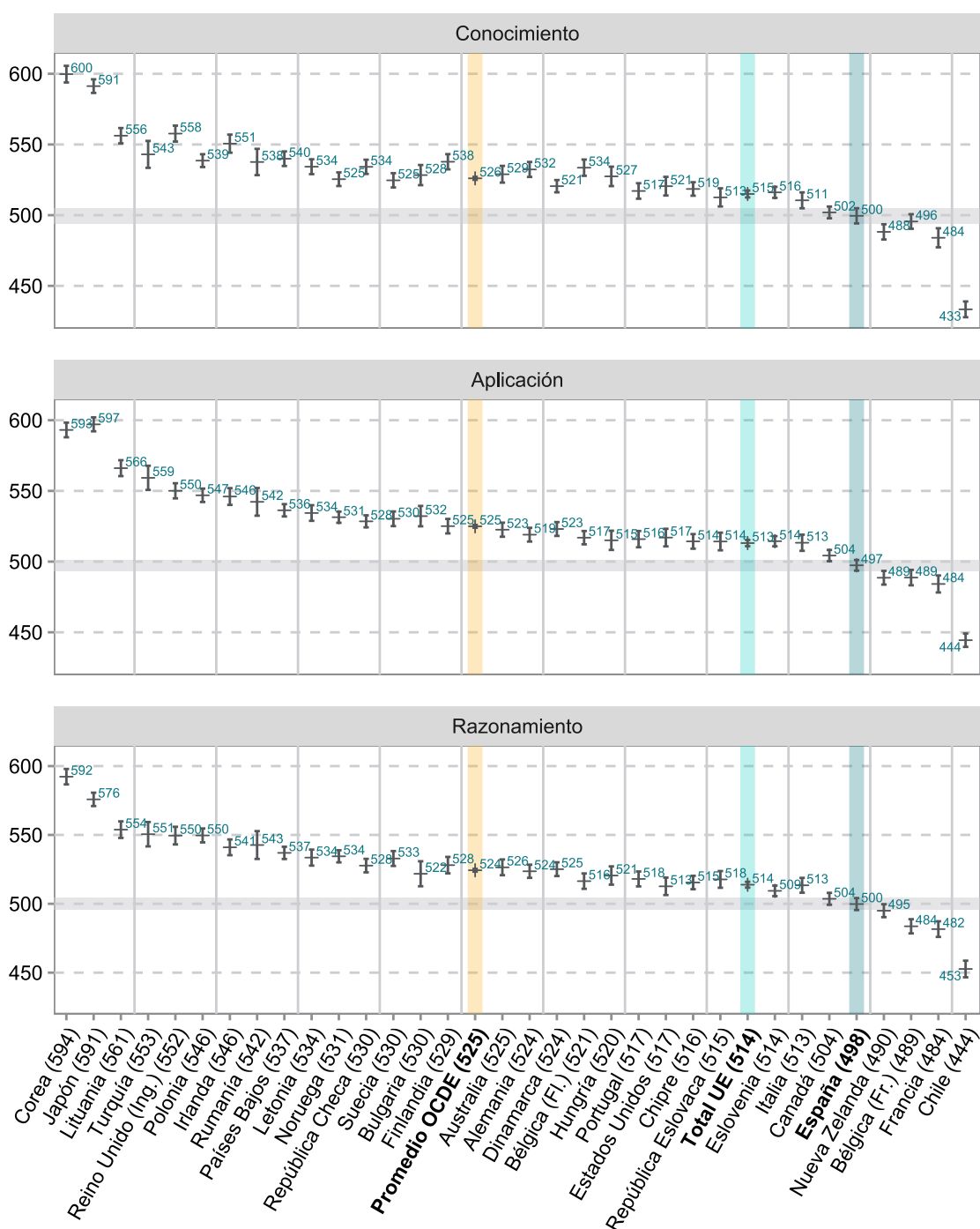
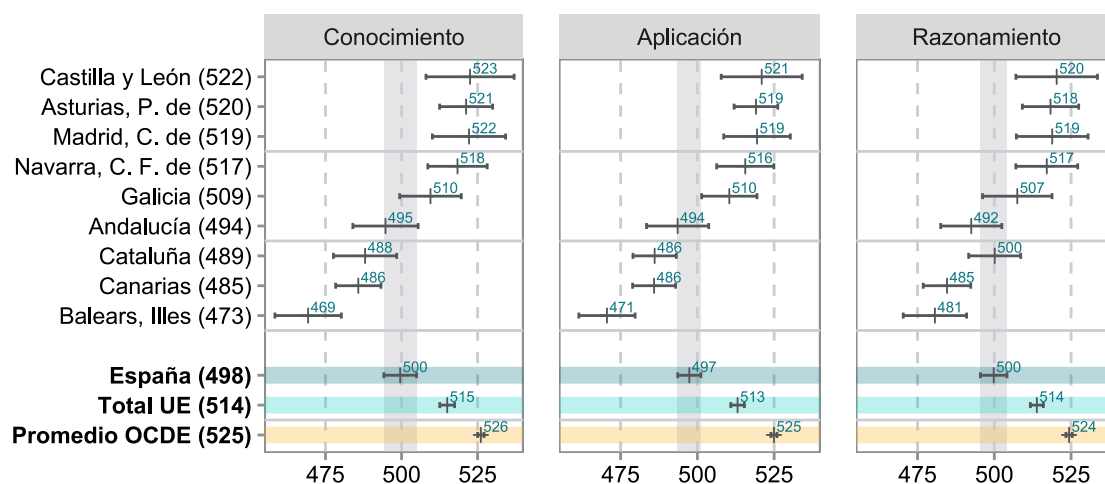


Figura 2.3.b. Puntuación media estimada con intervalo de confianza al 95 % para la media poblacional de los dominios cognitivos en matemáticas. Comunidades Autónomas, España, Total UE y Promedio OCDE



2.3. Escalas y niveles de rendimiento en matemáticas

Este apartado estará dedicado a presentar los resultados del alumnado por escalas y niveles de rendimiento en matemáticas. El mismo se divide en dos puntos. El primero describe los niveles de rendimiento en la escala de matemáticas de TIMSS 2023, mientras que el segundo mostrará la distribución del alumnado en los niveles de rendimiento descritos previamente.

2.3.1. Descripción de los niveles de rendimiento de la escala en matemáticas

El rendimiento en el área de matemáticas se clasifica en cuatro niveles: avanzado, alto, intermedio y bajo. A continuación, se describe lo que son capaces de hacer los estudiantes que alcanzan un determinado nivel de rendimiento en matemáticas junto con el intervalo de puntos correspondiente a cada uno de los niveles y un ejemplo ilustrativo de cada uno de los niveles. A los cuatro niveles que se describen habría que añadir un quinto nivel, que denominaremos como muy bajo, para aquellos estudiantes que no alcancen los 400 puntos en las pruebas de evaluación y que, por tanto, no son capaces de realizar las tareas del nivel bajo.

Nivel avanzado: 625 o más puntos

TIMSS 2023 ME71040

00:00

Un entrenador necesita poner 40 pelotas de tenis y 10 balones de fútbol en bolsas.
En cada bolsa caben 8 pelotas de tenis o 2 balones de fútbol.



¿Cómo puede calcular el entrenador el número total de bolsas que necesita?

(A) $40 + (10 + 2)$
 (B) $(40 + 2) + (10 + 8)$
 (C) $(40 + 8) + (10 + 2)$
 (D) $(40 + 8) + (10 + 2)$

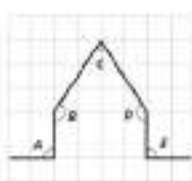
IEA TIMSS & PIRLS INTERNATIONAL

Los estudiantes que alcanzan este nivel son capaces de seleccionar y relacionar información para aplicar las operaciones adecuadas a la resolución de problemas. Pueden interpretar los resultados de los cálculos en contextos de resolución de problemas, diferentes expresiones y patrones, y relacionar fracciones y decimales. Saben estimar y relacionar medidas, aplicar conocimientos sobre figuras bidimensionales y tridimensionales, identificar propiedades sencillas de rectas y ángulos y mostrar una comprensión básica de la superficie y el perímetro de figuras sencillas. Además, los estudiantes pueden interpretar datos y tomar decisiones sobre los datos proporcionados en numerosos contextos.

Nivel alto: de 550 a menos de 625 puntos

TIMSS 2023 ME71040

00:00



Identifica de qué tipo es cada ángulo de la imagen. Marca tus respuestas en la tabla.
El ángulo A ya está resuelto.

	Ángulo recto	Menor que un ángulo recto	Mayor que un ángulo recto
A	☒	☐	☐
B	☐	☐	☐
C	☐	☐	☐
D	☐	☐	☐
E	☐	☐	☐

IEA TIMSS & PIRLS INTERNATIONAL

Los estudiantes en este nivel son capaces de relacionar conceptos o representaciones en contextos amplios. Pueden aplicar el conocimiento de las propiedades de los números naturales para justificar una solución. Demuestran comprensión de la recta numérica, los múltiplos, los divisores, el redondeo de números y las operaciones con fracciones y decimales. Los estudiantes pueden resolver tareas de medición en numerosos contextos. Pueden relacionar figuras bidimensionales con figuras tridimensionales desconocidas y demostrar una comprensión básica de los ángulos. Asimismo, los estudiantes pueden interpretar distintas características de representaciones de datos y, a su vez, representar datos en una variedad de gráficos.

Nivel intermedio: de 475 a menos de 550 puntos

Los alumnos de una clase han hecho tres animales de papel diferentes utilizando hojas azules, rojas y amarillas. La siguiente tabla muestra el número de animales que se han hecho de cada color.

Animal	Color del papel		
	Azul	Rojo	Amarillo
Tortuga	8	4	3
Jirafa	3	2	10
Pez			

Completa la tabla para que se cumpla lo siguiente:

- Hay el mismo número de peces azules que de jirafas amarillas.
- Hay el mismo número de peces rojos que de los otros dos animales rojos juntos.
- Hay 74 animales amarillos en total.

Los estudiantes en este nivel son capaces de demostrar sus conocimientos matemáticos en situaciones sencillas y relacionar representaciones. Pueden realizar cálculos con números naturales de tres cifras en diversas situaciones. Son capaces de sumar y ordenar decimales sencillos, así como medir distancias en línea recta y describir figuras tridimensionales. Además, son capaces de utilizar datos de múltiples fuentes para relacionar representaciones.

Nivel bajo: de 400 a menos de 475 puntos

Los estudiantes en este nivel muestran una comprensión matemática básica. Pueden sumar y restar números naturales de hasta tres cifras, multiplicar y dividir números naturales de una cifra y resolver problemas sencillos. Asimismo, son capaces de aplicar ideas básicas de medida y de las propiedades de las figuras geométricas comunes. Además, pueden leer datos de diferentes representaciones y completar gráficos de barras sencillos.

Nivel muy bajo: menos de 400 puntos

El alumnado de este nivel no es capaz de realizar las tareas del nivel bajo.

2.3.2. Distribución del alumnado por niveles de rendimiento

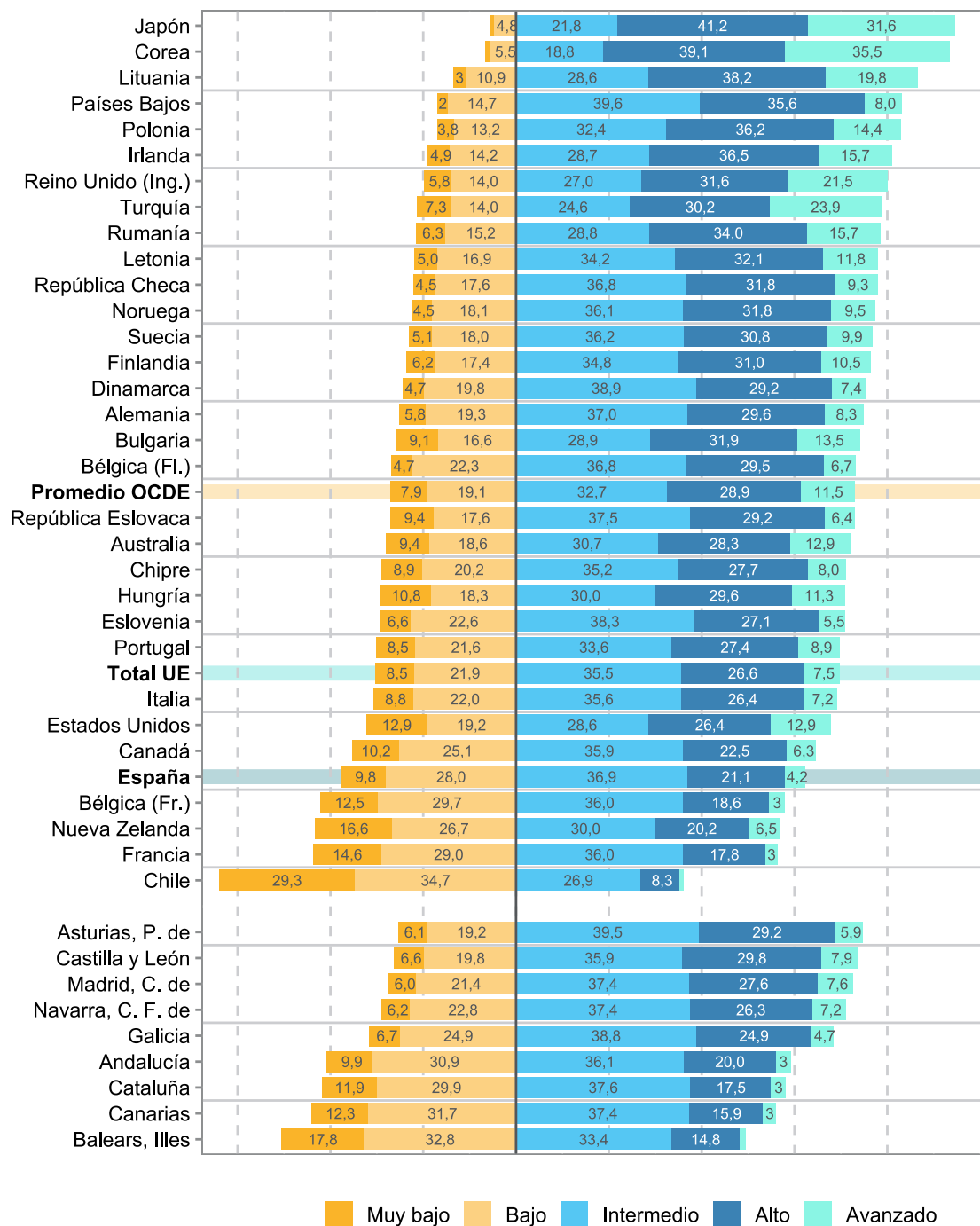
En la Figura 2.4 se representa la distribución del porcentaje de estudiantes de cada país, organismo internacional y comunidades autónomas españolas con muestra ampliada en los cinco niveles de la escala de rendimiento de matemáticas. Se han ordenado por los países y las comunidades de menor a mayor porcentaje de estudiantes que no alcanzan el valor intermedio, es decir, que están en los niveles bajo o muy bajo de la escala.

