



GOBIERNO
DE ESPAÑA

MINISTERIO
DE EDUCACIÓN, CULTURA
Y DEPORTE

inee

Instituto Nacional
de Evaluación
Educativa

Jornada Evaluación y mejora de la calidad educativa

Logroño, 13 de diciembre de 2017

Evaluar...¿para qué?

Elaboración de pruebas, marco muestral y análisis de datos

Francisco Javier García Crespo
Javier.gcrespo@mecd.es

Evaluaciones educativas, finalidad.

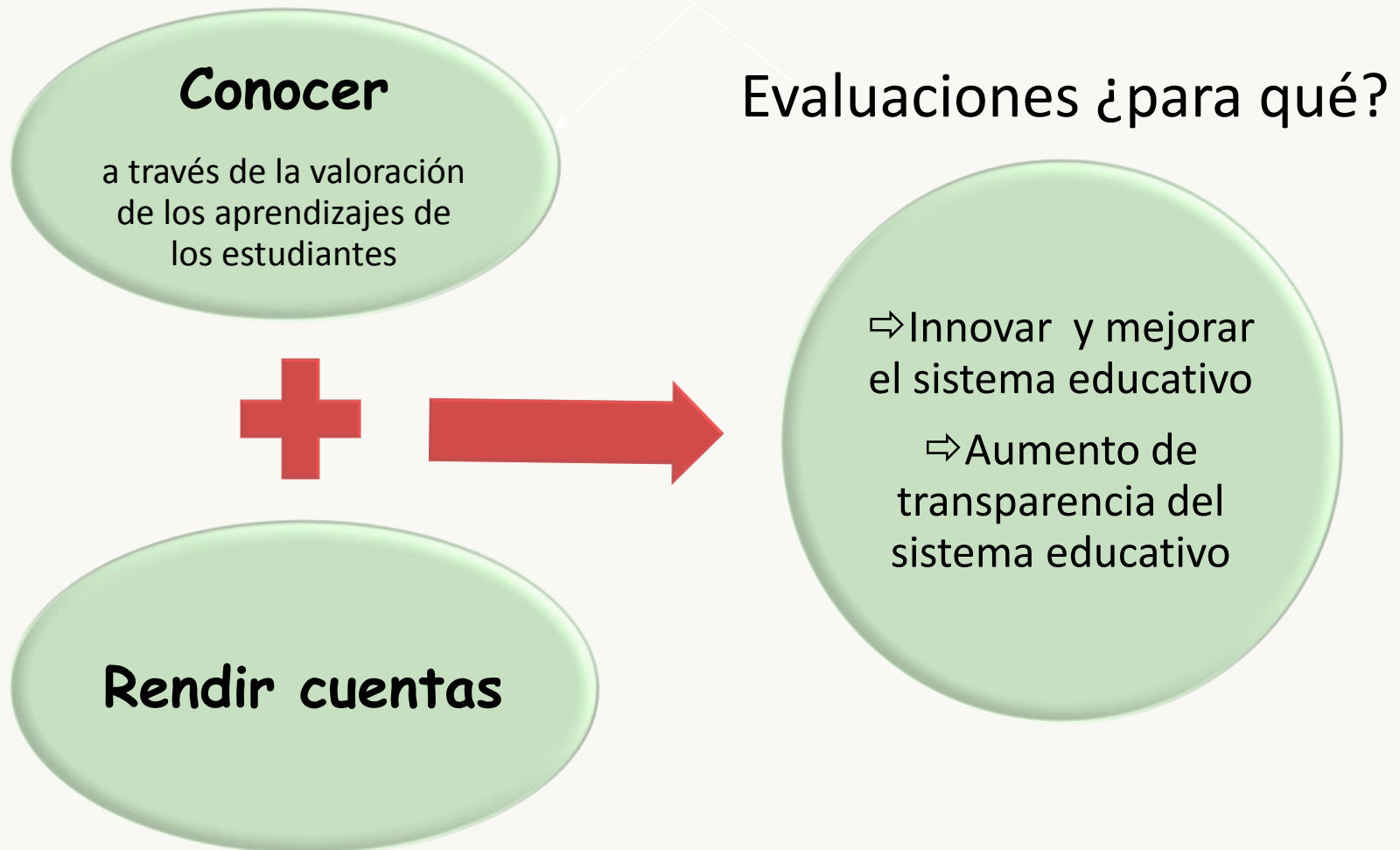
Marco de evaluación.

Diseño de las pruebas.

Diseño muestral.

Análisis de datos y presentación de resultados.

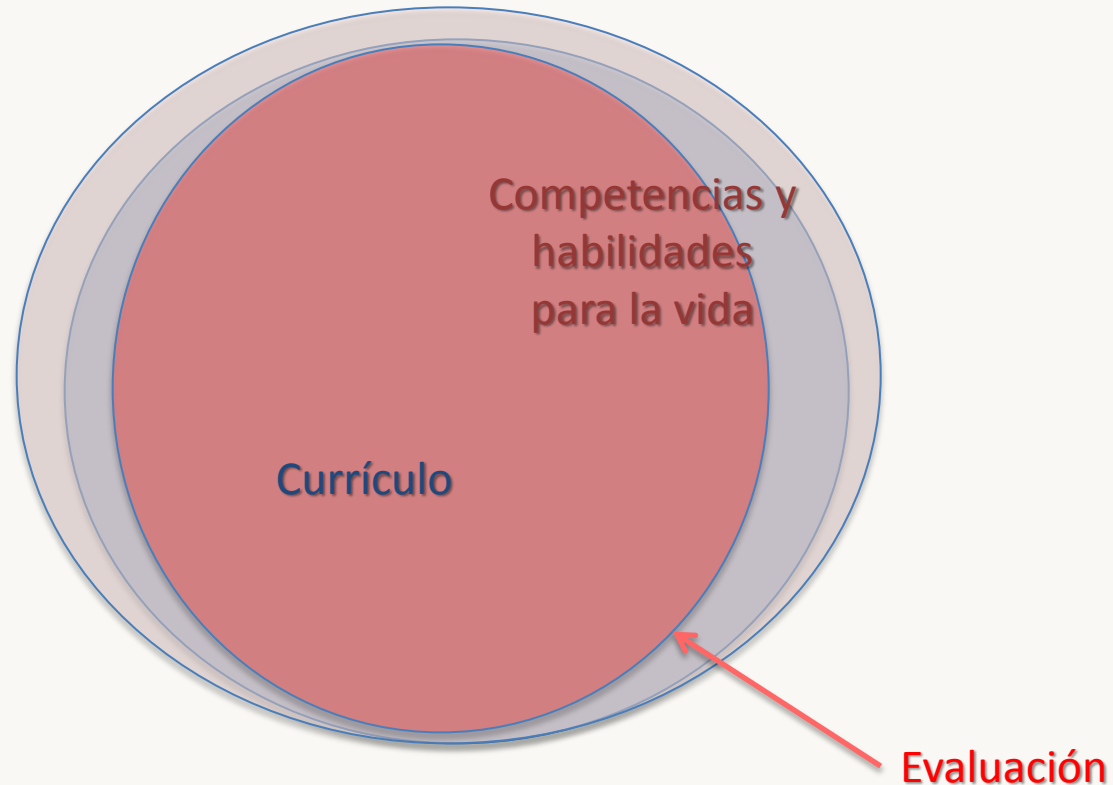
FINALIDAD DE LAS EVALUACIONES EDUCATIVAS



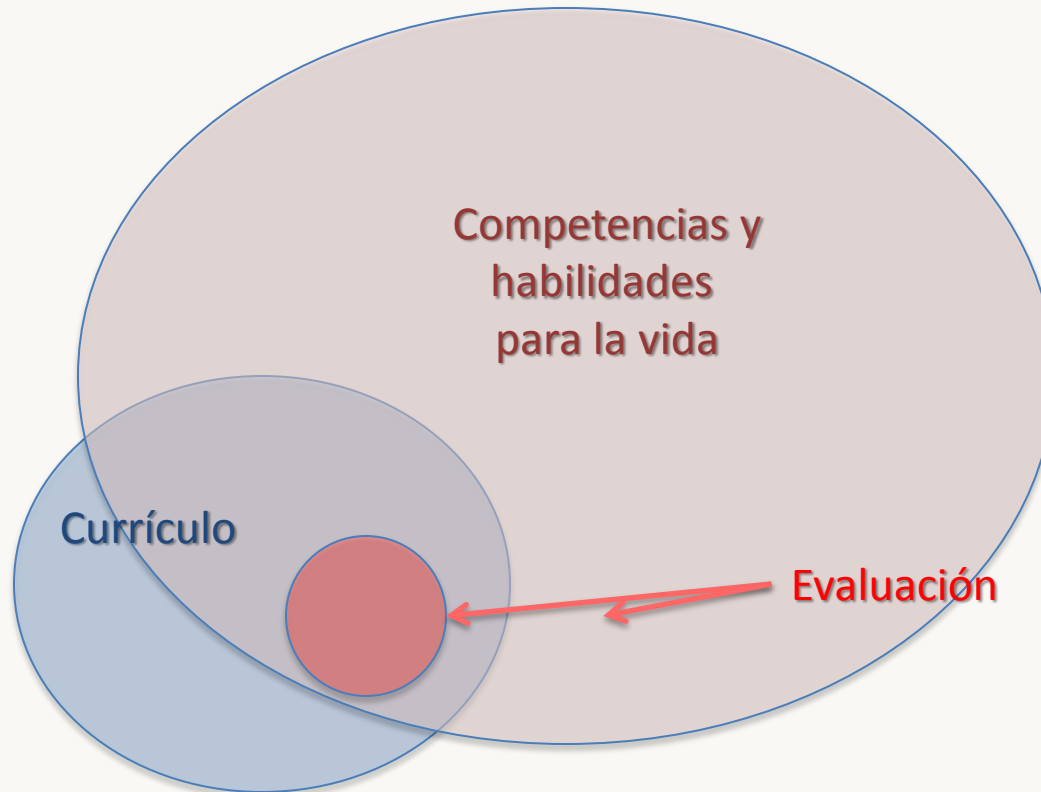
Rasgos comunes de los estudios de evaluación



