

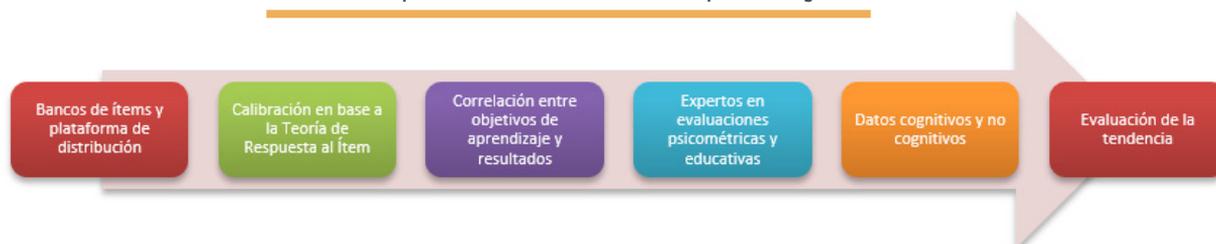
Pruebas adaptativas en las evaluaciones educativas internacionales

Evaluaciones adaptativas digitales

Con el surgimiento del área de la **Inteligencia Artificial** en la década de 1970, el **aprendizaje adaptativo** ganó popularidad y resultó ser la personalización de las técnicas de aprendizaje destinadas a identificar y diferenciar las necesidades específicas del alumnado. Dentro del modelo adaptativo, un elemento fundamental es la medición del progreso o **evaluación**. Las primeras pruebas adaptativas surgieron a partir de 1905 con el desarrollo del primer test de inteligencia por Alfred Binet (Binet IQ Test). Durante mucho tiempo, las evaluaciones se realizaron principalmente en formato papel hasta que, desde finales de los 80 y con la rápida expansión de los ordenadores en la educación, comenzaron las **evaluaciones adaptativas digitales** (*computerized adaptive testing*, CAT) cuya elaboración, aplicación y codificación o corrección tienen lugar por medio de un soporte informático¹.

De forma general, estas pruebas se componen de **ítems** o preguntas seleccionadas de una colección (banco de ítems) de manera que coincidan con el **nivel estimado de capacidad** (o aptitud) del estudiante, es decir, las preguntas se adaptan a las necesidades de los evaluados a partir de distintos aspectos. Este tipo de evaluación se sustenta en la **Teoría de Respuesta al Ítem** (TRI) que constituye un marco de medición ampliamente utilizado en el diseño y análisis de las evaluaciones educativas y psicológicas y que implica un procedimiento para estimar el nivel de habilidad del examinado y otro para seleccionar el ítem más adecuado de acuerdo con dicho nivel² (FIGURA 1).

FIGURA 1. Componentes de las evaluaciones adaptativas digitales³



Así, gracias a la digitalización y con el fin de adaptar la dificultad de las pruebas a la capacidad de las poblaciones objetivo, en las **evaluaciones educativas internacionales a gran escala** (*International Large-Scale Assessments*, ILSA) se han integrado actualmente distintos diseños:

- **Evaluaciones adaptativas de múltiples etapas** (*Multistage adaptive testing*, MST): en las que la adaptación se produce en las distintas secciones de la prueba dependiendo del rendimiento del evaluado en cada una de ellas o, también teniendo en cuenta sus antecedentes sobre ocupación, nivel socioeconómico o educativo.
- **Diseños adaptativos longitudinales**: en los cuales se crean pruebas de distinta dificultad dirigidas a poblaciones de diferente capacidad que son identificadas a partir de datos de rendimiento de evaluaciones educativas anteriores.

A continuación, se presentan las características de estos diseños aplicados en las principales evaluaciones educativas internacionales: **PIAAC**, **PISA** y **PIRLS**.

¹ Hambleton, R., Zaal, J. y Pieters, J. (1991). Computerized adaptive testing: Theory, applications and standards. En R. Hambleton y J. Zaal (Eds.), *Advances in educational and psychological testing: Theory and applications*. Boston: Kluwer Academic.

² Olea, J. y Ponsoda, V. (1996). Test adaptativos informatizados. En J. Muñiz (Ed.), *Psicometría* (pp. 729-783). Madrid: Universitas, S. A.