En la media de países de la OCDE, el 27,0 % de los estudiantes no alcanza el nivel intermedio (7,9 % nivel muy bajo; 19,1 % nivel bajo), 3,4 puntos porcentuales menos que el Total UE (8,5 % muy bajo; 21,9 % bajo). Dentro del conjunto de la OCDE, Japón y Corea son, con diferencia, los países con menor porcentaje de estudiantes con rendimiento bajo o muy bajo. En concreto, Japón alcanza solo un 5,4 % (0,6 % muy bajo; 4,8 % bajo) y Corea un 6,6 % (1,1 % muy bajo; 5,5 % bajo). También con un nivel relativamente bajo, menor del 20% de rendimiento bajo o muy bajo, están Lituania (13,9 %), Países Bajos (16,7 %), Polonia (17 %), Irlanda (19,1 %) y Reino Unido (Ing.) (19,8 %). Más allá del 27 % hay 13 países, destacando por niveles elevados de bajo o muy bajo rendimiento Bélgica (Fr.) (42,2 %: 12,5 % muy bajo; 29,7 % bajo), Nueva Zelanda (43,3 %: 16,6 % muy bajo; 26,7 % bajo) y Francia (43,6 %: 14,6 % muy bajo; 29,0 % bajo). Cabe destacar el elevado porcentaje de alumnos con bajo o muy bajo rendimiento de Chile (64,0 %: 29,3 % muy bajo; 34,7 % bajo). Finalmente, conviene resaltar el hecho de que Japón y Corea no solo tienen un número muy bajo de alumnos con bajo rendimiento, sino que, a la vez, son los dos países que mayor porcentaje tienen de alumnos con un rendimiento de nivel avanzado: Corea (35,5 %) y Japón (31,6 %).

En relación con los niveles Promedio de OCDE y Total UE, España presenta una proporción de estudiantes notablemente alta en los niveles bajos de la escala de matemáticas (37,8 %): 9,8 % en el nivel muy bajo y 28,0 % en el nivel bajo. Únicamente Chile, Francia, Nueva Zelanda y Bélgica (Fr.) obtienen porcentajes más altos en estos niveles de rendimiento (Figura 2.4). Por otra parte, cabe indicar que el porcentaje de estudiantes con rendimiento avanzado es pequeño (4,2 %), 3,3 puntos porcentuales menos que el del Total UE y 7,3 puntos porcentuales menos que el promedio de OCDE. Por último, exceptuando los cuatro países con peores resultados, España es el país de la OCDE que menor porcentaje tiene de estudiantes con un alto rendimiento en matemáticas (21,1 %).

En cuanto a los datos de las nueve comunidades autónomas que han participado en el estudio con muestra ampliada, cinco de ellas presentan niveles menores de rendimiento bajo o muy bajo respecto de la media de España y cuatro niveles mayores. En el lado positivo, están los datos del Principado de Asturias (25,3 %: 6,1 % muy bajo; 19,2 % bajo), Castilla y León (26,4 %: 6,6 % muy bajo; 19,8 % bajo), Madrid (27,4 %: 6,0 % muy bajo; 21,4 % bajo), Comunidad Foral de Navarra (29,0 %: 6,2 % muy bajo; 22,8 % bajo) y Galicia (31,6 %: 6,7 % muy bajo; 24,9 % bajo). En el lado negativo se encuentran Andalucía (40,8 %: 9,9 % muy bajo; 30,9 % bajo), Cataluña (41,8 %: 11,9 % muy bajo; 29,9 % bajo), Canarias (44,0 %: 12,3 % muy bajo; 31,7 % bajo) y, especialmente, Illes Balears, con más del 50 % de estudiantes con rendimiento bajo (32,8 %) o muy bajo (17,8 %). En cuanto al rendimiento intermedio, alto o avanzado, se observan datos similares (con escasas diferencias entre comunidades), excepto Illes Balears (cuyo porcentaje de alumnos es sensiblemente inferior al resto). En consecuencia, las diferencias se encuentran en el porcentaje de alumnos con un rendimiento alto y/o avanzado: mientras que Castilla y León, el Principado de Asturias o Comunidad de Madrid superan el 35 % de alumnos con rendimiento alto o avanzado, Cataluña, Canarias o Illes Balears no alcanzan el 20 %.

Figura 2.4. Porcentajes de estudiantes por niveles de rendimiento en matemáticas. Orden de menor a mayor en los niveles bajo y muy bajo



2.4. Rendimiento en ciencias

Al igual que los resultados en matemáticas, el rendimiento en ciencias se organiza en dos apartados. En el primero se comparará el rendimiento promedio de los países seleccionados y de las comunidades autónomas con datos representativos con los parámetros internacionales (Promedio OCDE y Total UE). En el segundo se mostrarán los resultados por los dominios de contenido y cognitivos que conforman la competencia científica en el programa de evaluación TIMSS.

2.4.1. Rendimiento promedio en ciencias

La Figura 2.5 recoge la distribución de las puntuaciones en ciencias de los países seleccionados, las comunidades con muestra comparable en el estudio y los parámetros internacionales: Promedio OCDE y Total UE. La figura también incluye la puntuación media estimada de ciencias junto con el intervalo de confianza al 95 % para la media poblacional. Los países están ordenados de mayor a menor puntuación media estimada.

Corea (583 puntos) presenta la puntuación media más alta entre los países seleccionados, y es significativamente más alta que la de Turquía (570 puntos), segundo país por puntuación media en ciencias. Además de Corea y Turquía, hay otros cuatro países (Reino Unido –Ing-, Japón, Australia y Polonia) con un promedio de 550 puntos o más. Por tanto, todos ellos sitúan sus puntuaciones promedio en el nivel de rendimiento alto de la escala TIMSS de ciencias.

El resto de los países, incluidas las regiones españolas y los parámetros internacionales, presenta puntuaciones que están dentro del rango del nivel de rendimiento intermedio, es decir, entre 475 y 549 puntos. No obstante, dentro de este rango de 75 puntos hay variaciones significativas. Así, los resultados de Finlandia (542) y Lituania (537) están cercanos al nivel de rendimiento alto y son, junto con la mencionada Polonia, los miembros de la UE con mejores puntuaciones medias. En el extremo contrario, Bélgica (Fl.), Francia, Chipre, Bélgica (Fr.) y Chile no alcanzan los 500 puntos. En los dos últimos casos, las puntuaciones numéricas ubican a su alumnado promedio cerca del límite del nivel de rendimiento bajo.

El rendimiento medio de España (504 puntos) es similar al de Italia y Portugal y estadísticamente inferior al Promedio OCDE (526 puntos) y del Total UE (513 puntos). Entre las comunidades autónomas con muestra representativa se observan variaciones: Principado de Asturias, Galicia, Castilla y León y Comunidad de Madrid presentan puntuaciones medias en ciencias por encima de 520 puntos, similares al Promedio OCDE y significativamente mejores que el Total UE. En el extremo contrario Canarias, Andalucía e Illes Balears no alcanzan los 500 puntos, y en este último caso el promedio está cerca del límite de puntuación que separa los niveles de rendimiento intermedio y bajo.

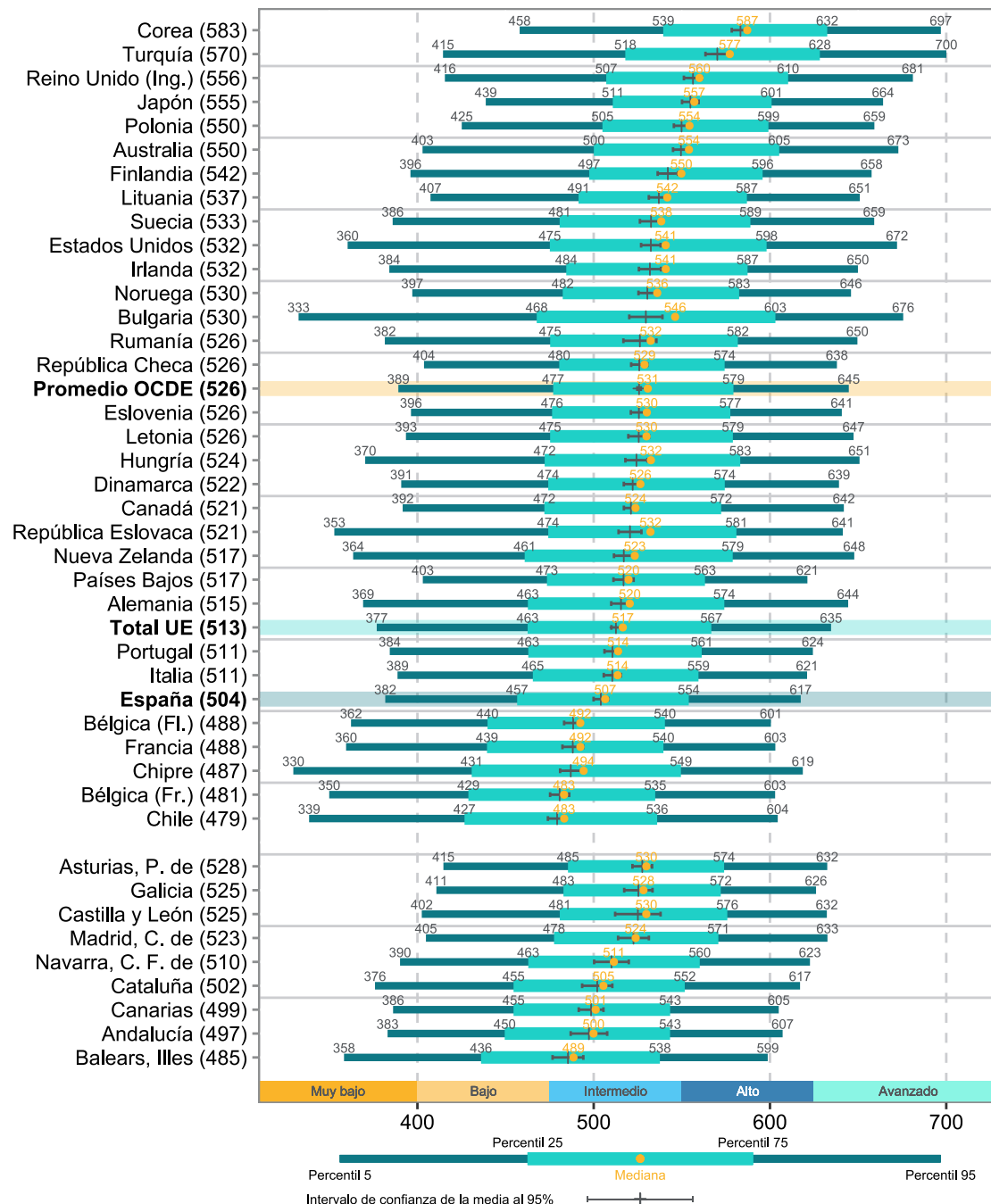
La Figura 2.5 muestra también las puntuaciones en ciencias correspondientes a los cuartiles de la distribución y a los percentiles 5 y 95. Estos últimos proporcionan información sobre la variabilidad de las puntuaciones, una vez descartado el 5 % de las puntuaciones más bajas y altas de la distribución. Las diferencias de puntuaciones correspondientes a los percentiles 95 y 5 pueden interpretarse como una medida de variabilidad en la distribución del rendimiento en ciencias.

Atendiendo al Promedio OCDE se observa que al percentil 5 corresponden 389 puntos, mientras que el percentil 95 se ubica en 645 puntos. Es otras palabras, en el conjunto de la OCDE el alumnado con puntuaciones más altas y más bajas está separado por 256 puntos, es decir, aproximadamente 2,6 desviaciones típicas en la escala $N(500,100)$ de TIMSS. Los países más equitativos en cuanto a la distribución del rendimiento en ciencias son Países Bajos (diferencia $p_{95} - p_5 = 218$ puntos) y Japón (225 puntos de diferencia). Por el contrario, Bulgaria tiene la variabilidad más alta entre los países seleccionados, casi 350 puntos de diferencia entre el percentil 95 y el percentil 5, indicando con ello importantes diferencias en el rendimiento de los estudiantes en ciencias.

España presenta una de las diferencias más pequeñas entre el alumnado con mejores y peores puntuaciones en ciencias: 235 puntos. Ello señala que la variabilidad en el dominio de los contenidos científicos evaluados por TIMSS es relativamente más pequeña que la observada en la mayoría de los países. Entre las comunidades

autónomas con muestras comparables, las diferencias más pequeñas se encuentran en Galicia (215 puntos), Principado de Asturias (217) y Canarias (219). Estos valores son similares a los ya mencionados de Países Bajos. La mayor diferencia corresponde a Cataluña (241 puntos) e Illes Balears (241). No obstante, el rango de las diferencias en todas las comunidades autónomas es inferior a los parámetros internacionales, lo que puede interpretarse como un rasgo de homogeneidad o cohesión dentro de las regiones.