- **Evaluaciones muestrales, cíclicas, a gran escala**
- **Cuadernillo de prueba + cuestionarios de contexto**
- **Rigor metodológico:**
 - Marco teórico: Qué evaluar, elaboración de ítems, distribución de ítems y cuadernillos...
 - Selección de la muestra
 - Traducción de materiales, aplicación, corrección, limpieza de datos...
 - Análisis TRI de las respuestas → Dificultad de los ítems y puntuación lograda por los alumnos
- **Informes comparativos**



LO DESEABLE DE UN SISTEMA EDUCATIVO



LO OBSERVADO CON UNA PRUEBA DE EVALUACIÓN

Transparencia

Rigurosidad

Evaluación
externa

Integrada en el proceso
educativo habitual

Los resultados deben
valorarse en su
contexto

Evaluaciones Educativas Internacionales

OCDE (Organización para la Cooperación y Desarrollo Económico)

	Destinatarios	Periodicidad	Última realizada
PISA (Programme for International Student Assessment)	Alumnos que cumplen 16 años en el año de realización	3 años	2015
TALIS, PISA-LINK (Teaching and Learning International Survey)	Profesores y directores de Educación Secundaria	5 años	2013
PIAAC (Programme for International Assessment of Adult Competences)	Población adulta (16 a 65 años)	10 años	2012

Evaluaciones Educativas Internacionales

IEA (International Association for the Evaluation of Educational Achievement)

	Destinatarios	Periodicidad	Última realizada
PIRLS (Progress in International Reading Literacy Study)	4º EP	5 años	2016
TIMSS (Trends in International Mathematics and Science Study)	4º EP	4 años	2015
ICILS (International Computer and Information Literacy Study)	2º ESO	No participamos	Piloto
ICCS (International Civic and Citizenship Education Study)	2º ESO	6 años	2009

Evaluaciones Educativas Nacionales

MECD/Comunidades Autónomas			
	Destinatarios	Periodicidad	Última realizada
EGD Primaria	4º EP		2009
EGD Secundaria	2º ESO		2010
Evaluaciones LOMCE	3º EP 6ºEP 4º ESO	Anual	2017

Evaluaciones educativas, finalidad.

Marco de evaluación.

Diseño de las pruebas.

Diseño muestral.

Análisis de datos y presentación de resultados.

Definición explícita sobre lo que pretende medir la evaluación:

- ✓ Descripción racional del programa de evaluación.
- ✓ Definición del área, materia, competencia...
- ✓ Descripción de las variables.
- ✓ Guía de construcción de las pruebas.
- ✓ Modelos de ítems o preguntas.

MARCO DE LA EVALUACIÓN

¿Qué evaluar?

Competencias

Medir el éxito con el que los estudiantes o adultos pueden extrapolar lo que han aprendido y aplicar sus conocimientos y habilidades en nuevos contextos

PISA, Ev.Nacionales, PIAAC

Currículo-Conocimientos

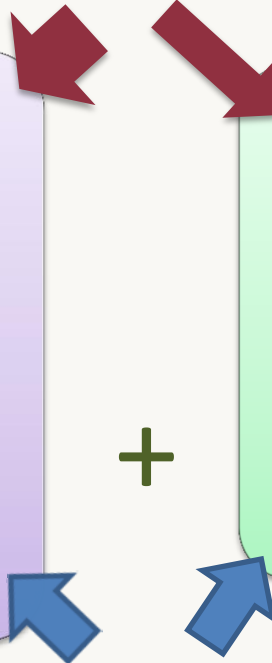
Medir lo que se espera que los estudiantes hayan aprendido

TIMSS, PIRLS

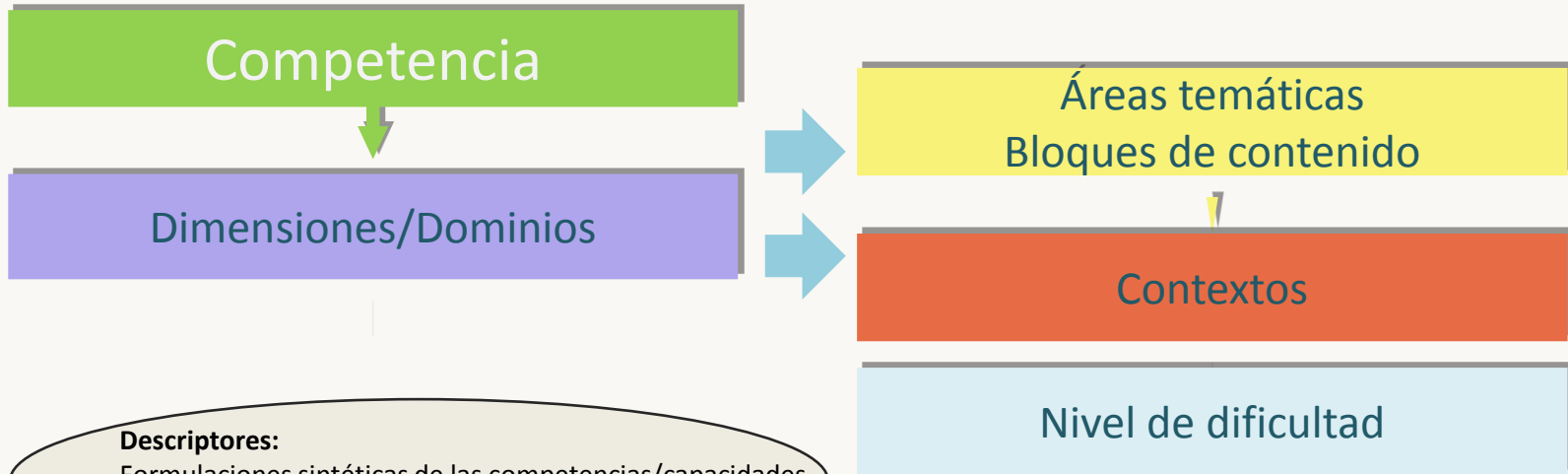
+

Contexto

Todos los estudios y TALIS



ANÁLISIS DE LA COMPETENCIA



Descriptor:
Formulaciones sintéticas de las competencias/capacidades que medirán los ítems (*estándares de aprendizaje*)

Dimensiones/Dominios	Procesos	Niveles cognitivos (Niveles de complejidad en la resolución de la tarea)			Ponderación
		Proceso I	Proceso II	Proceso III	
Dominio 1	Descriptor	Descriptor	Descriptor	Descriptor	%
Dominio 2	Descriptor	Descriptor	Descriptor	Descriptor	%
Dominio 3	Descriptor	Descriptor	Descriptor	Descriptor	%
Ponderación		%	%	%	

Ejemplo de marco de evaluación

La competencia científica en PISA



inee



Competencia Científica

“es la capacidad de interesarse e implicarse en temas científicos e ideas sobre la ciencia como ciudadano consciente y reflexivo”

Conocimiento científico de biología, geología, física, química y tecnología.

Tres sub-competencias

- Explicar fenómenos científicamente.
- Evaluar y diseñar experimentos y preguntas científicas (*scientific enquiry*).
- Interpretar datos y pruebas (*data and evidence*) científicamente.

La competencia científica en PISA



inee

Por tanto...

Una persona científicamente competente es aquélla que se muestra interesada por la ciencia, mantiene discursos racionales sobre cuestiones científicas.

Explica fenómenos científicamente:

Reconoce, ofrece y valora explicaciones para una variedad de fenómenos naturales u tecnológicos.

Evalúa y diseña experimentos y preguntas científicas:

Describe y evalúa la investigación científica y propone medios para elaborar y responder preguntas científicas.

Interpreta datos y pruebas científicamente:

Analiza y evalúa datos científicos, enunciados y argumentos en contextos diversos y extrae conclusiones adecuadas.

¿Qué tipo de conocimiento se evalúa?

■ Conocimiento sustantivo

- Sistemas físicos
- Sistemas biológicos
- Sistemas de la Tierra y el Espacio

■ Conocimiento procedimental

- Procedimientos y estrategias de cualquier forma de investigación científica

■ Conocimiento epistémico

- La manera en que las ideas se justifican y garantizan en ciencia

■ Actitudes

- Interés por la ciencia
- Valoración de enfoques científicos
- Conciencia medio-ambiental

- En 2006 se incluyeron en las preguntas cognitivas.
- En 2015 se han medido a través del cuestionario de contexto

TEMAS

	Personal	Local/nacional	Global
Salud y enfermedad	Mantenimiento de la salud, accidentes, nutrición.	Control de enfermedades, transmisión social, elección de alimentos, salud comunitaria.	Epidemias, propagación de enfermedades infecciosas.
Recursos naturales	El consumo personal de materiales y energía.	El mantenimiento de las poblaciones humanas, calidad de vida, seguridad, producción y distribución de alimentos, suministro de energía.	Sistemas naturales renovables y no renovables, crecimiento demográfico, uso sostenible de las especies.
Calidad medioambiental	Acciones favorables al medio ambiente, uso y desecho de materiales y dispositivos.	Distribución de la población, eliminación de residuos, impacto ambiental.	Biodiversidad, sostenibilidad ecológica, control de la contaminación, producción y pérdida de suelo/biomasa.
Riesgos naturales	Las evaluaciones de riesgo del estilo de vida.	Cambios rápidos (por ejemplo, terremotos, clima severo), cambios lentos y progresivos (por ejemplo, erosión costera, sedimentación), evaluación de riesgos.	Cambio climático, impacto de la comunicación moderna.
Fronteras entre la ciencia y la tecnología	Aspectos científicos de las aficiones personales, tecnología personal, música y actividades deportivas.	Los nuevos materiales, dispositivos y procedimientos, modificaciones genéticas, tecnología de la salud y transporte.	Extinción de especies, exploración del espacio, origen y estructura del universo.