³ Elaboración propia basada en: <https://www.compartirpalabramaestra.org/actualidad/articulos-informativos/aprendizaje-y-evaluacion-adaptativos-aprendizaje-adaptativo>

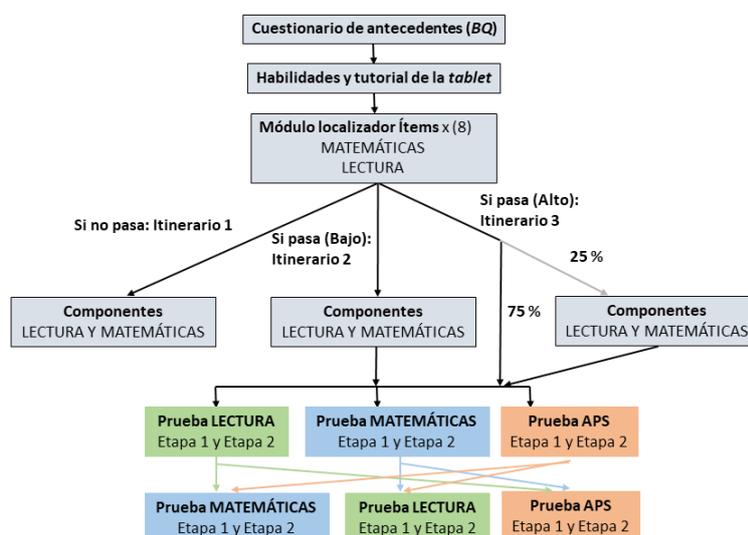
Evaluaciones adaptativas de múltiples etapas (MST): PIAAC CICLO-2

En este caso, la adaptación se produce en bloque después de que el evaluado reciba un conjunto de ítems o unidades, en vez de uno solo, de forma que el **rendimiento del examinado en cada etapa afectará a la elección del bloque** (conjunto de varias unidades) **de la siguiente etapa**.

Dado que el objetivo de las ILSA es describir **poblaciones** de interés, no evaluar individuos concretos, y puesto que sus mediciones se basan en **numerosos ítems** que han de ser traducidos en diferentes idiomas, el diseño adaptativo de múltiples etapas conlleva bastantes **ventajas** prácticas. Por un lado, consigue **reducir el error de medida** en poblaciones heterogéneas, sin sobrecargar a cada evaluado en particular, y, por otro lado, **mejora la comparabilidad entre países**⁴.

El **Programa para la Evaluación Internacional de las Competencias de la población Adulta** (PIAAC, por sus siglas en inglés), es una evaluación internacional de la **OCDE** que recoge datos a través de una encuesta domiciliaria a personas entre 16 y 65 años de edad. En este estudio, las pruebas adaptativas se aplicaron desde su primera edición, que se inició formalmente en 2008 y se aplicó en España entre 2011 y 2012, con la publicación de los resultados en 2013. Actualmente, en el segundo ciclo del estudio, la prueba digital administrada en **tablets** consiste en un **cuestionario de antecedentes** (*background questionnaire*, BQ) y una **prueba cognitiva** que mide **tres áreas**: comprensión lectora, matemáticas y solución adaptativa de problemas (*adapting problem solving*, APS). La **FIGURA 2** muestra el diseño de **evaluación adaptativa de múltiples etapas** para el Estudio Principal del **2.º ciclo de PIAAC**⁵.

FIGURA 2. Diseño MST para el Estudio Principal del 2.º ciclo de PIAAC



Por tanto, en este estudio la **adaptación** se basa en la información recogida del evaluado en el **cuestionario de antecedentes** (a través de un conjunto de preguntas relativas al dominio de la competencia tecnológica), en sus **habilidades con la tablet** y en su **nivel de competencia en lectura y matemáticas** (por medio del módulo denominado localizador).

Así, dependiendo de los datos recopilados, el evaluado podrá ser dirigido hacia tres **itinerarios** diferentes:

- **Itinerario 1:** Si el evaluado no pasa el localizador, realizará la prueba de componentes básicos de lectura y matemáticas.
- **Itinerario 2:** Si el evaluado pasa el localizador, pero con un rendimiento relativamente bajo, realizará la prueba de componentes además de un conjunto más breve y fácil de los módulos adaptados de comprensión lectora, matemáticas y APS, según una secuencia aleatoria.
- **Itinerario 3:** Si en el localizador se obtiene un rendimiento alto, el 25 % de los evaluados realizará primero la prueba de componentes antes de pasar a los módulos cognitivos adaptados, mientras que la mayoría de ellos (75 %) pasará directamente a dichos módulos.