Figura 2.5. Distribución de los resultados de ciencias (percentiles) y puntuación media estimada con intervalo de confianza al 95 % para la media poblacional



2.4.2. Dominios de contenido y cognitivos en ciencias

Como se mostró en el Capítulo 1, las especificaciones de ciencias en TIMSS toman la forma de una matriz organizada sobre dos ejes: dimensiones de contenido y dimensiones cognitivas. A su vez, los contenidos se estructuran en tres dominios: ciencias de la vida, ciencias físicas y ciencias de la Tierra, mientras que la dimensión cognitiva se organiza sobre tres dominios: conocimiento, aplicación y razonamiento (Cuadro 2.3). En este apartado se mostrarán los resultados por cada uno de los dominios de contenido y cognitivos.

Cuadro 2.3. Distribución de porcentajes de los dominios de contenido y cognitivos en la evaluación de ciencias

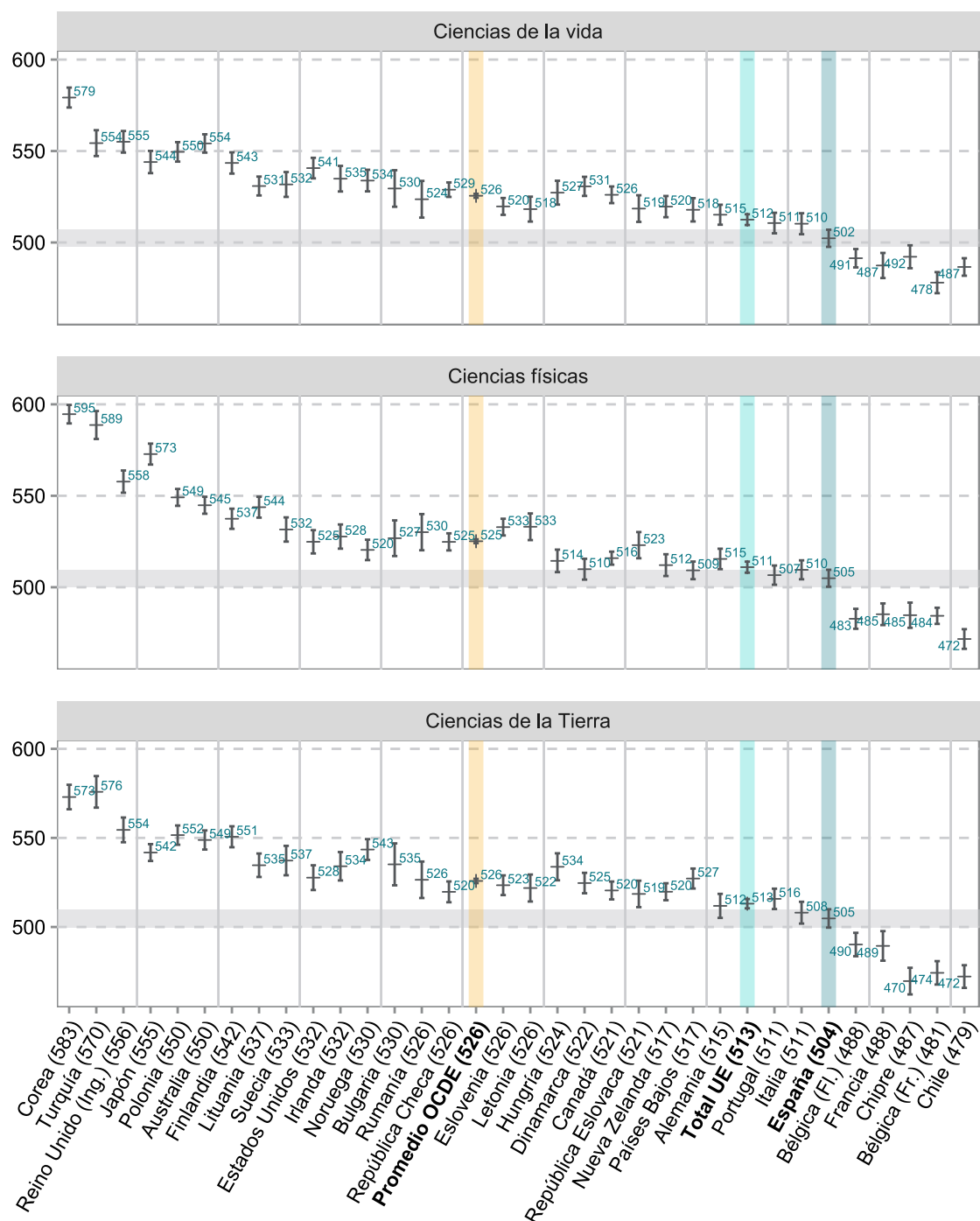
CIENCIAS	DOMINIOS DE CONTENIDO	%	DOMINIOS COGNITIVOS	%
	Ciencias de la vida	45 %	Conocimiento	40 %
	Ciencias físicas	35 %	Aplicación	40 %
	Ciencias de la Tierra	20 %	Razonamiento	20 %

2.4.2.1. Resultados en los dominios de contenido de ciencias

Las Figuras 2.6.a y 2.6.b comparan las puntuaciones numéricas en los tres dominios de contenido del área de ciencias. La primera figura contiene los resultados de España, los parámetros internacionales y los países seleccionados, mientras que, en la segunda, los países son reemplazados por las comunidades autónomas con muestras representativas. En todas las figuras las poblaciones comparadas se ordenan descendientemente por el promedio en ciencias.

En la Figura 2.6.a, la banda gris que recorre horizontalmente el gráfico señala el intervalo de confianza al 95 % para la media poblacional de ciencias de España en cada dominio de contenido. Este tipo de análisis permite detectar puntos fuertes y débiles de cada país. Por ejemplo, en Corea los resultados en ciencias físicas (595 puntos) son relativamente mejores que los obtenidos en los otros dominios. De igual modo, en Chipre la puntuación en ciencias de la Tierra (470 puntos) es relativamente peor que la mostrada en ciencias de la vida (492).

Figura 2.6.a. Puntuación media estimada con intervalo de confianza al 95 % para la media poblacional de los dominios de contenido en ciencias. Países, Total UE y Promedio OCDE

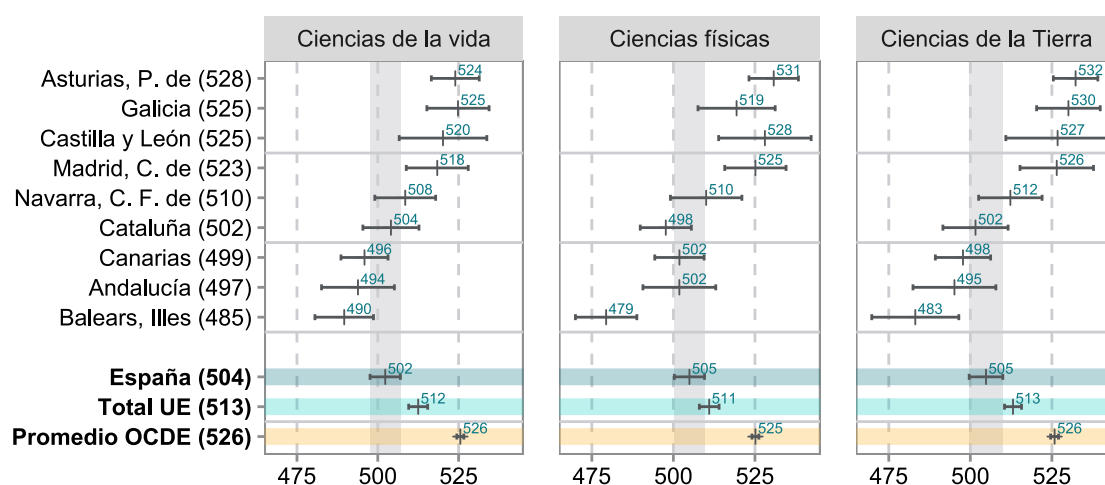


En España los resultados por dominios de contenido son muy similares: 502 en ciencias de la vida y 505 en ciencias físicas y ciencias de la Tierra. Sin embargo, dentro de algunos países las puntuaciones por dominios de contenido fluctúan considerablemente. Ello hace que la posición relativa de España no sea igual en las tres comparaciones. Así, en términos relativos el mejor resultado de España corresponde al dominio de las ciencias físicas. En dicho dominio no hay diferencias estadísticamente significativas entre la media española y la del conjunto de los socios europeos (Total UE). De hecho, en el dominio de las ciencias físicas los resultados

españoles son estadísticamente similares a los de Dinamarca, país que en la escala general de ciencias presenta resultados significativamente mejores que los de España. En ciencias de la Tierra la diferencia entre los resultados de España y Total UE está justo en la frontera de la significación estadística y, de hecho, no existen diferencias estadísticamente significativas entre Alemania (515 en la escala general de ciencias) y España (504 en la misma escala). Comparativamente hablando el dominio en el que España está peor posicionada es ciencias de la vida.

La Figura 2.6.b incluye a las comunidades autónomas con muestras comparables. En este caso la banda gris vertical señala el intervalo de confianza al 95 % para la media poblacional de España en cada uno de los dominios de contenido. Dentro de las regiones no hay grandes diferencias cuando los resultados se desglosan por dominios de contenido. Las mayores están en torno a 10 puntos y son las siguientes: Galicia con 530 puntos en ciencias de la tierra y 519 en ciencias físicas, e Illes Balears con 490 en ciencias de la vida y 479 en ciencias físicas. La comparativa de los resultados regionales por dominios de contenido permite hacer algunas agrupaciones: Principado de Asturias, Galicia, Castilla y León y Comunidad de Madrid conforman un bloque cuyos resultados en los tres dominios presentan las siguientes características en términos estadísticos: son similares al Promedio OCDE, superiores al promedio de España y, en general, mejores que el Total UE, aunque no siempre las diferencias son estadísticamente significativas. Un segundo bloque incluye a Comunidad Foral de Navarra, Cataluña, Canarias y Andalucía cuyo resultado en los tres dominios de contenido es similar al español y se encuentra por debajo del Promedio OCDE.

Figura 2.6.b. Puntuación media estimada con intervalo de confianza al 95 % para la media poblacional de los dominios de contenido en ciencias. Comunidades Autónomas, España, Total UE y Promedio OCDE



2.4.2.2. Resultados en los dominios cognitivos de ciencias

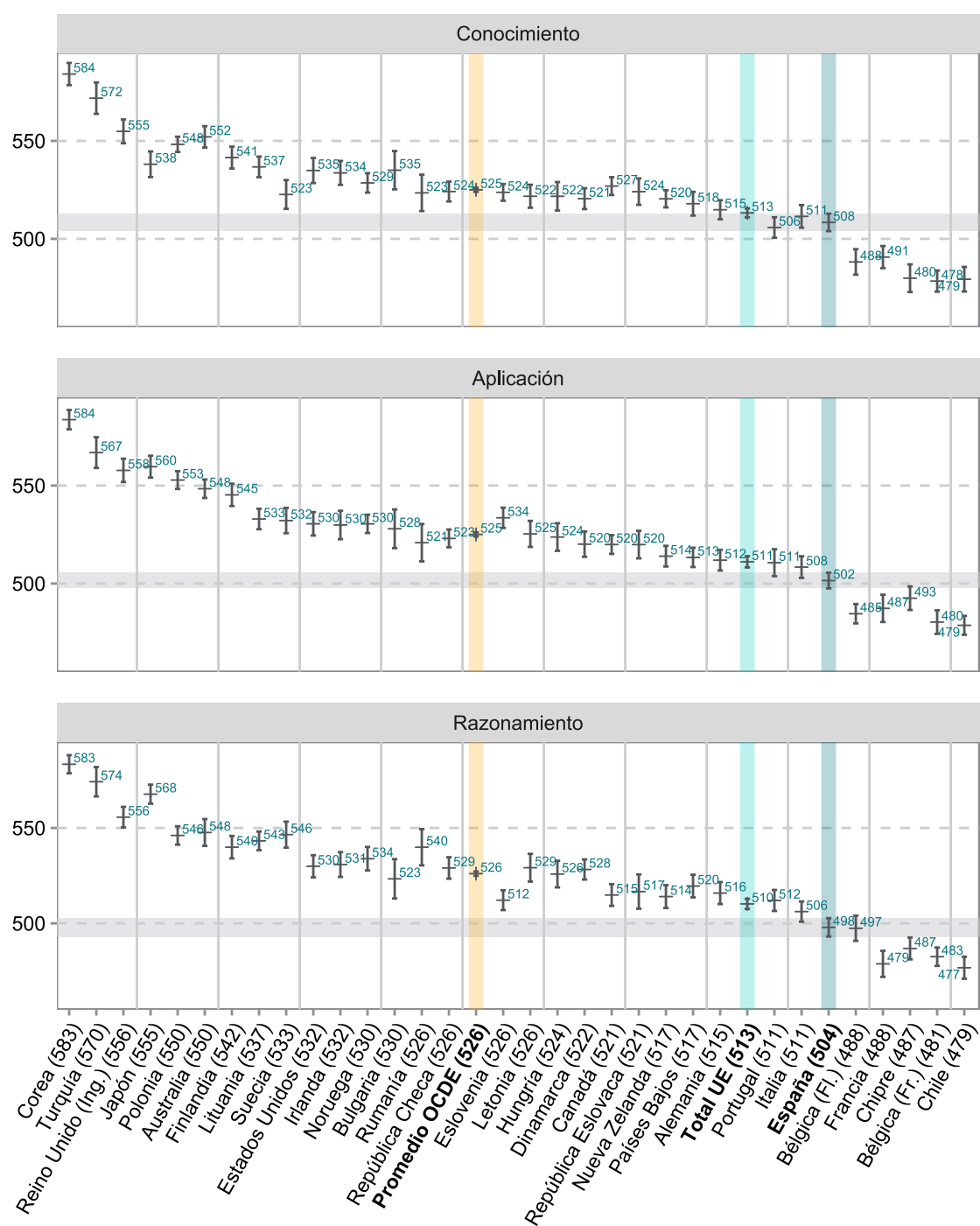
Las Figuras 2.7.a, y 2.7.b muestran las distribuciones de los resultados obtenidos en los dominios cognitivos de **conocimiento, aplicación y razonamiento**.