Evaluaciones educativas, finalidad.

Marco de evaluación.

Diseño de las pruebas.

Diseño muestral.

Análisis de datos y presentación de resultados.

ELABORACIÓN DE LOS ÍTEMS

¿Pruebas en papel o en ordenador?

- Anuncios
- Textos extraídos de los medios de comunicación
- Instrucciones
- Carteles informativos
- Diálogos, narraciones de hechos
- Descripción de situaciones reales o simuladas
- Textos literarios o científicos

TEXTOS ESCRITOS

Estímulos
Situaciones-problema

IMÁGENES

- Fotografías
- Dibujos
- Mapas
- Esquemas

Preguntas/ítems

Respuesta de elección múltiple
Respuesta cerrada
Respuesta abierta construida
Respuestas abiertas

Matriz de especificaciones

CONFIGURACIÓN DE LAS PRUEBAS

Generación de un banco de ítems

- Redacción de un número elevado de ítems
- Cobertura de diferentes niveles de desarrollo de la competencia
- Redacción de criterios de codificación, contemplando graduación de respuestas

Revisión de la calidad de los ítems

- Corrección
- Defectos en la redacción
- Adecuación de las respuestas previstas
- Graduación de las respuestas
- Adaptación al nivel al que se destinan
- Independencia

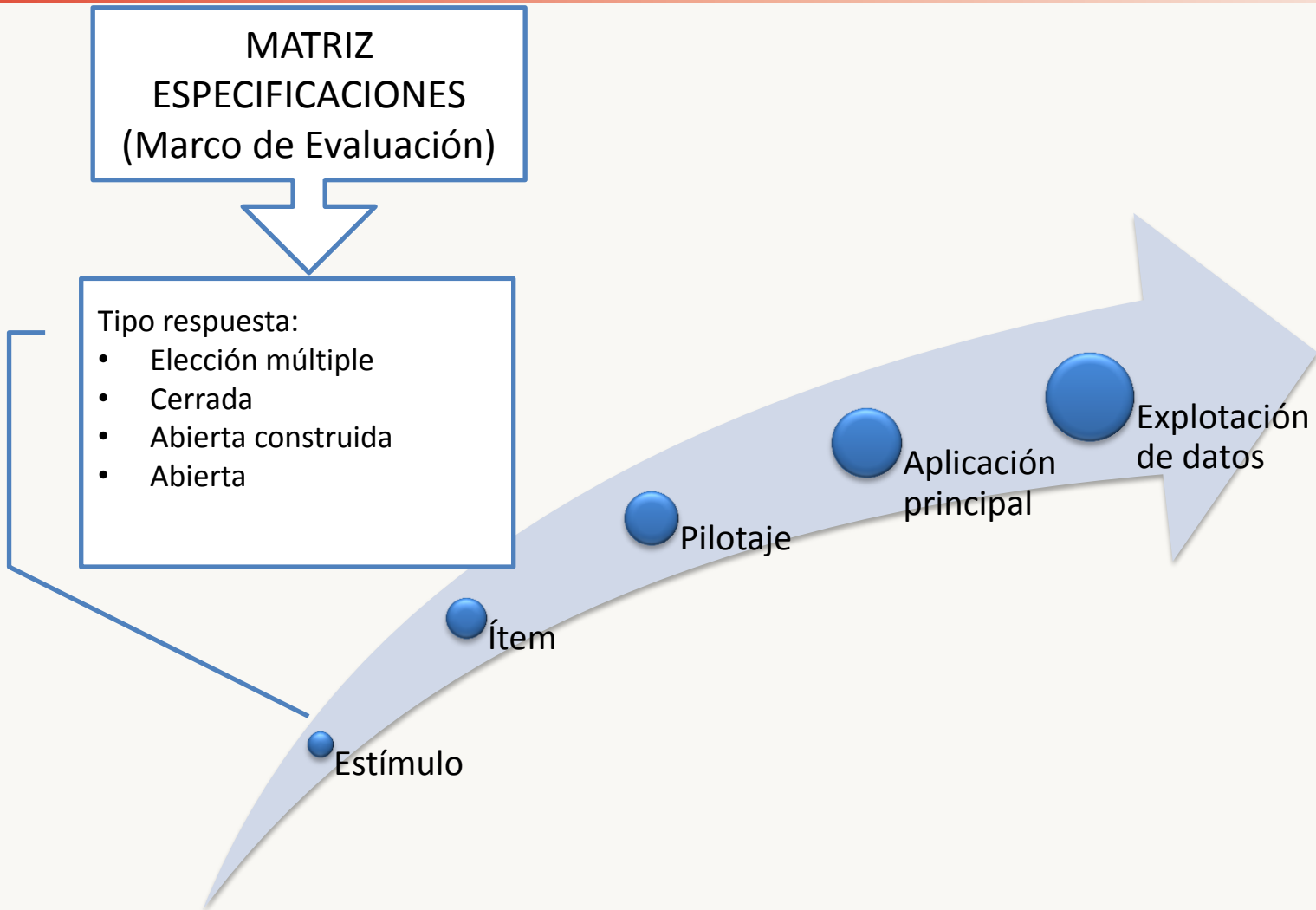
Pruebas prepiloto y piloto

CRITERIOS PARA LA SELECCIÓN DE ÍTEMES

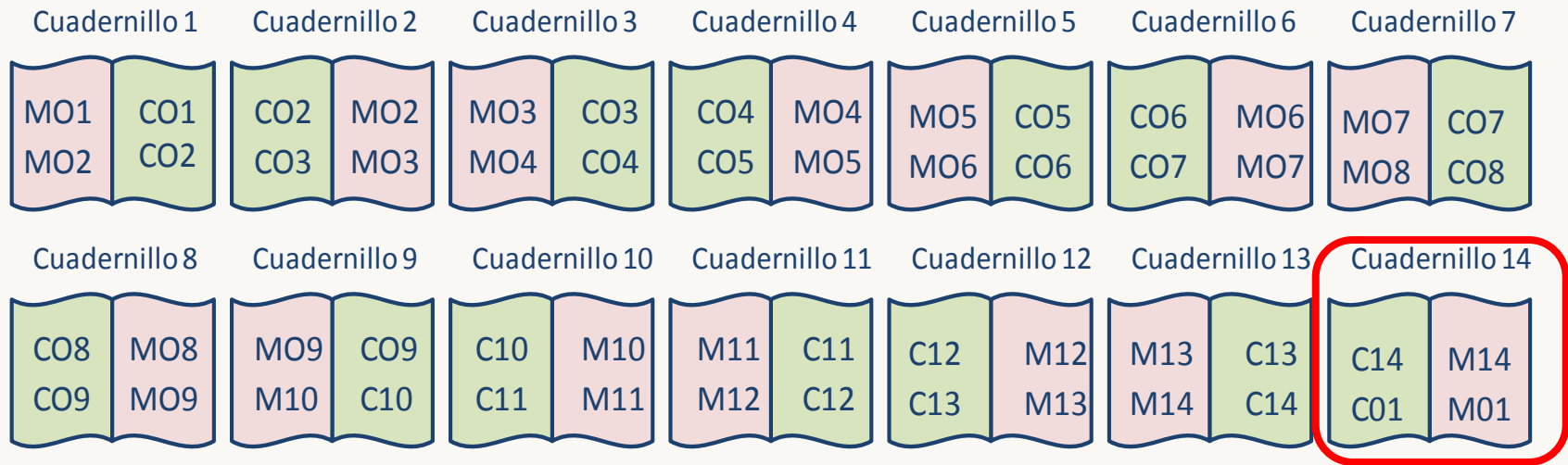
- Cobertura de las competencias objeto de evaluación
- Presencia de todas las categorías de respuesta
- Evitar ítems con bajos niveles de discriminación
- Evitar dificultad o facilidad extrema
- Inclusión de ítems de diferentes niveles de dificultad
- Equiparación con otros modelos de pruebas

Estudio principal

CONSTRUCCIÓN DE LAS PRUEBAS



DISEÑO DE TIMSS



- 14 bloques de matemáticas y 14 de ciencias
- 8 bloques de cada materia son de **anclaje**
- Cada bloque tiene entre 10 y 14 preguntas de dominios de contenido y cognitivos en proporción del marco teórico
- 175 ítems de matemáticas
- 172 ítems de ciencias

- **72 minutos**
- **unas 40 preguntas**

Llevaría 8 horas y media realizar toda la prueba cognitiva

DISTRIBUCIÓN DE TAREAS A LOS ESTUDIANTES

- El diseño de las pruebas cognitivas de PISA 2015 se realiza mediante un proceso aleatorio de asignación en dos fases
- 1. Formato base (N: del 31 al 96)
 - Se sortea el formato básico del test que corresponde a un alumno determinado.
- 2. Número aleatorio (S: del 1 al 36)
 - Se sortea la combinación de bloques de Ciencias que deberá responder el alumno.