⁴ Yamamoto, K., Shin, H. J. y Khorramdel, L. (2018). Introduction of multistage adaptive testing design. *Educational Measurement: Issues and Practice*, 37(4), 16-27. Recuperado de https://www.researchgate.net/publication/328611047_Multistage_Adaptive_Testing_Design_in_International_Large-ScaleAssessments

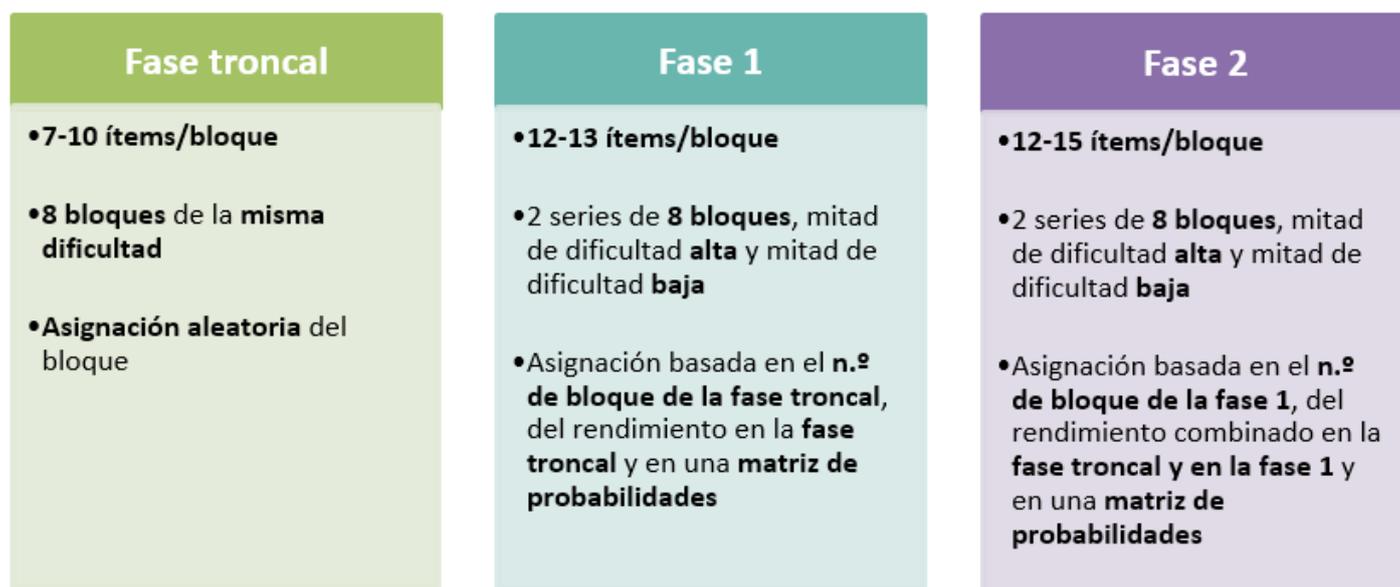
⁵ OECD. (2021). *The Assessment Frameworks for Cycle 2 of the Programme for the International Assessment of Adult Competencies*, OECD Skills Studies. Paris: OECD Publishing. <https://doi.org/10.1787/4bc2342d-en>

Evaluaciones adaptativas de múltiples etapas (MST): PISA 2018

En el caso del **Programa para la Evaluación Internacional de los Estudiantes** (PISA, por sus siglas en inglés), evaluación de la OCDE dirigida al alumnado de 15-16 años, el diseño adaptativo MST tuvo lugar por primera vez en el **dominio principal de lectura** en su edición de 2018. Este diseño presentó diferencias respecto al de PIAAC al ser **más complejo** en el enrutamiento de los diferentes itinerarios y al **no integrar** en los mismos **información relativa al contexto** de los evaluados. La necesidad de discriminar mejor en los **niveles más altos y más bajos** de la distribución del rendimiento del alumnado y, por tanto, **mejorar en la precisión** de las medidas, es lo que condujo a la evaluación adaptativa de lectura en PISA.