En la Figura 2.7.a, la banda gris que recorre horizontalmente el gráfico señala el intervalo de confianza al 95 % para la media poblacional de ciencias de España en cada uno de los dominios cognitivos. De nuevo, la comparación de las puntuaciones de los dominios cognitivos dentro de cada país permite identificar puntos fuertes y débiles de sus sistemas educativos. Por ejemplo, en Japón el promedio en los dominios razonamiento (568 puntos) y aplicación (560 puntos) es bastante superior al mostrado en el dominio conocimiento (538 puntos).

En el caso de España se advierte una diferencia de 10 puntos entre el promedio del dominio conocimiento (508 puntos) y el dominio razonamiento (498 puntos), quedando el dominio aplicación en una situación intermedia

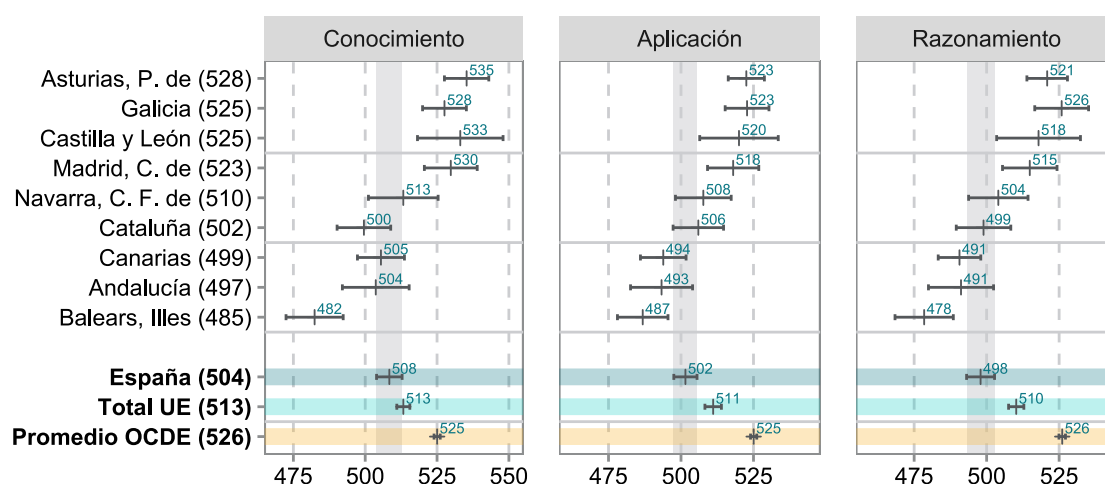
(502 puntos). Estas evidencias replican resultados previos de España, que habitualmente se encuentra mejor situada en las competencias de carácter reproductor (conocer, recordar, ...) que en las competencias que demandan el uso de razonamientos. De este modo, el promedio de España en el dominio conocimiento es, desde el punto estadístico, similar al Total UE. La peor posición relativa de España corresponde al dominio de razonamiento y, de hecho, desde el punto de vista estadístico sus resultados solo son similares a los de Bélgica (Fl).

Figura 2.7.a. Puntuación media estimada con intervalo de confianza al 95 % para la media poblacional de los dominios cognitivos en ciencias. Países, Total UE y Promedio OCDE



La Figura 2.7 b contiene las comunidades autónomas con muestras representativas y la banda gris vertical señala el intervalo de confianza al 95 % para la media poblacional de ciencias de España en cada uno de los dominios cognitivos. Los resultados regionales reproducen la pauta nacional. Así, en todas las regiones (salvo Illes Balears y Cataluña) el dominio conocimiento es el que presenta mejores resultados. Igualmente, en todas las regiones (a excepción de Galicia) el dominio cognitivo con resultados más bajos es razonamiento. Hay cinco comunidades autónomas (Andalucía, Principado de Asturias, Castilla y León, Comunidad de Madrid y Canarias) donde la diferencia entre el promedio de los dominios conocimiento y razonamiento oscila entre 10 y 15 puntos. En cuanto a la posición relativa de las regiones por dominios cognitivos, la lectura es muy similar a la señalada al hablar de los dominios de contenido. Las cinco primeras regiones de la Figura 2.7.b muestran puntuaciones similares a la OCDE y superiores al promedio de España, mientras que Illes Balears es la única región ubicada por debajo de la media española en los tres dominios cognitivos.

Figura 2.7.b. Puntuación media estimada con intervalo de confianza al 95 % para la media poblacional de los dominios cognitivos en ciencias. Comunidades Autónomas, España, Total UE y Promedio OCDE

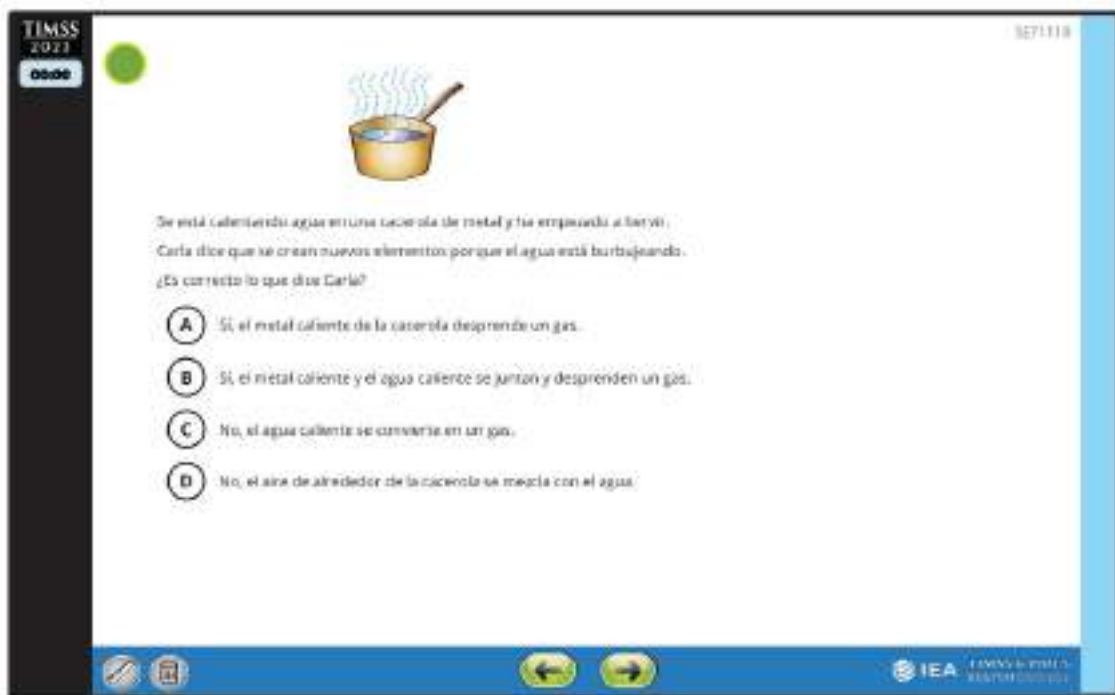


2.5. Escalas y niveles de rendimiento en ciencias

Este apartado estará dedicado a presentar los resultados del alumnado por escalas y niveles de rendimiento en ciencias. El mismo se divide en dos puntos. El primero describirá los niveles de rendimiento en la escala de ciencias de TIMSS 2023, mientras que el segundo mostrará la distribución del alumnado en los niveles de rendimiento descritos previamente.

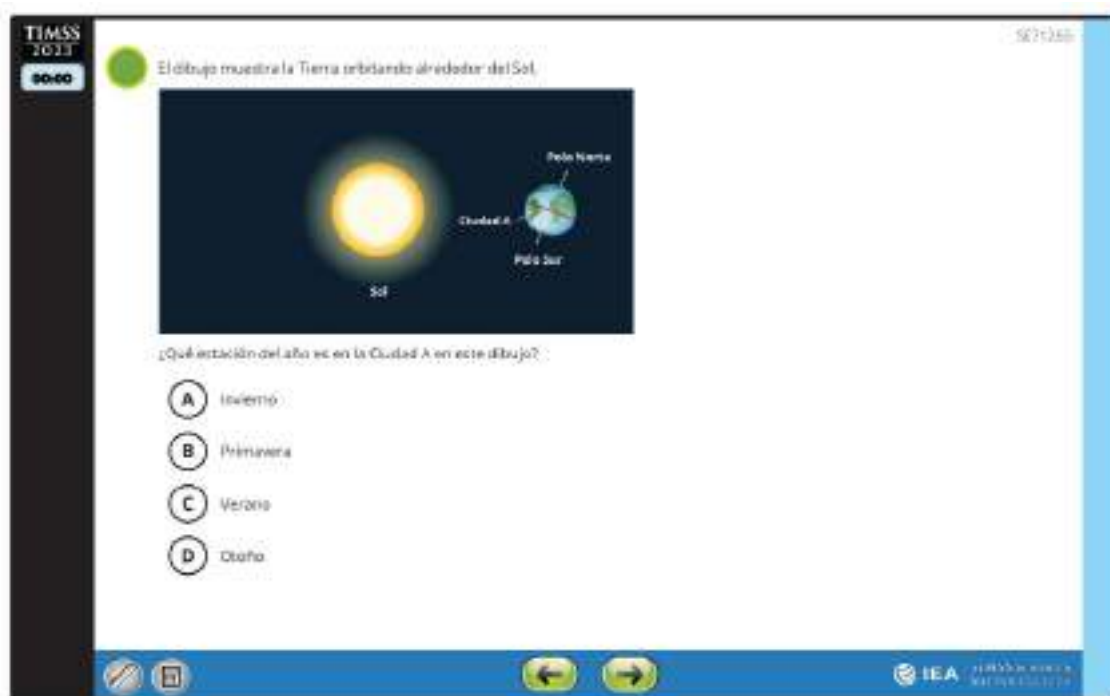
2.5.1 Descripción de los niveles de rendimiento de la escala en ciencias

IMSS clasifica el rendimiento en el área de ciencias en cuatro niveles: avanzado, alto, intermedio y bajo. A continuación, se describe lo que son capaces de hacer los estudiantes que alcanzan un determinado nivel de rendimiento en ciencias junto con el intervalo de puntos correspondiente a cada uno de los niveles y un ejemplo que ilustra la dificultad de las preguntas de cada uno de los niveles. A los cuatro niveles descritos habría que añadir un quinto nivel, que denominamos muy bajo, para aquellos estudiantes que no alcancen los 400 puntos en las pruebas de evaluación y que, por tanto, no son capaces de resolver las cuestiones relativas al nivel bajo de la escala.

Nivel avanzado: 625 puntos

El alumnado del nivel avanzado puede mostrar, aplicar y comunicar sus conocimientos sobre ciencias de la vida, ciencias físicas y ciencias de la Tierra, y participar en múltiples prácticas de investigación científica. Así mismo demuestra conocer las características de los seres vivos y puede construir y reflexionar a partir de diagramas y representaciones sobre las relaciones entre organismos en los ecosistemas. Demuestra conocimiento sobre la herencia, la eliminación de gérmenes y la contaminación ambiental; conoce las propiedades de la materia y los cambios de estado de la materia, y razona sobre las tasas de disolución en un entorno de laboratorio. El alumnado puede comunicar sus conocimientos sobre las características físicas y procesos de la Tierra, y de cómo los humanos usan e impactan los recursos naturales de la Tierra. Finalmente, demuestra conocer el movimiento y la posición relativa de la Tierra, la Luna y el Sol y puede diseñar pruebas para predecir resultados y evaluar posibles conclusiones.

Nivel alto: de 550 a menos de 625 puntos



El alumnado de este nivel conoce y aplica conocimientos sobre ciencias de la vida, ciencias físicas y ciencias de la Tierra, y participa en algunas prácticas de investigación científica. Puede distinguir entre seres vivos y no vivos, demuestra conocimientos sobre la reproducción y supervivencia de plantas y animales, y puede aplicar conocimientos sobre algunas características de las plantas y animales y sus ciclos de vida. Este alumnado también puede aplicar conocimientos sobre la propagación de gérmenes, los estados y propiedades de la materia, imanes, sonido y calor, y pueden razonar utilizando conocimientos sobre tasas de disolución en un contexto cotidiano.

Este alumnado también demuestra y puede aplicar algunos conocimientos sobre fuerzas y movimiento. Así mismo, conoce varios hechos sobre las características físicas de la Tierra y aplica sus conocimientos sobre los diferentes climas de la Tierra y los cambios a lo largo del tiempo. Finalmente puede aplicar sus conocimientos sobre el sistema Tierra-Sol y demuestra conocimientos básicos sobre las fases de la Luna. Los estudiantes describen observaciones e interpretan modelos y representaciones gráficas.

Nivel intermedio: de 475 a menos de 550 puntos

TIMSS 2023
00:00

5671005

Juan investigó el efecto de diferentes cantidades de luz en dos plantas idénticas. Puso las plantas en dos maceteros idénticos con el mismo tipo de tierra y la misma cantidad de agua.

Juan colocó la Planta A cerca de la ventana y la Planta B dentro de un armario con la puerta cerrada.