TABLA BASE PARA LA ASIGNACIÓN ALEATORIA

Percentage of Assignment	Base Test Forms	Cluster 1	Cluster 2	Cluster 3	Cluster 4
33% (No CPS: 46%)	31	S	S	R01	R02
	32	S	S	R02	R03
	33	S	S	R03	R04
	34	S	S	R04	R05
	35	S	S	R05	R06ab
	36	S	S	R06ab	R01
	37	R01	R03	S	S
	38	R02	R04	S	S
	39	R03	R05	S	S
	40	R04	R06ab	S	S
	41	R05	R01	S	S
	42	R06ab	R02	S	S
33% (No CPS: 46%)	43	S	S	M01	M02
	44	S	S	M02	M03
	45	S	S	M03	M04
	46	S	S	M04	M05
	47	S	S	M05	M06ab
	48	S	S	M06ab	M01
	49	M01	M03	S	S
	50	M02	M04	S	S
	51	M03	M05	S	S
	52	M04	M06ab	S	S
	53	M05	M01	S	S
	54	M06ab	M02	S	S
4% (No CPS: 8%)	55	S	S	M01	R01
	56	S	S	R02	M02
	57	S	S	M03	R03
	58	S	S	R04	M04
	59	S	S	M05	R05
	60	S	S	R06ab	M06ab
	61	R01	M01	S	S
	62	M02	R02	S	S
	63	R03	M03	S	S
	64	M04	R04	S	S
	65	R05	M05	S	S
	66	M06ab	R06ab	S	S

Percentage of Assignment	Base Test Forms	Cluster 1	Cluster 2	Cluster 3	Cluster 4
4% (No CPS: NA)	67	S	S	C01	M01
	68	S	S	M02	C02
	69	S	S	C03	M03
	70	S	S	M04	C03
	71	S	S	C02	M05
	72	S	S	M06ab	C01
	73	M01	C02	S	S
	74	C03	M02	S	S
	75	M03	C01	S	S
	76	C01	M04	S	S
	77	M05	C03	S	S
	78	C02	M06ab	S	S
4% (No CPS: NA)	79	S	S	R01	C01
	80	S	S	C02	R02
	81	S	S	R03	C03
	82	S	S	C03	R04
	83	S	S	R05	C02
	84	S	S	C01	R06ab
	85	C02	R01	S	S
	86	R02	C03	S	S
	87	C01	R03	S	S
	88	R04	C01	S	S
	89	C03	R05	S	S
	90	R06ab	C02	S	S
22% (No CPS: NA)	91	S	S	C01	C02
	92	S	S	C02	C03
	93	S	S	C03	C01
	94	C02	C01	S	S
	95	C03	C02	S	S
	96	C01	C03	S	S

R01-R06 → Lectura

M01-M06 → Matemáticas

S → Ciencias

C01-C03: Resolución de problemas colaborativos

TABLA PARA ASIGNACIÓN ALEATORIA CIENCIAS

- 36 posibles combinaciones

Science Cluster Combination		
N	S	S
1	S01	S07
2	S01	S10
3	S02	S08
4	S03	S09
5	S03	S12
6	S04	S07
7	S04	S10
8	S05	S11
9	S06	S12
10	S07	S06
11	S08	S01
12	S08	S05
13	S09	S02
14	S09	S06
15	S10	S03
16	S11	S02
17	S11	S04
18	S12	S05

Science Cluster Combination		
N	S	S
19	S07	S08
20	S07	S09
21	S07	S11
22	S08	S10
23	S08	S12
24	S09	S08
25	S09	S11
26	S10	S07
27	S10	S09
28	S10	S12
29	S11	S08
30	S11	S10
31	S12	S07
32	S12	S09
33	S12	S11
34	S02	S04
35	S05	S01
36	S06	S03

Ejemplo

Asignación aleatoria del número N=34
R4 y R05 (tabla base)

Asignación aleatoria del número S = 4
S03 y S09 (tabla Ciencias)

Base Form (CC)	Random number (S)					
	1	2	3	4	5	6
34	35	4	7	19	23	30

Resultado del sorteo



Cluster 1	Cluster 2	Cluster 3	Cluster 4
S03	S09	R4	R05

Science Cluster Combination		
N	S	S
1	S01	S07
2	S01	S10
3	S02	S08
4	S03	S09
5	S03	S12
6	S04	S07
7	S04	S10
8	S05	S11
9	S06	S12
10	S07	S06
11	S08	S01
12	S08	S05
13	S09	S02
14	S09	S06
15	S10	S03
16	S11	S02
17	S11	S04
18	S12	S05

Evaluaciones educativas, finalidad.

Marco de evaluación.

Diseño de las pruebas.

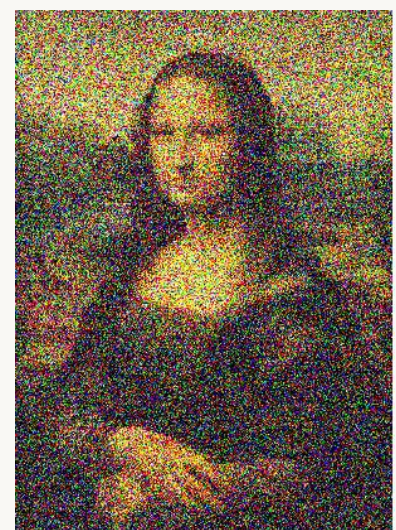
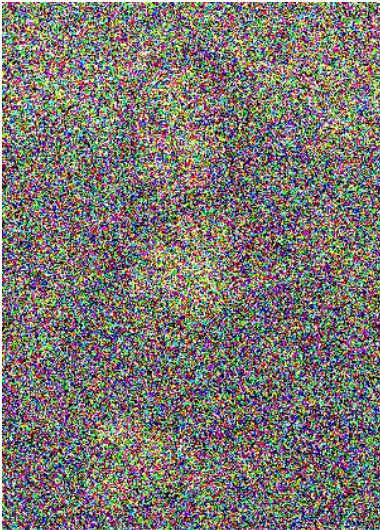
Diseño muestral.

Análisis de datos y presentación de resultados.

SELECCIÓN DE LA MUESTRA: Tamaño

- Tiempo
- Presupuesto
- Sencillez del diseño del muestreo y del trabajo de campo

- Precisión de los estimadores
- Tamaño de la muestra

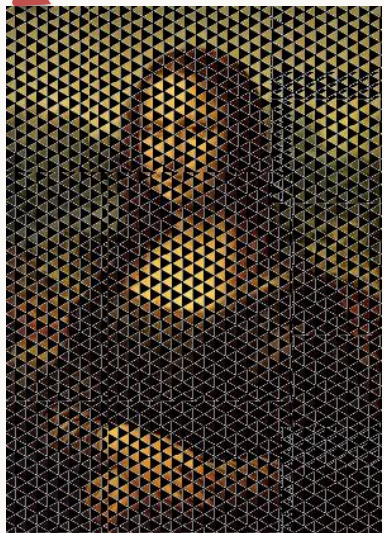
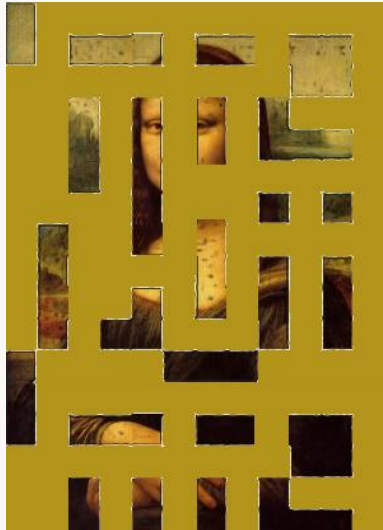
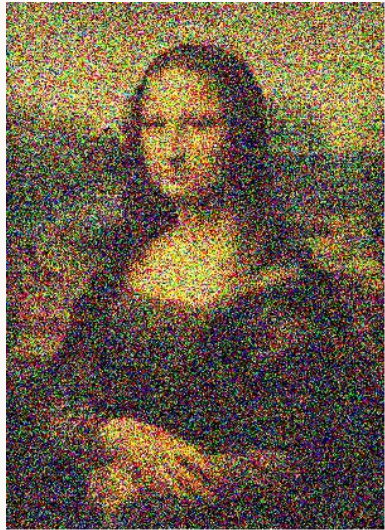


SELECCIÓN DE LA MUESTRA: Individuos frente a centros

Muestreo por individuos (alumnos)

Muestreo por conglomerados (centros educativos)

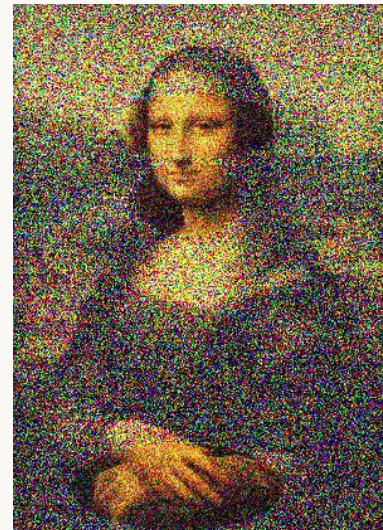
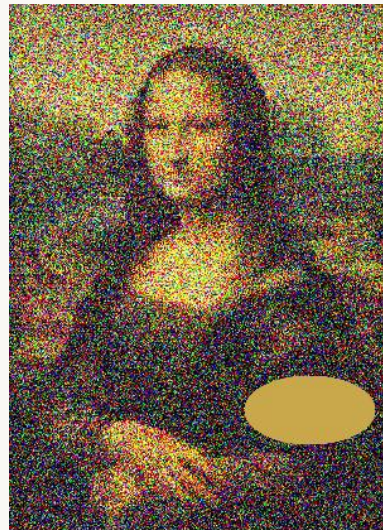
Selección sistemática de centros



SELECCIÓN DE LA MUESTRA: Exclusiones y ampliaciones de la muestra

**¡Cuidado con las exclusiones!
¡Algunas pueden
ser importantes!**

**Ampliación de
muestras
(por algún aspecto de
especial interés)**



SELECCIÓN DE LA MUESTRA: características generales

El diseño de la muestra es complejo y ello incide en el cálculo de los estimadores (**Pesos**: variables de ponderación)

Muestra **representativa del conjunto de estudiantes**. Procedimiento probabilístico para evitar sesgos

✓ Muestreo **estratificado**

E. explícita: selección en diferentes subpoblaciones o **estratos** predefinidos (CCAA, titularidad de los centros, etc.)