La prueba contó con un diseño de **tres fases**: troncal, fase 1 y fase 2. Los estudiantes realizaban primero la **fase troncal**, que consistía en un bloque de entre 7 y 10 ítems o preguntas, donde la mayoría (80 %) se codificaba automáticamente. Así, los estudiantes se clasificaban provisionalmente como de **nivel bajo, medio o alto**, dependiendo del número de respuestas correctas. Después, en las **fases 1 y 2**, había **dos niveles de dificultad**: comparativamente fácil y comparativamente difícil (8 bloques de cada). Cada bloque de la **fase 1** se componía de 12 a 13 preguntas, de las que de 8 a 11 se codificaban automáticamente y en la **fase 2**, de 12 a 15 preguntas, de las que de 6 a 12 se codificaban por ordenador. El alumnado clasificado con un **nivel medio** en la fase troncal, tenía las **mismas probabilidades** de que se le asignara un nivel más fácil o más difícil en la fase 1. Los que se clasificaron con **nivel bajo**, tenían una probabilidad del **90 %** de que se les asignara un nivel **más fácil** en la fase 1 y un **10 %** de probabilidad de que se les asignara un nivel **más difícil**. Los clasificados como de nivel alto en la fase troncal, tenían una probabilidad del **90 %** de que se les asignara un nivel más difícil en la fase 1 y un **10 %** de probabilidad de que se les asignara un nivel **más fácil**. Por último, para la **fase 2**, el procedimiento era parecido como puede observarse en la FIGURA 3⁶.

FIGURA 3. Diseño MST para la prueba de lectura en PISA 2018



Este **diseño de evaluación adaptativa** seguirá aplicándose en la **prueba de lectura** de la próxima edición de **PISA 2022** donde también se va a realizar la prueba de **matemáticas** de forma adaptativa. Se espera que en **PISA 2025** las **tres competencias troncales** evaluadas (lectura, matemáticas y ciencias) sean adaptativas.

⁶ Para más información, consultar el Informe Técnico de PISA 2018: <https://www.oecd.org/pisa/data/pisa2018technicalreport/>

Diseños adaptativos longitudinales: PIRLS 2021

El diseño adaptativo por grupo en el caso del **Estudio Internacional para el Progreso de la Comprensión Lectora** (PIRLS, por sus siglas en inglés), evaluación de la IEA dirigida al alumnado de 4.º grado (4.º de Educación Primaria en España), permite personalizar la prueba según el **promedio de habilidad lectora obtenido por los países participantes en los ciclos previos** de la evaluación o, en su caso, a través de los estudios piloto que realizan aquellos países que participan por primera vez.

Así, los **18 textos** que componen la evaluación se dividen en **tres niveles de dificultad** (difícil, media y fácil) y se combinan dando lugar a **dos niveles de dificultad en los cuadernillos** que realizan los evaluados, resultando:

- **9 cuadernillos más difíciles** compuestos por 2 textos difíciles o de uno difícil y otro de dificultad media
- **9 cuadernillos menos difíciles** compuestos por 2 textos fáciles o de uno fácil y otro de dificultad media

Con este diseño, **todos los países participantes administrarán los 18 textos, pero en distintas proporciones** con el **objetivo** de lograr una **mejor correspondencia entre la dificultad** de la evaluación y el **rendimiento** del alumnado en cada país. Para ello, en los países con un **nivel de rendimiento alto** en los ciclos anteriores del estudio se administrarán proporcionalmente una **mayor cantidad de cuadernillos más difíciles**, mientras que en los países que tuvieron un **rendimiento más bajo** se distribuirán proporcionalmente **más cuadernillos de los de menor dificultad** (FIGURA 4)⁷.

FIGURA 4. Asignación de cuadernillos según el rendimiento de los países participantes en ciclos anteriores



Se puede observar en la FIGURA 4 cómo es el **porcentaje de distribución de los cuadernillos** en los diferentes países según los **puntos obtenidos en la escala de rendimiento de PIRLS** y que los clasificará por su desempeño en alto, medio o bajo, cuando estén por encima de los 550 puntos, entre los 450 y los 550 o por debajo de los 450 puntos, respectivamente. Este diseño adaptativo por grupo en PIRLS 2021 implica el cambio del procedimiento utilizado en ciclos anteriores en el que los cuadernillos se asignaban aleatoriamente a los estudiantes de cada país en la misma proporción y pretende **mejorar la precisión de la medida, proporcionar ventajas prácticas y operativas y aumentar la tasa de respuesta** del alumnado al obtener mayor compromiso y reducir su frustración.

Por último, el **Estudio Internacional de Tendencias en Matemáticas y Ciencias** (TIMSS, por sus siglas en inglés) de la IEA en su próxima edición de 2023 también contará con una prueba adaptativa de características similares a la de PIRLS 2021.

⁷ Para más información, consultar los Marcos de Evaluación PIRLS 2021: <https://www.educacionyfp.gob.es/inee/evaluaciones-internacionales/pirls/pirls-2021.html>