Después de dos semanas las plantas estaban así:



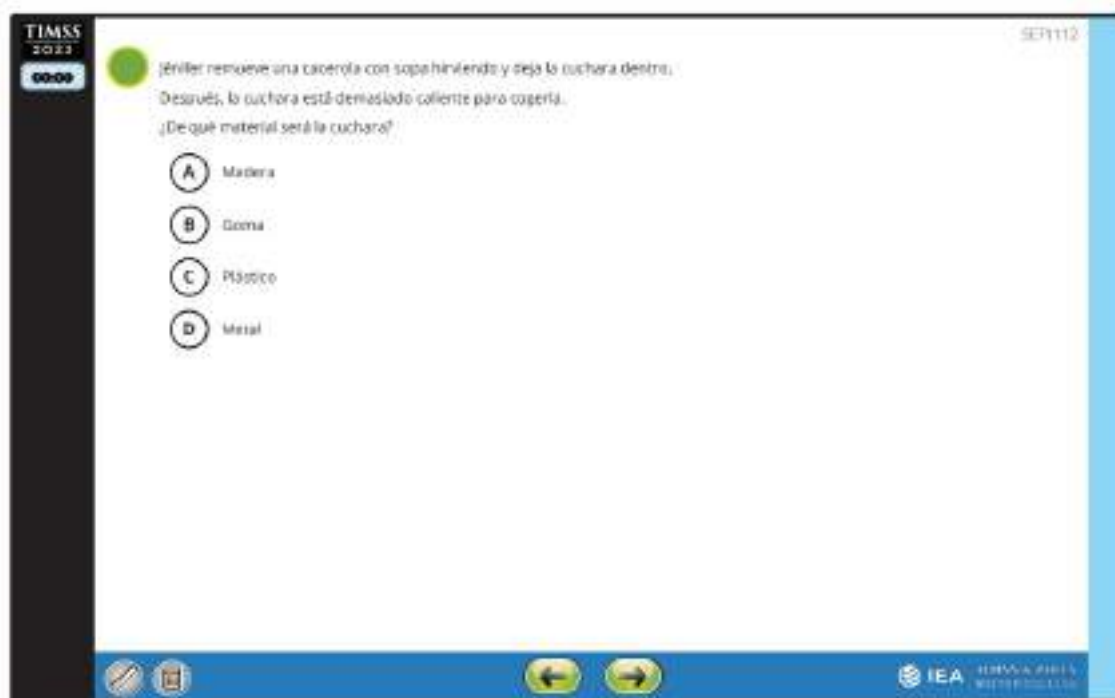
Planta A Planta B

¿Por qué la Planta B, que estuvo en el armario, está peor que la Planta A, que estuvo al lado de la ventana?

IEA TIMSS & PIRLS
INTERNATIONAL EDUCATIONAL ASSESSMENT AT TENDENCY STUDIES

El alumnado de este nivel conoce y aplica conceptos científicos sobre plantas y animales, y tiene conocimientos básicos sobre la salud humana, las propiedades de la materia, la energía y la luz y muestra una comprensión básica de la superficie de la Tierra. Este alumnado puede aplicar conocimientos básicos sobre fuerzas y movimiento, realiza descripciones parciales de observaciones y relaciona observaciones y datos con hechos científicos.

Nivel bajo: de 400 a menos de 475 puntos



Los estudiantes en este nivel muestran conocimientos básicos sobre plantas, animales y el medio ambiente; conocen algunas propiedades de la materia en situaciones cotidianas y saben que las turbinas proporcionan electricidad a algunas regiones. También conocen las características de la Tierra, su clima y sus cambios a lo largo del tiempo.

Nivel muy bajo: menos de 400 puntos

El alumnado de este nivel no es capaz de realizar las tareas del nivel bajo.

2.5.2. Distribución del alumnado por niveles de rendimiento

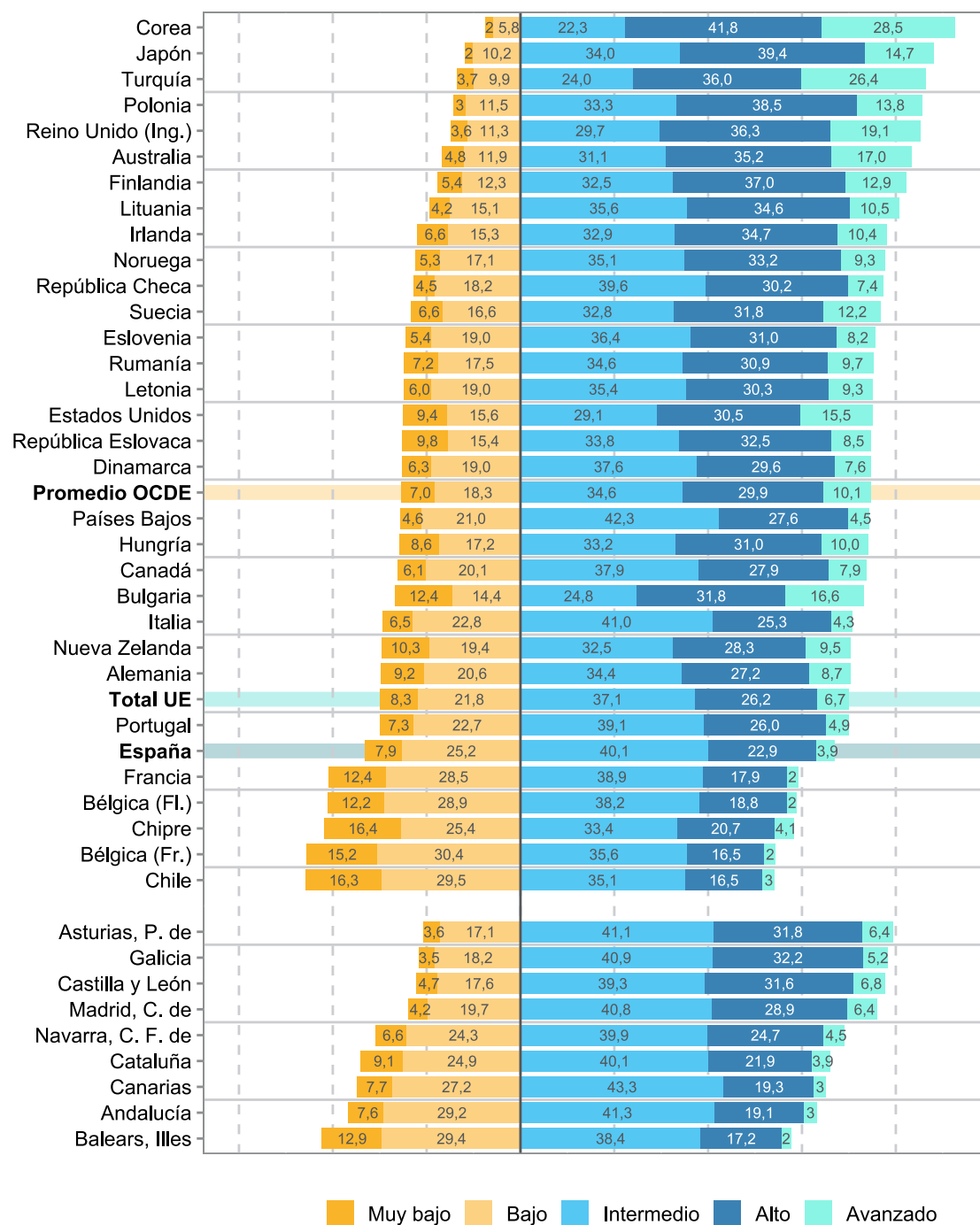
En la Figura 2.8 se representa la distribución del porcentaje de estudiantes de cada país, organismo internacional y comunidades autónomas en los cinco niveles de rendimiento de la escala de ciencias. Los países y comunidades están ordenados de menor a mayor porcentaje de estudiantes que no alcanzan el valor intermedio, es decir, proporción de estudiantes en los niveles de rendimiento bajo o muy bajo.

En el conjunto de la OCDE, 1 de cada 4 estudiantes no alcanza el nivel intermedio (7,0 % nivel muy bajo; 18,3 % nivel bajo). Por su parte, en el Total UE la proporción de estudiantes en los niveles de rendimiento inferiores supera el 30 % (8,3 % muy bajo; 21,8 % bajo). Como ya ocurría en matemáticas, dentro de la OCDE, los países asiáticos son los que presentan menor proporción de estudiantes en los niveles de rendimiento bajo o muy bajo. En Corea menos del 8 % y en Japón algo más del 12 %. En el caso coreano, los buenos resultados son ratificados por el hecho de ser el país con mayor proporción de estudiantes en el nivel avanzado, casi el 30 %. Dentro de la UE los países con menor proporción de estudiantes en los dos niveles inferiores de rendimiento son Polonia (14,5 %) y Finlandia (17,7 %) (Figura 2.8).

En relación con los niveles Promedio de OCDE y Total UE, España presenta una proporción de estudiantes notablemente alta en los niveles bajos de la escala de ciencias (33,1 %): 7,9 % en el nivel muy bajo y 25,2 % en el nivel bajo. Únicamente Chile, Bélgica (Fr.), Chipre, Bélgica (Fl.) y Francia obtienen porcentajes más altos en estos niveles de rendimiento (Figura 2.8). Por otra parte, cabe indicar que el porcentaje de estudiantes con rendimiento avanzado es pequeño (3,9 %), casi 3 puntos porcentuales menos que el Total UE y algo más de 6 menos que el Promedio de OCDE.

En cuanto a los datos de las nueve comunidades autónomas que han participado con muestra ampliada en el estudio, cinco de ellas presentan niveles menores de rendimiento bajo o muy bajo respecto de la media de España y cuatro niveles mayores. En el lado positivo, están los datos del Principado de Asturias (20,7 %: 3,6 % muy bajo; 17,1 % bajo), Galicia (21,7 %: 3,5 % muy bajo; 18,2 % bajo), Castilla y León (22,3 %: 4,7 % muy bajo; 17,6 % bajo), Comunidad de Madrid (23,9 %: 4,2 % muy bajo; 19,7 % bajo) y Comunidad Foral de Navarra (30,9 %: 6,6 % muy bajo; 24,3 % bajo). En el polo negativo se encuentran Cataluña (34,0 %), Canarias (34,9 %), Andalucía (36,8 %) y, especialmente, Illes Balears (42,3 %), con más del 40 % de estudiantes con rendimiento bajo o muy bajo. En cuanto al rendimiento intermedio, alto o avanzado, se observan datos similares (con escasas diferencias entre comunidades) en cuanto al rendimiento intermedio. En consecuencia, las diferencias se encuentran en el porcentaje de alumnos con un rendimiento alto y/o avanzado: mientras que el Principado de Asturias, Galicia, Castilla y León y Madrid superan el 35 % de alumnos con rendimiento alto o avanzado, Canarias, Andalucía o Illes Balears alcanzan escasamente el 20 %.

Figura 2.8. Porcentajes de estudiantes por niveles de rendimiento en ciencias. Orden de menor a mayor en los niveles bajo y muy bajo



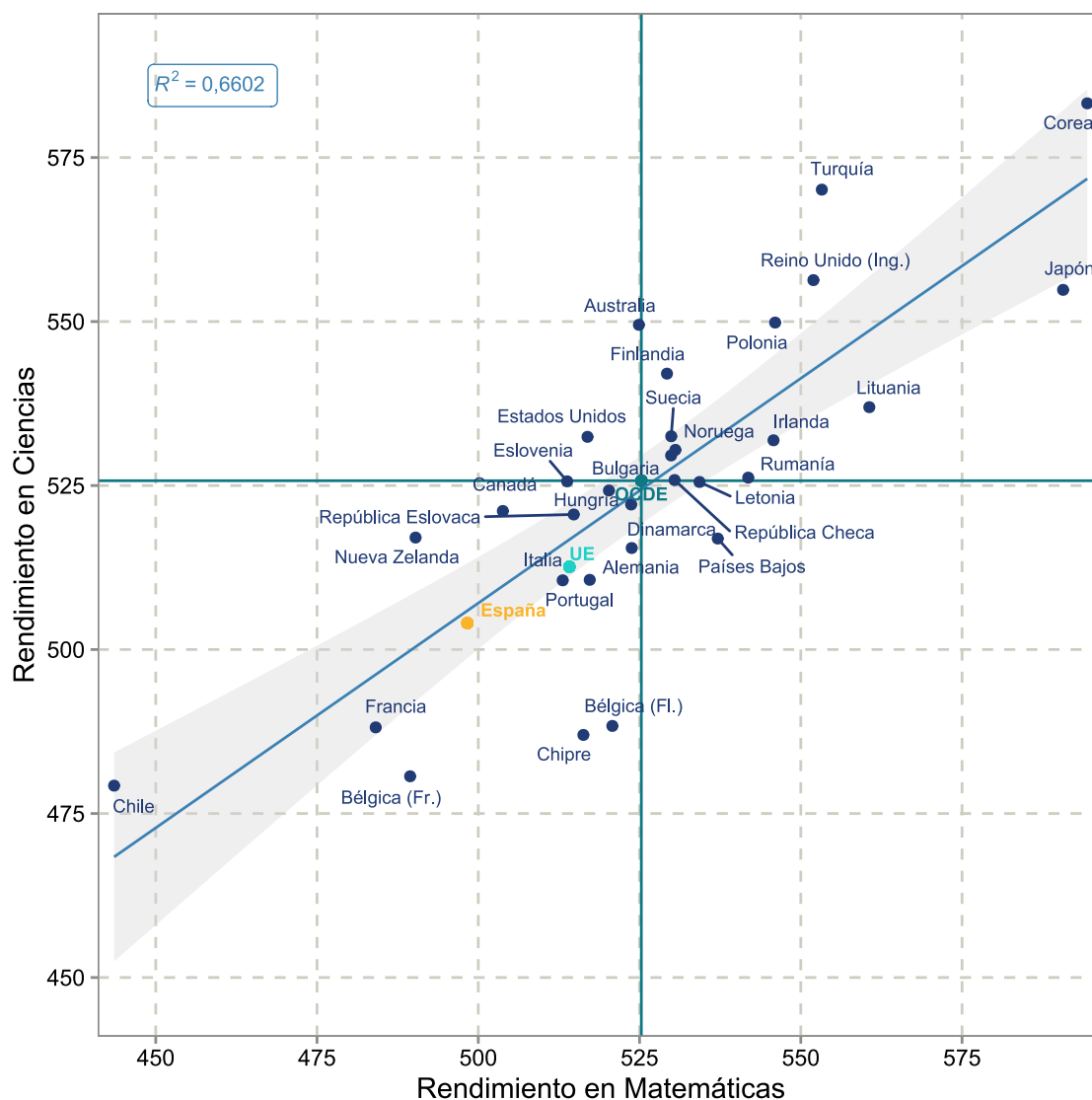
2.6. Relación entre los resultados de matemáticas y ciencias

2.6.1. Relación entre las puntuaciones medias de ambas competencias

La Figura 2.9.a es un gráfico de dispersión, donde el eje horizontal representa los valores de la escala de puntuaciones en matemáticas y el eje vertical las puntuaciones en ciencias. El plano cartesiano también contiene dos líneas perpendiculares en torno a 525 puntos que coinciden con la media OCDE en las dos materias y que dividen el plano en cuatro cuadrantes. Los puntos del plano corresponden al par de puntuaciones en matemáticas y ciencias de cada país. De este modo los puntos ubicados en el cuadrante superior derecho señalan aquellos países cuyo promedio en ambas materias supera la media OCDE, mientras que los puntos situados en el cuadrante inferior izquierdo señalan el caso contrario.