E. implícita: asociada a un muestreo sistemático. La muestra debe reproducir la distribución de los porcentajes de la población.

✓ Selección **sistemática** dentro de cada estrato

Muestreo en dos etapas

1.
Selección
centros
(PSU)

- Muestreo proporcional al tamaño del centro

2.
Selección
alumnos
dentro del
centro
(SSU)

- 42 alumnos elegidos aleatoriamente

Procedimiento sistemático para seleccionar los centros

1. **Ordenar** los centros por tamaño (dentro de cada estrato)

2. Determinar el **intervalo de muestreo**:

$$I_e = \frac{N}{n_c}$$

3. Elegir al azar un número del intervalo $c \in (0,1)$

4. Multiplicar $c \cdot I_e$

5. Se elige **el primer centro** que contenga, de **forma acumulada**, $[c \cdot I_e]$ estudiantes

6. El **segundo centro y siguientes** son los que contengan (acumulados) $[c \cdot I_e] + I_e, [c \cdot I_e] + 2I_e, [c \cdot I_e] + 3I_e$, etc.

Ejemplo



POBLACIÓN N = 180
 MUESTRA n = 12
 Elegir 3 centros /
 4 alumnos por centro

Probabilidad
del centro

$$\frac{15}{180} \cdot 3 = \frac{1}{4}$$

15 alumnos



30 alumnos



45 alumnos



Probabilidad
del alumno en
el centro

$$\frac{4}{15}$$



$$\frac{4}{30} = \frac{2}{15}$$



$$\frac{4}{45}$$



Probabilidad
final
del alumno

$$\frac{1}{4} \cdot \frac{4}{15} = \frac{1}{15}$$

$$\frac{1}{2} \cdot \frac{2}{15} = \frac{1}{15}$$

$$\frac{3}{4} \cdot \frac{4}{45} = \frac{1}{15}$$

Peso
base
alumno

15

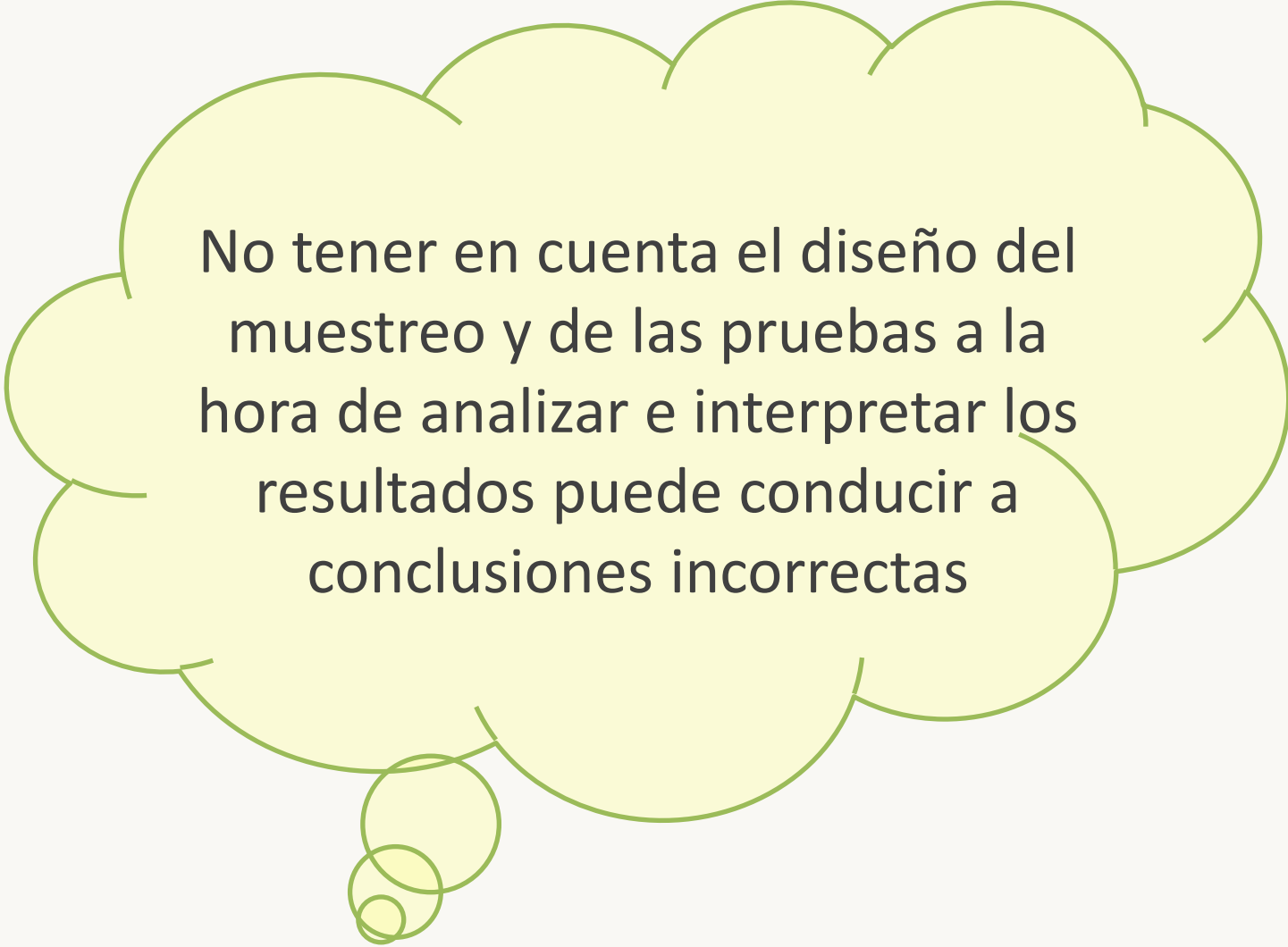
15

15

inee

La suma de pesos de los alumnos es el tamaño de la población

$$\sum \text{pesos} = 12 \cdot 15 = 180$$



No tener en cuenta el diseño del muestreo y de las pruebas a la hora de analizar e interpretar los resultados puede conducir a conclusiones incorrectas

Evaluaciones educativas, finalidad.

Marco de evaluación.

Diseño de las pruebas.

Diseño muestral.

Análisis de datos y presentación de resultados.

Teoría clásica de los tests

- Puntuación del alumno según el porcentaje de respuestas correctas a la prueba.
- Dificultad del ítem según el porcentaje de alumnos que lo responde correctamente.
- Dependencia de los estadísticos del test de la muestra (población)
- Dependencia de las puntuaciones de un alumno del conjunto particular de ítems incluidos en la prueba.

Teoría de respuesta al ítem

- En la misma escala se estima la puntuación del alumno y la dificultad de cada ítem.
- La dificultad del ítem no depende del grupo de alumnos.
- Supuestos: unidimensionalidad e independencia local. La probabilidad de contestar correctamente a un ítem es una función monótona no decreciente de la habilidad.

TEORÍA DE RESPUESTA AL ÍTEM (TRI)

Se calcula:

- > Dificultad de cada ítem.
- > Discriminación de cada ítem.
- > Estadísticos de ajuste de los ítems.

Modelo:

- > Modelo de Rasch para ítems dicotómicos.
- > Modelo de crédito parcial (Masters) para ítems de respuesta 0-1-2.

Teoría de la Respuesta al Ítem (TRI)

Dificultad de un ítem

Relación con la probabilidad de respuesta correcta, dada la “habilidad” del individuo

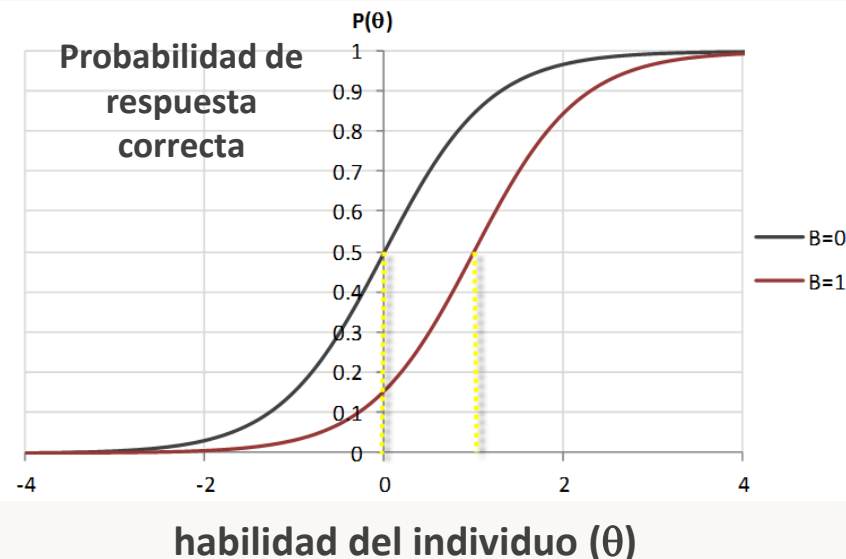
“Habilidad” del individuo

Relación con la verosimilitud de dar una respuesta correcta dada la dificultad del ítem

La respuesta a un ítem depende de la interacción entre la “habilidad” de alumno y la dificultad del mismo.

No se obtienen puntuaciones ligadas a una población en particular o a un grupo de individuos concretos, sino sobre la relación entre la “habilidad” del individuo y la dificultad de la tarea

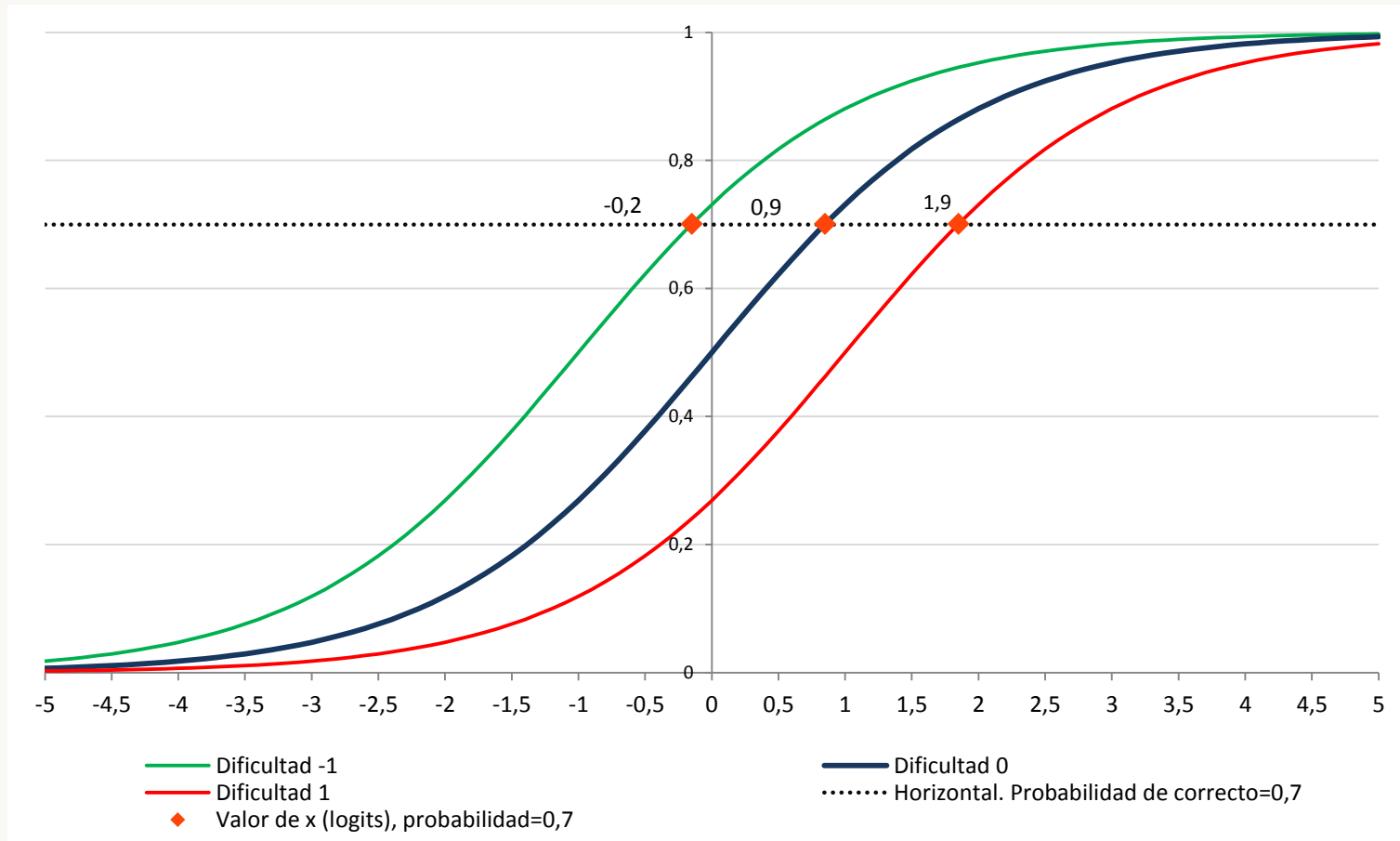
Curva característica del ítem (modelo de un parámetro)