En términos generales la Figura 2.9.a indica que existe una asociación fuerte y positiva entre las puntuaciones en ciencias y matemáticas de los países seleccionados. Gráficamente esta idea está representada por la línea de tendencia que atraviesa diagonalmente el gráfico. El trazado de esta diagonal, que se eleva sobre la vertical a medida que se desplaza hacia la derecha del eje horizontal, indica que los países que obtienen altas puntuaciones en matemáticas tienden a presentar buenos resultados en ciencias (y viceversa). El valor del coeficiente de determinación (R^2) del gráfico (0,6602), que depende del grado de inclinación de la línea de regresión, cuantifica la fuerza de dicha relación. Este valor se interpreta del siguiente modo: aproximadamente el 66% de las diferencias en las puntuaciones de los países en ciencias viene explicada por los resultados de matemáticas (y viceversa). Este valor de R^2 puede transformarse en un coeficiente de correlación de Pearson, cuya magnitud en este caso sería 0,81. Ello confirma la consistencia de la relación entre las puntuaciones de los países en matemáticas y ciencias.

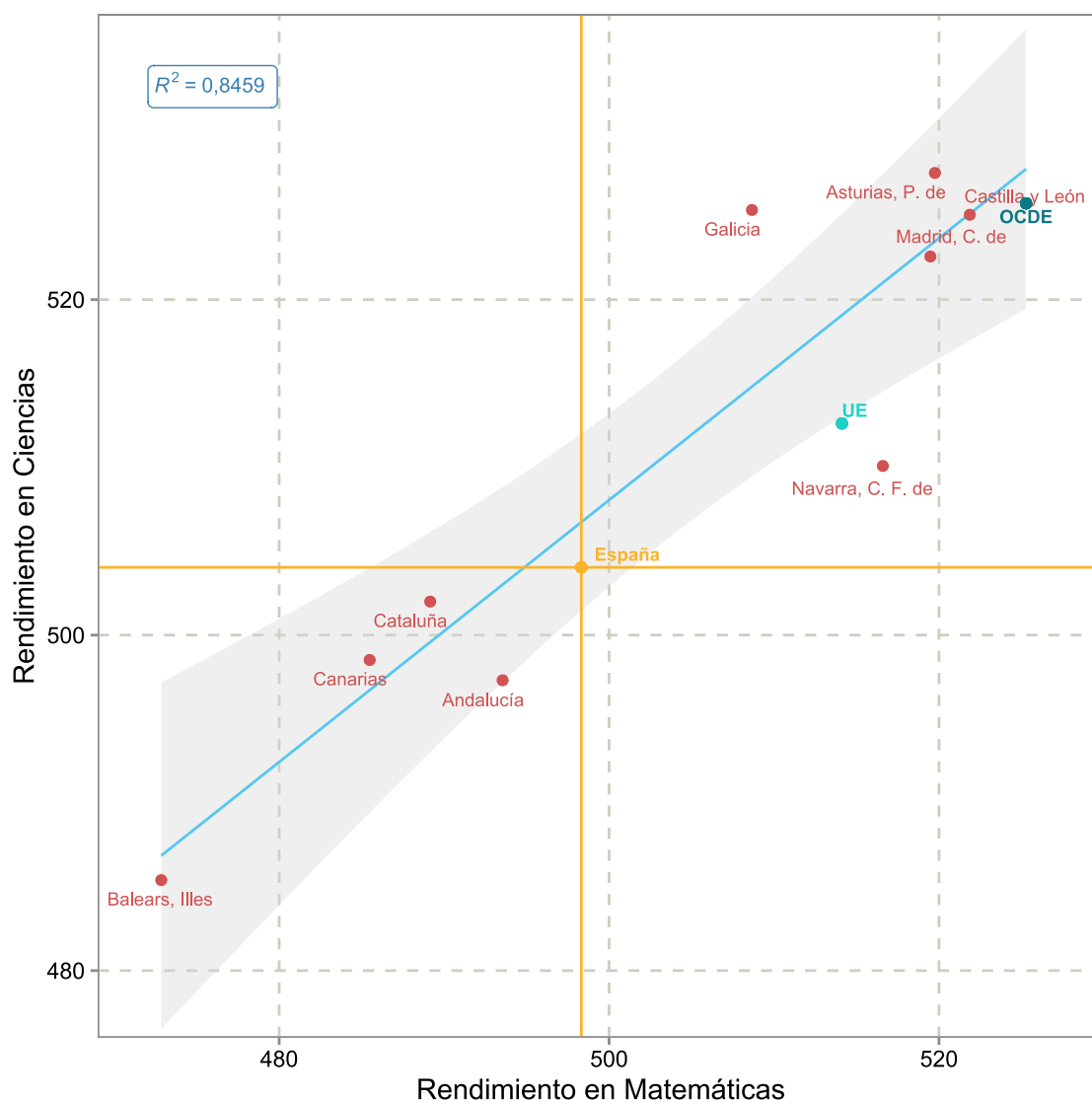
Figura 2.9.a. Relación a nivel de sistemas educativos entre las puntuaciones medias estimadas de ciencias y matemáticas. Países, Total UE y Promedio OCDE



Finalmente, cabe señalar que, en la Figura 2.9.a, la banda o intervalo de confianza asociada a la recta de regresión de las puntuaciones medias de ciencias sobre las de matemáticas representa la puntuación media en ciencias esperada en función de la estimada en matemáticas, con un 95 % de confianza. Los países cuyos puntos se sitúan por encima del intervalo de confianza son aquellos cuyo resultado en ciencias está por encima de lo esperado teniendo en cuenta su resultado en matemáticas. Tal sería en el caso de Australia, cuyo desempeño en ciencias (en torno a 550 puntos) está muy por encima del que sería esperado por su media en matemáticas (en torno a 525 puntos). Reino Unido (Ing.), Finlandia o Nueva Zelanda, entre otros, son ejemplos de países cuyo resultado en ciencias es comparativamente mejor que su resultado en matemáticas. Por el contrario, los puntos situados por debajo de la banda o intervalo de confianza señalan aquellos países cuyo resultado en ciencias es más bajo del esperado por su media en matemáticas. Este sería el caso de las dos comunidades belgas, Países Bajos y Chipre o Lituania y, en menor medida, Japón.

El punto que representa a España se sitúa dentro de la franja del intervalo de confianza y, de hecho, prácticamente roza la recta de regresión. Ello indica que la puntuación de España en ciencias es muy similar a la que cabría esperar por el desempeño en matemáticas del alumnado español.

Figura 2.9.b. Relación a nivel de sistemas educativos entre las puntuaciones medias estimadas de ciencias y matemáticas. CCAA, España, Total UE y Promedio OCDE



La Figura 2.9.b replica el análisis anterior, aunque considerando las nueve comunidades autónomas con datos representativos y los parámetros internacionales. En este caso, las líneas perpendiculares que delimitan los cuadrantes coinciden con la media de España en matemáticas y ciencias. La interpretación es similar a la descrita en el gráfico anterior: los promedios de las regiones en las dos materias serán mejores cuando más arriba y más a la derecha se sitúen los puntos y peores en la medida en que los puntos se sitúen más abajo y más a la izquierda. Por su parte, el trazado de la diagonal indica que las puntuaciones en ambas materias están fuerte y positivamente asociadas. En el caso de las regiones españolas la fuerza de la asociación es muy elevada: el valor de R^2 (0,8459) indica que aproximadamente el 85 % de las diferencias en las puntuaciones

de las comunidades autónomas en ciencias se explica por los resultados de matemáticas (y viceversa), lo que supone un coeficiente de correlación de 0,92.

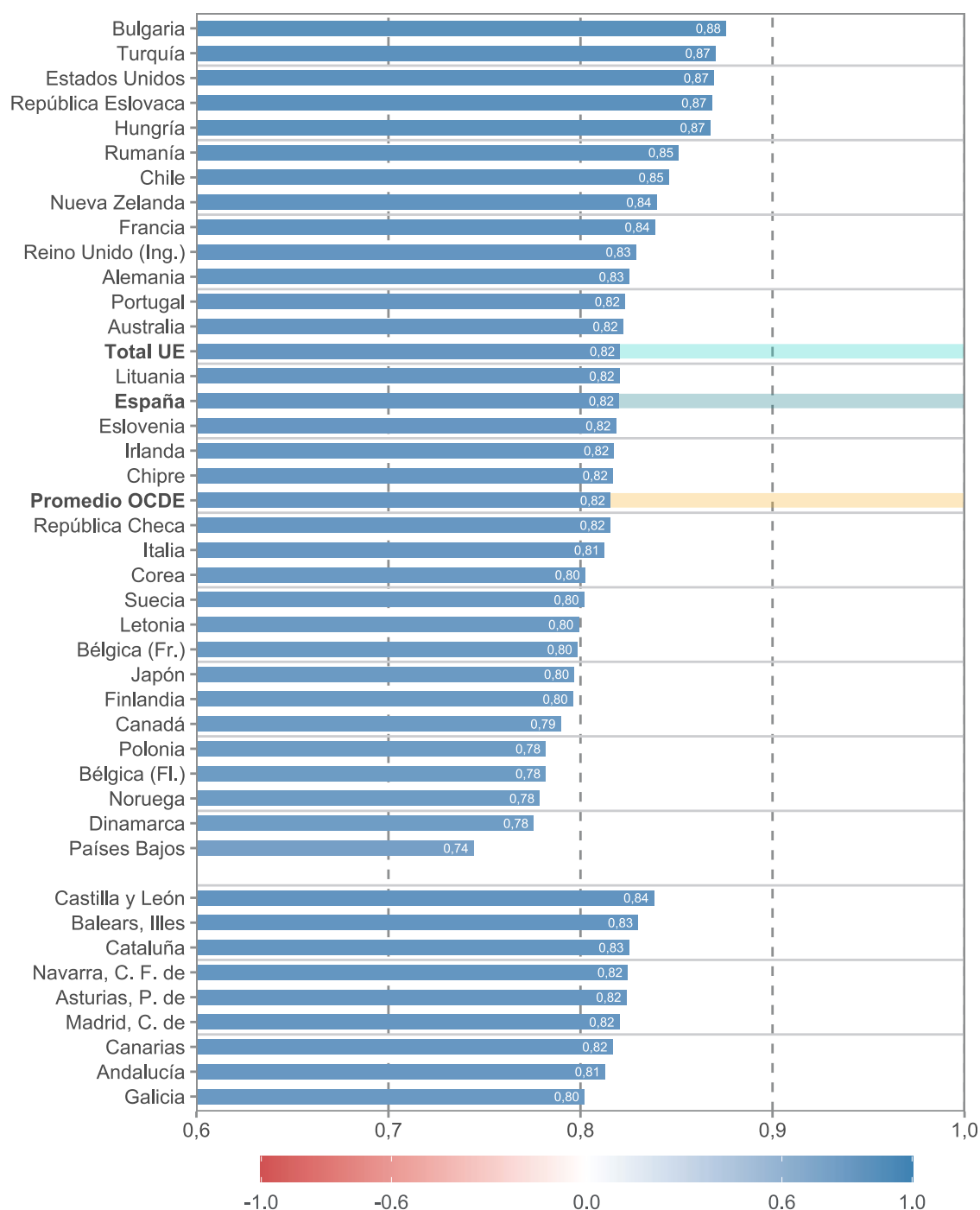
Finalmente se observa que la mayoría de los puntos que representan a las comunidades autónomas se ubican dentro de la franja del intervalo de confianza, indicando que los resultados en ciencias son similares a los esperados por las puntuaciones de cada región en matemáticas. La anterior afirmación presenta dos excepciones: Galicia, cuyo resultado en ciencias es comparativamente mejor que el obtenido en matemáticas, y Comunidad Foral de Navarra en la situación opuesta, es decir, con una puntuación comparativamente mejor en matemáticas que en ciencias (Figura 2.9.b).

2.6.2. Correlación entre resultados

En los gráficos anteriores se mostró que, cuando los datos se agrupan por países o comunidades autónomas, la relación entre los resultados de matemáticas y ciencias es fuerte y positiva. Este mismo análisis se ha realizado dentro de cada país o región española, es decir, con los datos considerados a nivel de alumnado. En este caso, las unidades a correlacionar serían las puntuaciones estimadas para cada estudiante del país o comunidad autónoma considerada. Como era esperable, los resultados obtenidos por el alumnado en ambas áreas también presentan una asociación fuerte y positiva. La Figura 2.10 resume esta idea, puesto que las barras representan la magnitud del coeficiente de correlación de Pearson entre las puntuaciones en matemáticas y ciencias en cada país y región.

En el promedio de países OCDE y en el Total UE, el valor del coeficiente de correlación está ligeramente por encima de 0,80 puntos, indicando que algo más del 64 % de la variabilidad observada en los resultados de ciencias es explicada por los resultados de matemáticas (y recíprocamente). La mayoría de los países seleccionados presentan un coeficiente de correlación de 0,80 o superior, incluyendo a España ($r = 0,82$). Se observa también que en 7 países el coeficiente de correlación es de al menos 0,85, significando que más del 72 % de la variabilidad de los resultados de ciencias en esos países es explicada por el rendimiento en matemáticas de sus estudiantes.

Figura 2.10. Correlación entre los resultados de ciencias y matemáticas



En el caso de las comunidades autónomas con muestras representativas, los valores del coeficiente de correlación oscilan entre 0,80 y 0,84, es decir, valores similares a los parámetros internacionales. Todo ello confirma que, dentro de cada país o región los estudiantes que obtienen buena puntuación en matemáticas tienden a presentar también buen resultado en ciencias (Figura 2.10).

En general, la magnitud de los coeficientes de correlación en TIMSS 2023 es muy similar a la estimada en TIMSS 2019. La similitud de estos datos a lo largo de las diferentes ediciones del estudio confirma la consistencia y estabilidad de este tipo de estimadores en las evaluaciones internacionales (Fernández-Alonso y Muñiz, 2022).

2.7. Evolución del rendimiento

España arrancó su participación en el estudio TIMSS con muestras de 4.º de Educación Primaria en el año 2011. Ello permite disponer ahora de datos comparables de cuatro ediciones de estudio (TIMSS 2011, 2015, 2019 y 2023) que, en conjunto, abarcan más de una década de tendencias de rendimiento. En el siguiente subapartado se compara la evolución de los resultados en matemáticas y ciencias de España y del promedio de la OCDE en las ediciones mencionadas. En un subapartado posterior se analizará la tendencia en los porcentajes de alumnado ubicado en los niveles bajos (muy bajo y bajo) y altos (alto o avanzado) de las escalas de rendimiento de matemáticas y ciencias.

2.7.1. Evolución del rendimiento en matemáticas y ciencias

La Figura 2.11 compara la tendencia de resultados en matemáticas y ciencias de España y del Promedio OCDE. En matemáticas el parámetro internacional se ha mantenido prácticamente estable en las diferentes ediciones. El Promedio OCDE ha oscilado dentro de una pequeña franja de 5 puntos (entre 522 en TIMSS 2011 y 527 en TIMSS 2019).

Por su parte, el promedio español en matemáticas muestra mayor variación entre ediciones. España presentó sus peores resultados en el año 2011 y, de hecho, la distancia con el parámetro internacional aquel año fue de 40 puntos. Cuatro años más tarde (TIMSS 2015), el promedio español en matemáticas experimentó una subida estadísticamente significativa que redujo a la mitad la diferencia de puntuación con respecto al parámetro internacional. A partir de 2015 la media española ha experimentado un ligero descenso de 7. Ello ha hecho que, en las dos últimas ediciones la distancia con respecto al promedio de la OCDE se haya estabilizado en torno a un cuarto de desviación típica, esto es, aproximadamente 25 puntos de diferencia (Figura 2.11).

Figura 2.11. Evolución del rendimiento en ciencias y matemáticas



En el caso de ciencias el parámetro internacional también describe una línea prácticamente plana: en las cuatro ediciones el promedio oscila en un rango de apenas 5 puntos. Por su parte, la tendencia de España en ciencias presenta algunas fluctuaciones que merecen ser comentadas. En el año 2011 la diferencia entre el promedio de

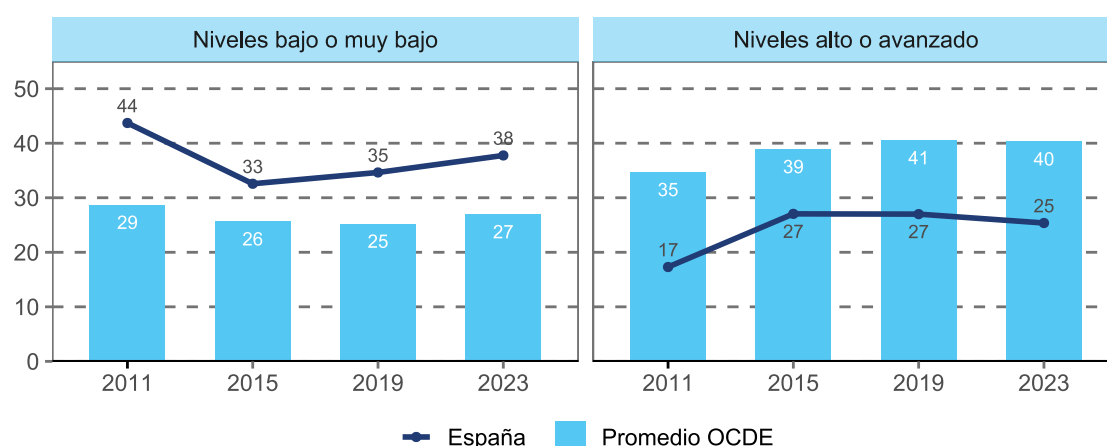
España y el parámetro internacional estaba en torno un 20 % en términos de desviación típica, es decir, algo menos de 20 puntos en la escala TIMSS. En 2015, que fue la mejor edición en términos de puntuaciones promedio para España, esa diferencia se redujo prácticamente a la mitad. Desde entonces los resultados españoles han sufrido una corrección a la baja estadísticamente significativa, retrocediendo a los niveles de 2011 y devolviendo la diferencia con respecto al Promedio OCDE al rango de los 20 puntos en la escala TIMSS (Figura 2.11).

2.7.2. Evolución del porcentaje en los niveles bajos y altos de la escala de matemáticas y ciencias

Una forma alternativa de estudiar la tendencia de los resultados es atender a los porcentajes de estudiantes ubicados en los niveles bajos (muy bajo o bajo) y en los niveles altos (alto o avanzado) de las correspondientes escalas de matemáticas y ciencias.

La Figura 2.12 compara la proporción de estudiantes en los niveles bajo y alto de matemáticas en España y en el Promedio OCDE. En lógica relación con los datos presentados en el subapartado anterior, los porcentajes para el conjunto de la OCDE son bastante estables. Así, a lo largo de la serie histórica el porcentaje de alumnado en los niveles bajos de desempeño ha oscilado entre 29 y 25 %, mientras que en los niveles altos se ha estabilizado en torno al 40 % a partir de 2015.

Figura 2.12. Evolución del porcentaje de estudiantes en los niveles bajos y altos de matemáticas



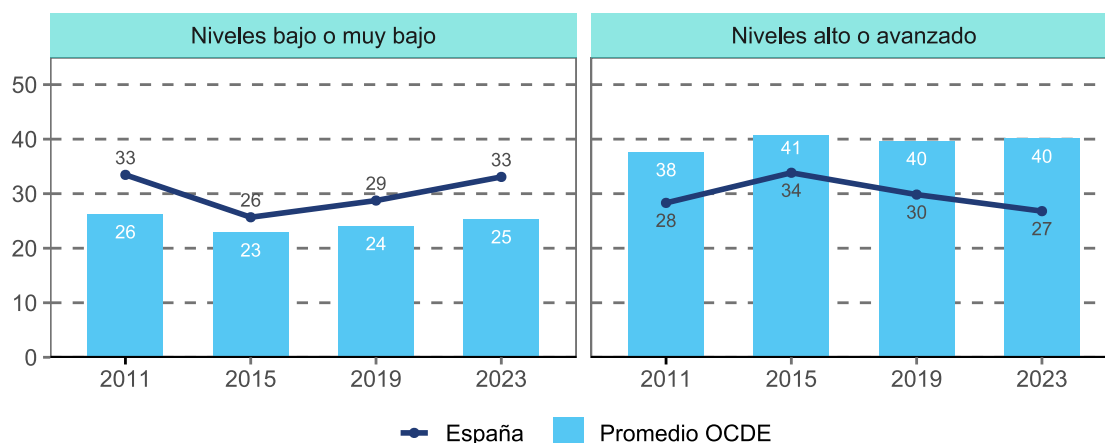
En España la variación de las proporciones es mayor y las líneas de tendencia de los grupos bajo y alto reflejan la subida que experimentaron los resultados nacionales entre las evaluaciones de 2011 y 2015. Así, entre TMSS 2011 y TMSS 2015 se aprecia un fuerte descenso (11 %) del alumnado ubicado en los niveles más bajos de desempeño. Desde entonces, dicha proporción ha subido 5 puntos porcentuales, acercándose al 40 %. En los niveles altos la tendencia española señala que entre 2011 y 2015 se produjo una subida de 10 puntos porcentuales y, desde entonces, este porcentaje se estabilizó en torno al 25 % (Figura 2.12).

La lectura global de estas tendencias, en especial a partir de 2015, permite afirmar que, en un aula promedio de España compuesta por 24 estudiantes cabría esperar que entre 8 y 9 se encuentren en los niveles bajos de desempeño matemático y otros 6 alcancen los niveles altos. Por su parte, si se tomara un aula promedio de la OCDE del mismo tamaño, lo esperable sería encontrar 6 estudiantes en los niveles bajos de matemáticas y entre 9 y 10 estudiantes en los niveles altos.

La Figura 2.13 compara la proporción de estudiantes en los niveles bajo y alto de ciencias en España y en el Promedio OCDE. En el conjunto de los países participantes de la OCDE las tendencias son bastante estables.

A lo largo de la serie histórica la proporción de alumnado en los niveles bajos ha oscilado en torno al 25 % y en torno al 40 % en los niveles altos.

Figura 2.13. Evolución del porcentaje de estudiantes en los niveles bajos y altos de ciencias



En España las variaciones son ligeramente mayores y reflejan el hecho de que en TIMSS 2023 el promedio nacional en ciencias ha vuelto a los niveles de TIMSS 2011, perdiendo la ganancia experimentada entre 2011 y 2015. Así el descenso de 7 puntos porcentuales en el alumnado de bajo rendimiento entre 2011 y 2015 se ha recuperado en las dos ediciones posteriores, por lo que 1 de cada 3 estudiantes españoles se ubica en el nivel bajo de desempeño en ciencias. En el caso de los niveles altos la lectura es similar, aunque en este caso se trata de una bajada de la proporción, desde el 34 % en 2015 hasta el 27 % en 2023 (Figura 2.13).

Los datos evolutivos permiten estimar que, en un aula promedio de España compuesta por 24 estudiantes, cabría esperar que entre 7 y 8 se encuentren en los niveles bajos de desempeño en ciencias y otros 6 alcancen los niveles altos. Por su parte, si se tomara un aula promedio de la OCDE del mismo tamaño, lo esperable sería encontrar 6 estudiantes en los niveles bajos de ciencias y entre 9 y 10 estudiantes en los niveles altos.

2.8. Rendimiento en conciencia medioambiental

Los resultados en conciencia medioambiental se organizan en dos apartados. En el primero se comparará el rendimiento promedio de los países seleccionados y de las comunidades autónomas con muestra suficiente para permitir la comparación con los parámetros internacionales (Promedio OCDE y Total UE). El segundo estará dedicado a presentar los resultados por escalas y niveles de rendimiento.

2.8.1 Rendimiento promedio en conciencia medioambiental

La Figura 2.14 muestra un gráfico de barras flotantes con los siguientes resultados en conciencia medioambiental: media estimada y su intervalo de confianza al 95 % para la media poblacional, así como las puntuaciones correspondientes con los cuartiles y los percentiles 9 y 95. Como en el resto del capítulo, los países y parámetros internacionales (Promedio OCDE y el Total UE) están ordenados descendientemente por su puntuación media en la competencia, y lo mismo ocurre con las comunidades autónomas en la parte baja de la figura.

De nuevo, el alumnado coreano alcanza la mayor puntuación media (575 puntos), que es significativamente más alta que la de cualquiera de los países seleccionados. Por detrás de Corea se encuentran Polonia y Reino Unido (Ingl.), que obtienen 557 puntos y exhiben los mejores resultados entre los países europeos. La media

de estos tres países, junto con Turquía (553 puntos) y Australia (550), se encuentra en el nivel de rendimiento alto en la escala de conciencia medioambiental. En el extremo contrario se encuentran Francia, Bélgica (Fl.), Chile, Bélgica (Fr.) y Chipre, todos ellos por debajo de 500 puntos. De hecho, en los dos últimos casos, los resultados están muy cerca del límite del nivel de rendimiento bajo. La puntuación promedio de España (510 puntos) es similar a la de países como Alemania o República Eslovaca. El resultado de España está ligeramente por debajo del Total UE (515 puntos) y es significativamente inferior al promedio de la OCDE (525 puntos) (ver Figura 2.14).

Por su parte, la situación de las nueve comunidades autónomas con datos comparables presenta cierta variabilidad. Se aprecia que hay cuatro regiones con datos por encima de los parámetros internacionales. Galicia (533) y Principado de Asturias (532) superan los 530 puntos. En el extremo contrario, los promedios de Andalucía y las dos comunidades insulares están en torno a 500 puntos.