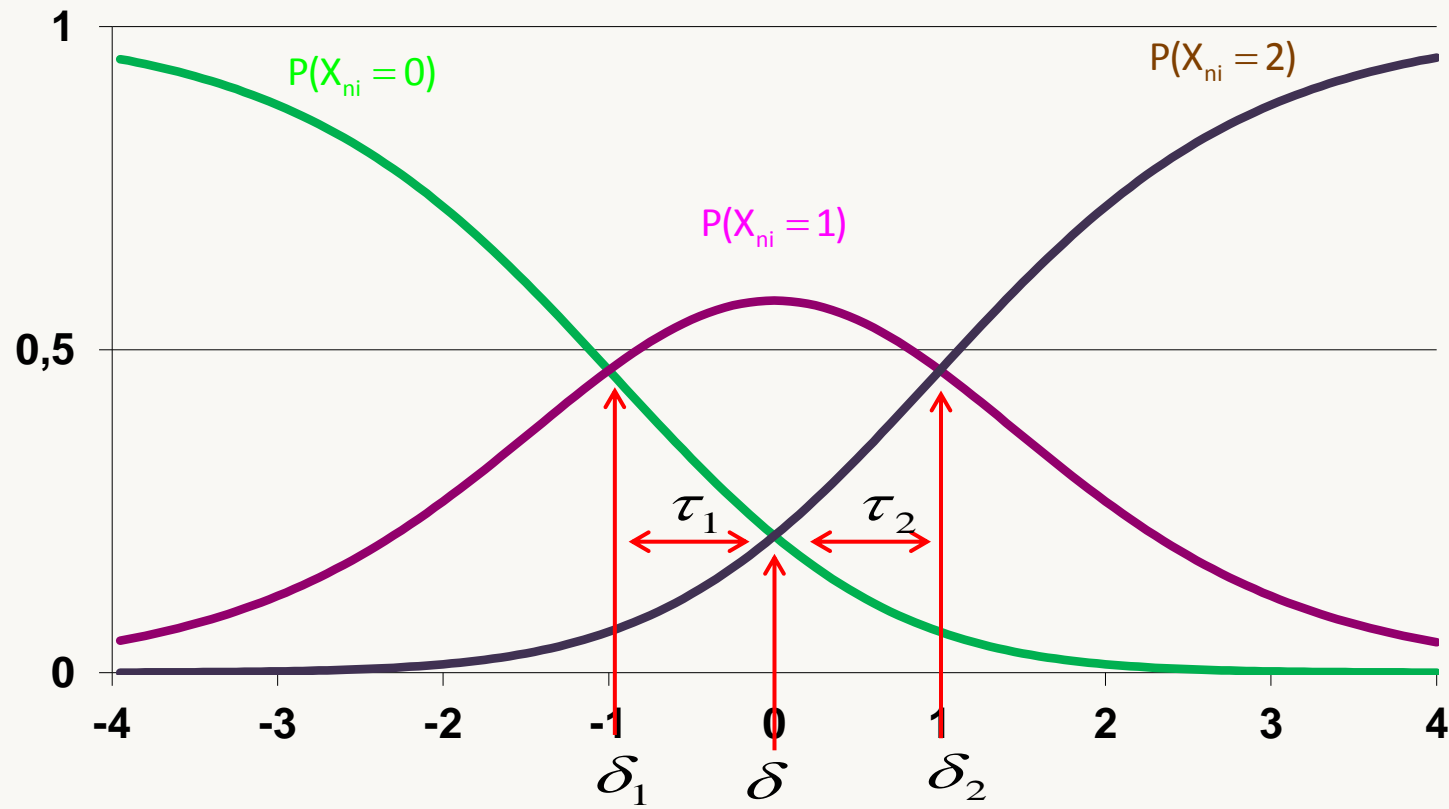
MODELO DE RASCH

Modelo logístico de un parámetro

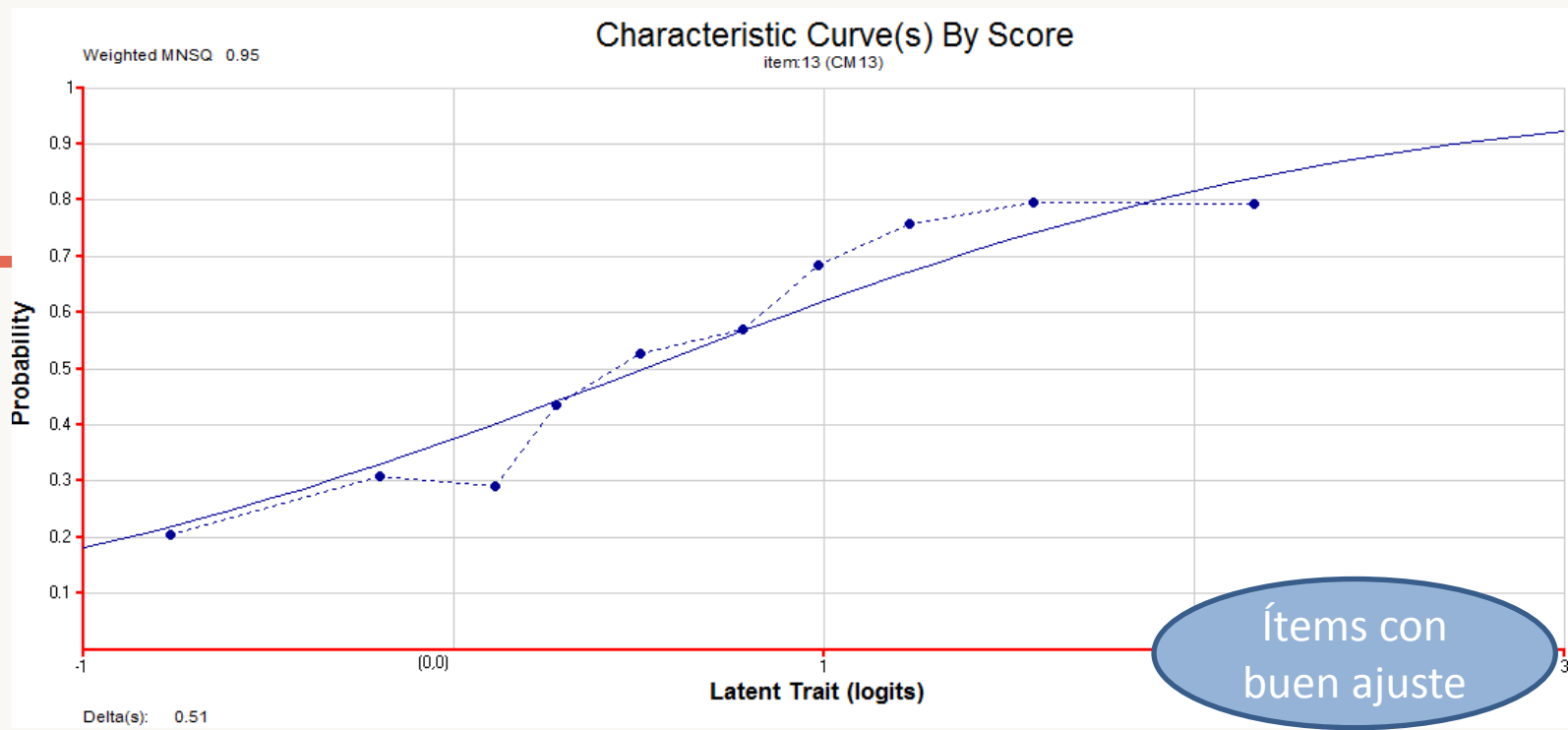
Curva característica del ítem



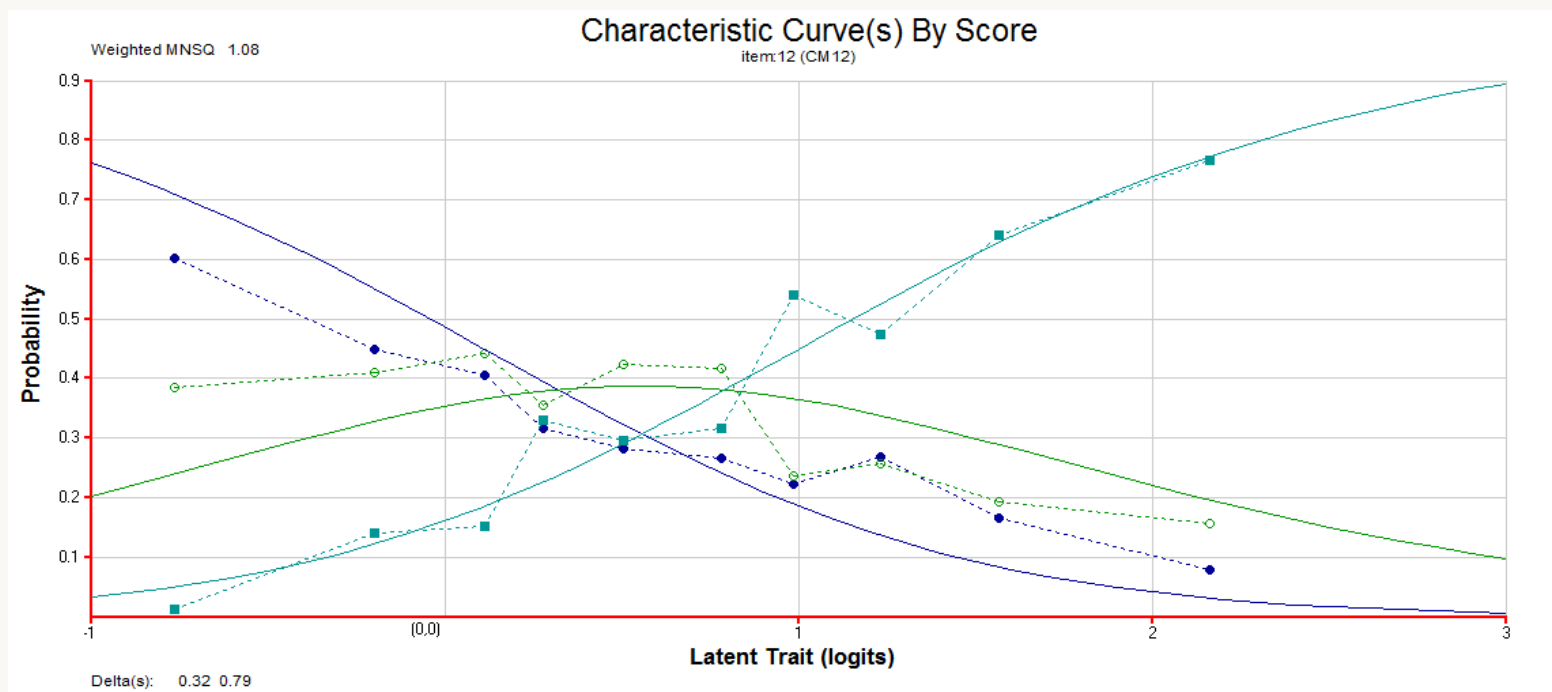
Curva característica del ítem



Tipo 0-1



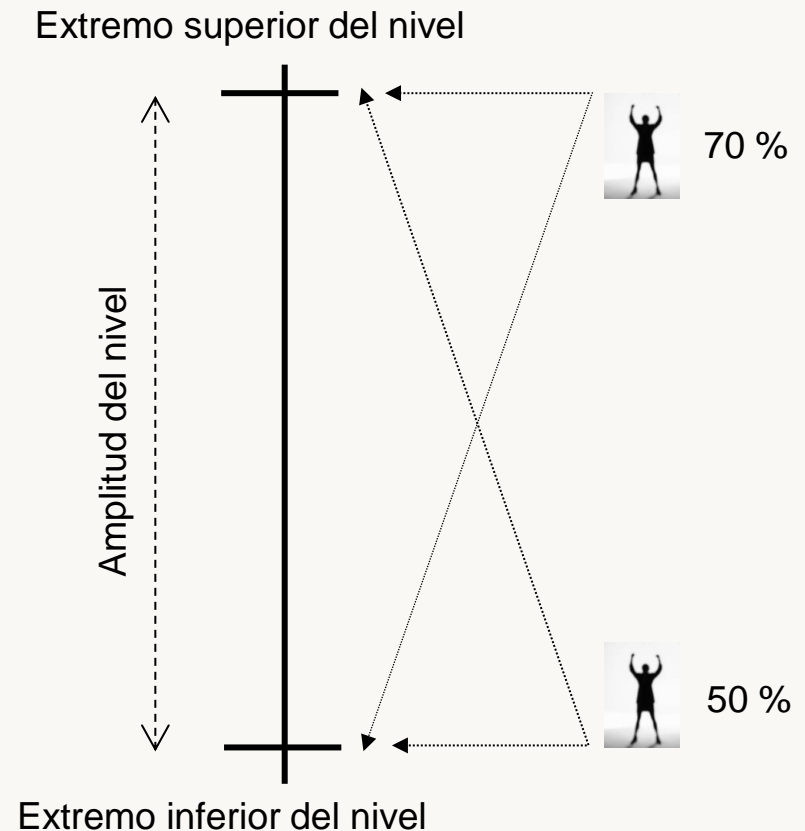
Tipo 0-1-2



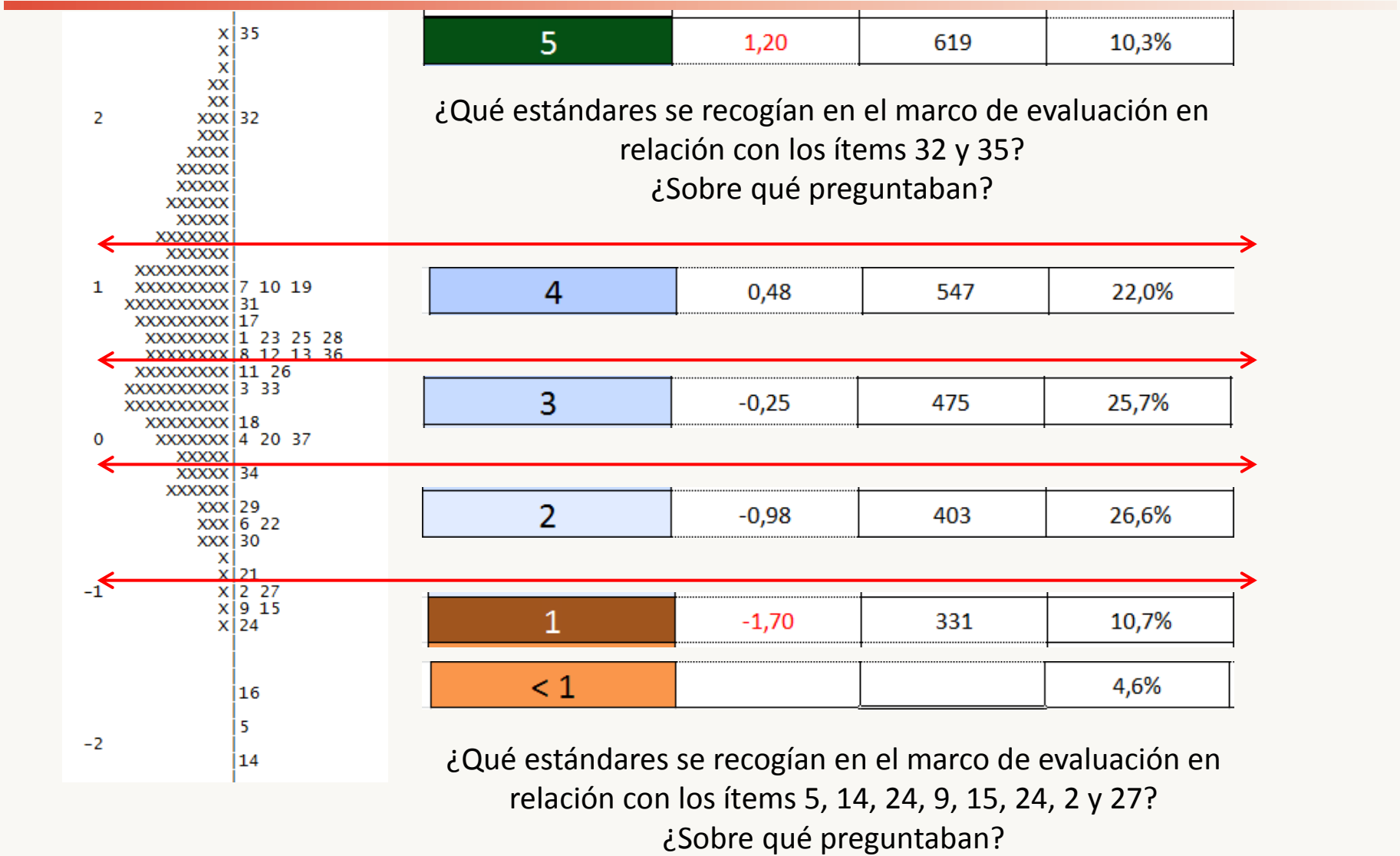
PUNTOS DE CORTE PARA OBTENER LOS NIVELES:

Combinación de análisis cuantitativo y cualitativo

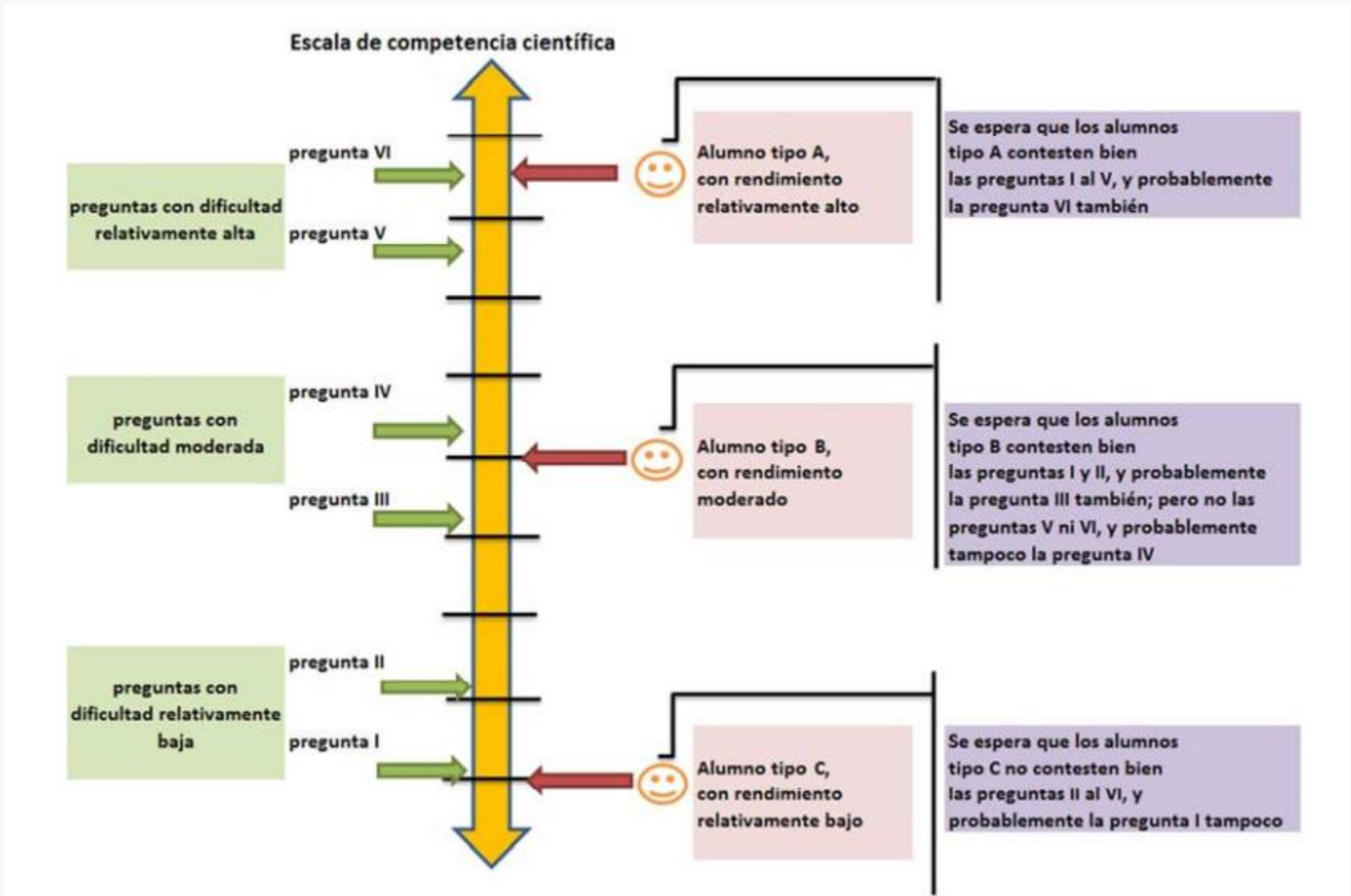
- Un alumno con habilidad en el **extremo inferior** del nivel tiene al menos una probabilidad p_1 de responder correctamente a los ítems de los niveles inferiores
- Un alumno con habilidad en el **extremo superior** del nivel tiene al menos una probabilidad p_2 de responder correctamente a los ítems del nivel (y, por supuesto, de los niveles inferiores)
- Por comodidad, se suele fijar:
 $p_1 = 0,5$
 $p_2 = 0,7$



DESCRIPCIÓN DE NIVELES



Escala de rendimiento



Valores plausibles

A cada estudiante que ha participado en las pruebas se le asignan unas puntuaciones (5 o 10) en cada competencia evaluada, que se denominan **valores plausibles**.

Se pueden describir como una representación del **rango de habilidades** que un estudiante puede, razonablemente, tener.

Para cada estudiante se obtiene una distribución de probabilidad del estimador de sus habilidades. Los valores plausibles son una selección aleatoria de valores de dicha distribución.

Son estimadores insesgados de los parámetros poblacionales de interés (y en los estratos considerados).

RESULTADOS DE LA EVALUACIÓN



¿EN QUÉ SE TRADUCE ESTO A MENUDO?

- En pensar que **la evaluación se limita a un Ranking**
- En que se **puede convertir en un arma arrojadiza**
- En **que se genere "mala prensa"** sobre la educación española para la opinión pública
- En un **cierto malestar entre los profesores**: “¿qué hemos hecho mal?”



¿Son de fiar los resultados del Informe PISA en Andalucía?

EL DIARIO VASCO .COM

vocento5 | oferplan | Entradas | Kirolprobak | Promociones DV | Gastronomía | Blogs | Esquejas | Directorio de empresas

GIPUZKOA SOCIEDAD POLÍTICA ECONOMÍA MUNDO REAL SOCIEDAD DEPORTES CULTURA PLANES GENTE TECNOLOGÍA

SOCIEDAD

SOCIEDAD EDUCACIÓN

Euskadi suspende en PISA

En Lectura también se sitúa por debajo de la media española y de la OCDE. Sólo en Matemáticas obtiene el aprobado

MARTA FDEZ. VALLEJO

6 diciembre 2016 18:43

50



Los resultados de la evaluación internacional PISA muestran un retroceso de Euskadi en las tres materias a examen,



VÍDEO ▶

LA VANGUARDIA

Vida Nat

CATALUNYA

EDUCACIÓN

Cataluña enciende la alarma

El sistema ca

En los exámenes

SOCIEDAD

Info a la Cier Lect

PISA 2015 la parte b diferencia

6 diciembre,

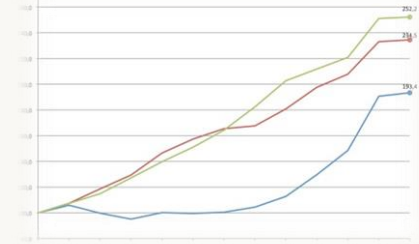
Mejora en los re

• Qué tener educ finan

De esta forma en la línea Alemania, Bélgica, Irlanda Reino Unido Dinamarca Australia,

Según la comparación de los resultados de la orientación educativa de los últimos años en el Informe Educación compromiso de dirección.

¿CUÁLES SON LOS OBJETIVOS?



Proporcionar indicadores de la **eficacia, la equidad y la eficiencia** de los sistemas educativos

Establecer puntos de referencia para la comparación nacional/internacional

Seguir la evolución de los datos en el tiempo

Para alcanzar estos objetivos no solo **se necesitan medidas fiables y válidas** para evaluar el rendimiento cognitivo, **sino también información sobre qué variables influyen en dichos resultados.**

Es importante explicar los resultados cognitivos mediante variables de contexto.
Principales factores asociados al rendimiento

¿CÓMO SE CONSIGUE ESTO?



Se elaboran los **cuestionarios de contexto**, complementarios de las pruebas cognitivas



La profesora o el profesor conoce bien el **contexto individual y social** de sus alumnos, en cada grupo, en cada centro escolar



Los responsables de política educativa (*policy makers*) tienen que **conocer el contexto de la población** estudiantil que está evaluando a nivel del sistema educativo

EN RESUMEN: ¿QUÉ ES REALMENTE UNA EVALUACIÓN?



¡¡Muchas gracias!!

+ info: <https://www.mecd.gob.es/inee/portada.html>

Síguenos en Twitter:
@educaINEE

En nuestro blog:
<http://blog.educalab.es/inee/>