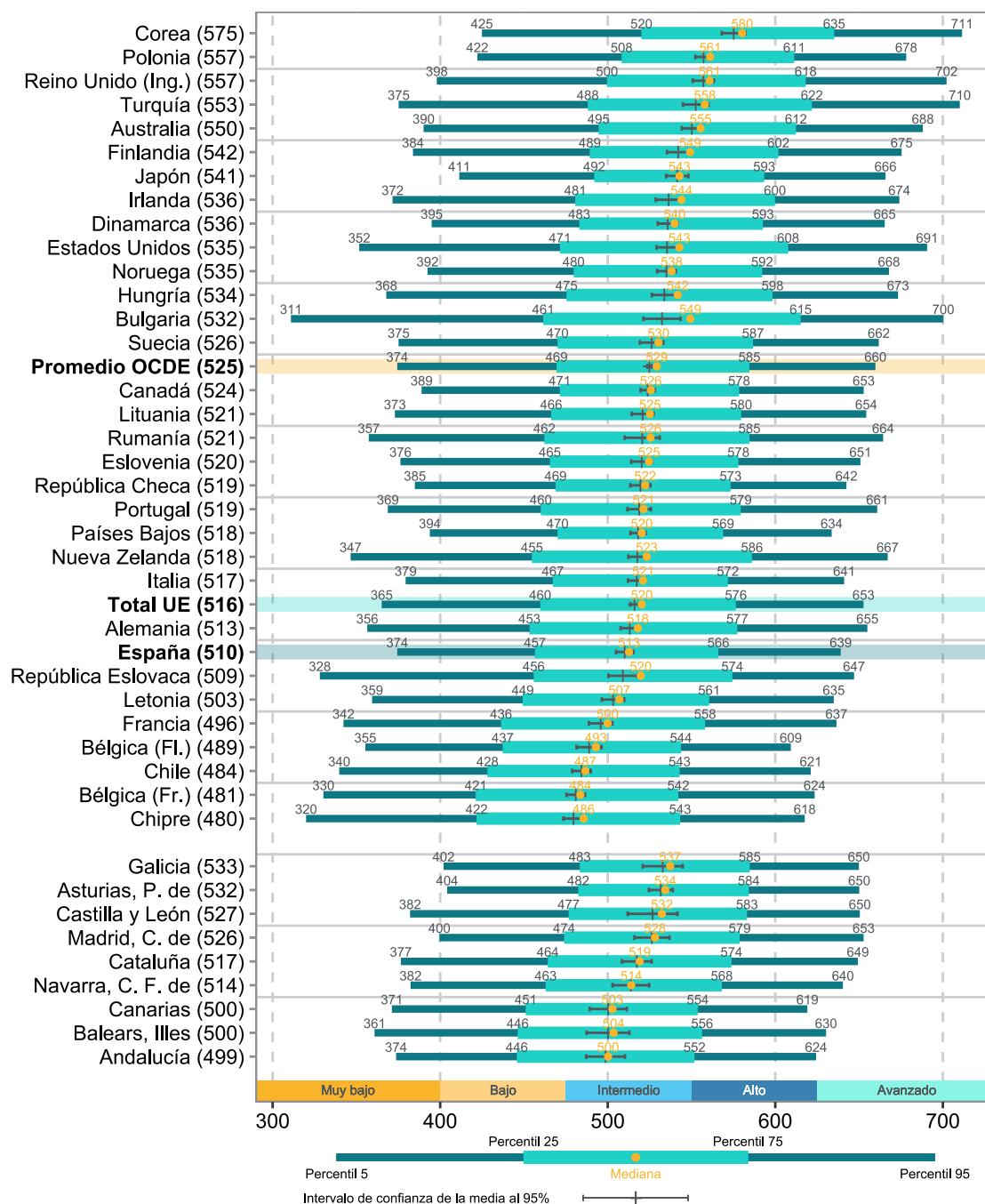
Como ya se mencionó, la Figura 2.14 muestra también los cuartiles (percentiles 25, 50 y 75) y los percentiles 5 y 95 de la distribución de las puntuaciones en conciencia medioambiental. Estos valores ofrecen información sobre la variabilidad de las puntuaciones dentro de cada país y comunidad autónoma. De este modo la diferencia entre las puntuaciones correspondientes a los percentiles 75 y 25 permite estimar el rango de puntuaciones entre las que se encuentra el 50 % central del alumnado. De igual manera, los percentiles 5 y 95 descartan las puntuaciones más extremas (ya sean más altas o bajas), y de este modo se puede estimar el rango que separa al 90 % central de las puntuaciones.

Atendiendo a los parámetros internacionales, se observa que en la OCDE y en la UE los mejores y peores estudiantes (descontando los extremos) están separados por 286 y 288 puntos, respectivamente. Se trata de una distancia importante, de casi 3 desviaciones típicas en la escala TIMSS.

Bulgaria es el país de entre los seleccionados donde la diferencia entre las mejores y peores puntuaciones es más elevada (389 puntos), lo que supone una distancia de casi 4 desviaciones típicas en el rendimiento en conciencia medioambiental. En el extremo contrario, Países Bajos presenta los resultados más homogéneos, ya que la diferencia entre los percentiles extremos es de 240 puntos, es decir, prácticamente 2,5 desviaciones típicas en la escala TIMSS.

España presenta una distribución de resultados más homogénea que los parámetros internacionales. La diferencia entre los percentiles 95 y 5 es de 265 puntos. En las comunidades autónomas los rangos de diferencias entre los estudiantes con mejores y peores resultados son, por lo general, pequeños, y oscilan con puntuaciones entre 246 puntos del Principado de Asturias y 258 puntos de la Comunidad Foral de Navarra. La anterior afirmación presenta dos excepciones: Illes Balears, cuyo rango es de 269 puntos, y Cataluña, con 272 puntos, donde los rangos son algo mayores y, en todo caso, similares al parámetro nacional. En definitiva, el análisis de las diferencias de puntuación entre los desempeños más extremos en conciencia medioambiental confirma que el sistema educativo español presenta unos indicadores de homogeneidad relativamente satisfactorios, ya que las diferencias entre el alumnado son relativamente más pequeñas que las observadas en el conjunto de los países seleccionados.

Figura 2.14. Distribución de los resultados en conciencia medioambiental (percentiles) y puntuación media estimada con intervalo de confianza al 95% para la media poblacional



2.8.2. Escalas y niveles de rendimiento en conciencia medioambiental

Este apartado estará dedicado a analizar la distribución del alumnado en los niveles de rendimiento de la competencia medioambiental.

La Figura 2.15 recoge la distribución del porcentaje de estudiantes de cada país, organismo internacional y comunidades autónomas en los cinco niveles de la escala conciencia medioambiental. Tanto los países, por un lado, como las comunidades autónomas, por otro, se han ordenado de menor a mayor porcentaje de estudiantes que no alcanzan el valor intermedio, es decir, que están en los niveles bajo o muy bajo de la escala.

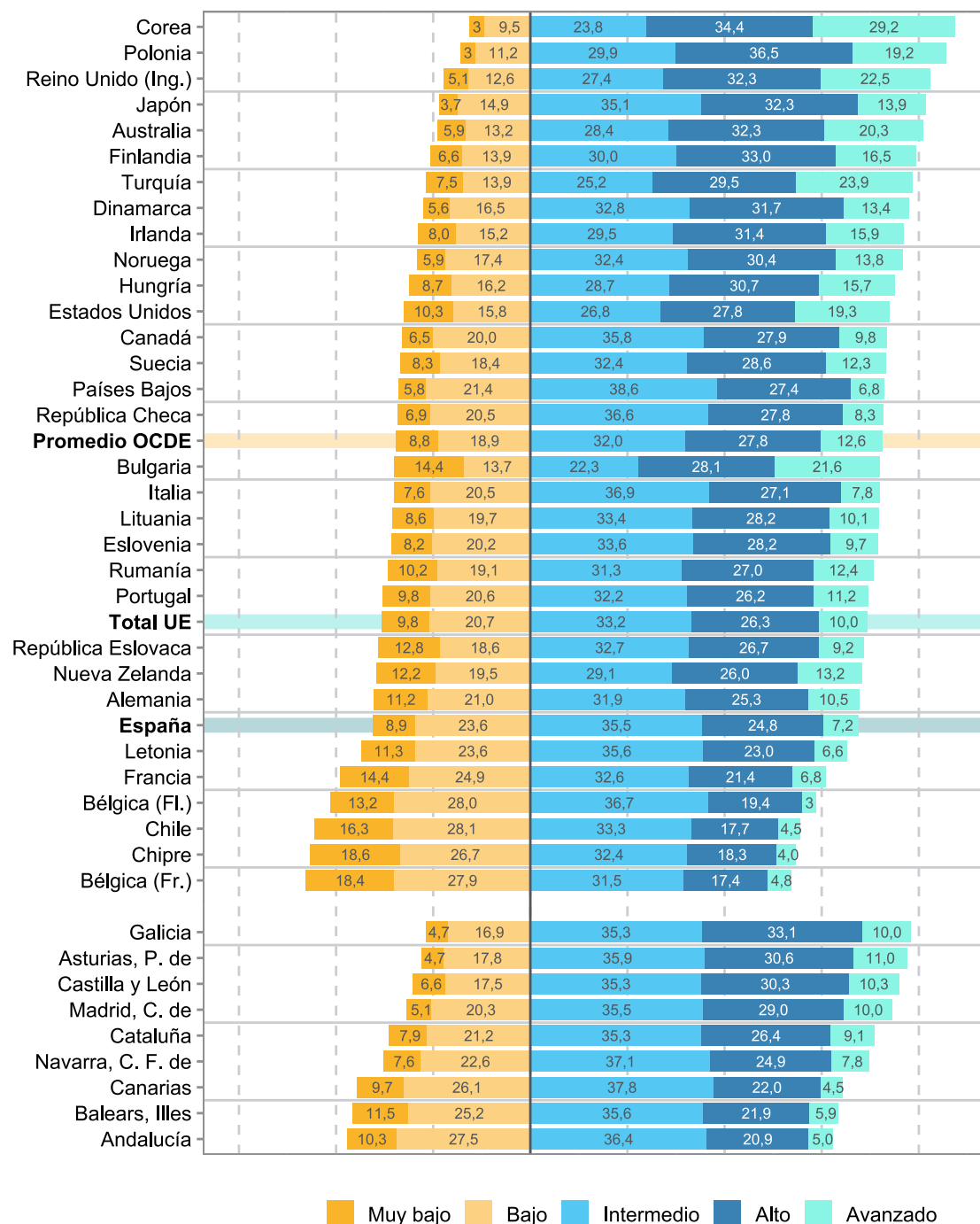
Se estima que en el conjunto de la OCDE algo menos del 28 % del alumnado no alcanza el nivel intermedio (8,8 % nivel muy bajo; 18,9 % nivel bajo). En la UE dicho porcentaje supera el 30 % (9,8 % muy bajo; 20,7 % bajo). Los países mejor ubicados en esta comparación son, por este orden, Corea, Polonia, Reino Unido (Ingl.), Japón y Australia. Todos ellos tienen porcentajes inferiores al 20 % en los niveles más bajos de la escala de rendimiento en conciencia medioambiental. En la situación opuesta se encuentran Chile, Chipre y las dos poblaciones belgas, donde más de 40 % del alumnado se encuentra en los dos niveles más bajos de rendimiento. En España es posible afirmar que uno de cada 3 estudiantes se encuentra en los niveles bajos de rendimiento (8,9 % muy bajo; 23,6 % bajo) (Figura 2.15).

Por su parte, las comunidades autónomas muestran resultados bastante dispares. Así, Galicia y Asturias presentan porcentajes relativamente pequeños (22 %) de estudiantes en los niveles bajos de rendimiento. En el extremo contrario, Andalucía (37,8 %), Illes Balears (36,7 %) y Canarias (35,8 %) presentan porcentajes cercanos a 40 %.

Con respecto a los niveles superiores, alto y avanzado, de la escala de conciencia medioambiental, en la media de países de la OCDE, el 40,4 % de los estudiantes supera el nivel intermedio (27,8 % nivel alto; 12,6 % nivel avanzado), 4 puntos porcentuales más que el Total UE (26,3 % alto; 10,0 % avanzado). En 5 países entre los seleccionados –Corea, Polonia, Reino Unido (Inglaterra), Australia y Turquía– al menos la mitad del alumnado supera el nivel intermedio de la escala de rendimiento en conciencia ciudadana.

En España, aproximadamente 1 de cada 3 estudiantes supera el nivel intermedio en esta competencia, si bien el porcentaje de alumnado en el nivel avanzado es relativamente modesto en el conjunto de la comparación. En este sentido, España presenta uno de los mayores porcentajes de estudiantes en el nivel intermedio. Por su parte, hay tres comunidades autónomas, Galicia, Castilla y León y Principado de Asturias que acumulan más del 40 % del alumnado en los niveles alto y avanzado, mientras que, en Andalucía, Illes Balears y Canarias el porcentaje de alumnado en los dos niveles de rendimiento superiores no alcanza el 30 %.

Figura 2.15. Porcentajes de estudiantes por niveles de rendimiento en conciencia medioambiental. Orden de menor a mayor en los niveles bajo y muy bajo



2.9. Referencias

- Fernández-Alonso, R., y Muñiz, J. (2019). Calidad de los sistemas educativos: Modelos de evaluación. *Propósitos y Representaciones*, 7(SPE), e347. <https://doi.org/10.20511/pyr2019.v7nSPE.347>
- Frey, A., Hartig, J., y Rupp, A. A. (2009). An NCME instructional module on booklet designs in large-scale assessments of student achievement: Theory and practice. *Educational Measurement: Issues and Practice*, 28(3), 39-53. <http://edmeasurement.net/8225/Frey-Hartig-Rupp-2009-booklet-design.pdf>
- Martin, M. O., y Kelly, D. L. (1997). *TIMSS technical report: Vol. II. Implementation and analysis: Primary and middle school years*. TIMSS & PIRLS International Study Center, Boston College
- Mazzeo, J. (2018). Large-scale group-score assessments. En W. J. van der Linden (Ed.) *Handbook of Item Response Theory. Vol. 3: Applications* (pp. 297-311). CRC Press.
- Mullis, I. V. S., Cotter, K. E., Centurino, V. A. S., Fishbein, B. G., y Liu, J. (2016). Using scale anchoring to interpret the TIMSS 2015 achievement scales. En M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (pp. 14.1-14.47). Recuperado de: <http://timss.bc.edu/publications/timss/2015-methods/chapter-14.htm>
- Muñiz, J. (2018). *Introducción a la psicometría*. Pirámide
- OECD (2009). *PISA Data Analysis Manual: SPSS*, Second Edition, OECD Publishing. <https://doi.org/10.1787/9789264056275-en>.
- von Davier, M., Gonzalez, E., y Mislevy, R. (2009). What are plausible values and why are they useful? *IERI Monograph Series, Volume 2*, 9-36.

