



# **Fundamentos de la teoría de respuesta al ítem**

**Satoshi Kusaka**

Universidad de Educación de Naruto

Hoy les hablaré acerca de los fundamentos de la teoría de respuesta al ítem.

Me llamo Kusaka, de la Universidad de Educación de Naruto.

Mucho gusto.

## Presentación personal

### Satoshi Kusaka

Universidad de Educación de Naruto, Curso de educación global  
Catedrático (Educación en Matemáticas)



2002–2004	Enseñanza de matemáticas en la República Dominicana como voluntario para la cooperación japonesa en el extranjero
2004–2005	Maestro de primaria en la Primaria Municipal de Ogino, Atsugi, prefectura de Kanagawa
2008–2011	Maestro de la Corporación Educacional New International School
2011–2021	Participo en proyectos educativos de JICA como experto en educación en matemáticas
2021–actualidad	Catedrático de la Universidad de Educación de Naruto

#### Pasatiempos:

- ✓ Viajar (ya sea dentro o fuera del país)
- ✓ Estudiar programación (estudiando desde cero)
- ✓ Resolver problemas de matemáticas de bachillerato (no me gustaba en esa época, pero ahora sí)

2

Me llamo Satoshi Kusaka, de la Universidad de Educación de Naruto. Mi especialidad es la educación matemática. A la fecha, tengo experiencia de trabajar como docente y he participado como experto en educación en matemáticas en los proyectos educativos de JICA en diversos países. Desde abril de 2021, soy catedrático de la Universidad de Educación de Naruto. Mucho gusto.

## **Contenido**

1. Los límites de la teoría clásica de los tests (TCT)
2. Qué es la teoría de respuesta al ítem (TRI)
3. Métodos de pruebas con la TRI
4. Curva característica de un ítem (CCI)
5. Desarrollo de un banco de ítems
6. Hacia la implementación de la TRI
  - Desafíos
  - Aplicación

El contenido es el siguiente. En primer lugar, reflexionaremos de forma práctica acerca de las limitaciones de la teoría de los tests hasta ahora. A partir de eso, explicaré acerca de la «teoría de respuesta al ítem (TRI)» y los «métodos de pruebas con la TRI». En la teoría de respuesta al ítem, la curva característica de un ítem (CCI) es sumamente importante. Lo explicaré detalladamente mediante ejercicios. Para finalizar, abordaremos también los desafíos para adoptar y aplicar la TRI.

## 1. Las limitaciones de la teoría clásica de los tests



### Problema:

Supongamos que el alumno de la secundaria A da la prueba X y el de la secundaria B da la Y (las pruebas X e Y son de la misma asignatura y contenidos, pero las preguntas son diferentes). En tal caso,

- P1. Pedro, de la secundaria A obtiene 70 puntos y Ana, de la B, también obtiene 70. ¿Tendrán ambos el mismo nivel de rendimiento académico? ¿Por qué?
- P2. El promedio para la prueba X fue de 60 puntos y para la prueba Y, también fue de 60 puntos. ¿Se podría afirmar que el grado de dificultad es igual en ambas pruebas? ¿Por qué?

Vamos a pensar acerca de las limitaciones de la teoría clásica de los tests. Veamos este ítem. «Supongamos que el alumno de la secundaria A da la prueba X y el de la secundaria B da la Y (las pruebas X e Y son de la misma asignatura y contenidos, pero las preguntas son diferentes)». En tal caso, «P1. Pedro, de la secundaria A obtiene 70 puntos y Ana, de la B, también obtiene 70. ¿Tendrán ambos el mismo nivel de rendimiento académico? ¿Por qué?»

Tienen 30 segundos para pensarlo.

Preguntar a dedo a algunos participantes para que respondan

Cierto. Es «no se puede afirmar que ambas personas tengan el mismo nivel de rendimiento académico.», porque las pruebas X e Y no necesariamente tienen el mismo grado de dificultad. Dado que las preguntas en ambas pruebas son diferentes, nada garantiza que los 70 puntos en la prueba X y los 70 en la Y representen un nivel de capacidad equivalente.

Luego, «P2. El promedio para la prueba X fue de 60 puntos y para la prueba Y, también fue de 60 puntos. ¿Se podría afirmar que el grado de dificultad es el mismo en ambas pruebas? ¿Por qué?»

Tienen 30 segundos para pensarlo.

Preguntar a dedo a algunos participantes para que respondan

Bien. Es «no se puede afirmar que ambas pruebas tengan el mismo grado de dificultad», porque nada garantiza que las escuelas A y B tengan el mismo nivel de aptitud académica. Por ejemplo, en la escuela A podrían alcanzar quizás un promedio de 80 puntos en la prueba Y. (Esto cuando el nivel de aptitud académica de la escuela A es más alto en comparación al de la escuela B).

**Respuesta:**

- R1. Es «no se puede afirmar que ambas personas tengan el mismo nivel de rendimiento académico», porque las pruebas X e Y no necesariamente tienen el mismo grado de dificultad. Dado que las preguntas en ambas pruebas son diferentes, nada garantiza que los 70 puntos en la prueba X y los 70 en la Y representen un nivel de habilidad equivalente.
- R2. Es «no se puede afirmar que ambas pruebas tengan el mismo grado de dificultad», porque nada garantiza que las escuelas A y B tengan el mismo nivel de aptitud académica. Por ejemplo, en la escuela A podrían alcanzar quizás un promedio de 80 puntos en la prueba Y. (Esto cuando el nivel de aptitud académica de la escuela A es más alto en comparación al de la escuela B).



**A partir de esto, pensemos qué se puede afirmar.**

Tenemos aquí las respuesta, lo mismo que ustedes acaban de decir. Entonces, ¿qué se puede afirmar a partir de esto?

De estos ejemplos se desprende que:

✓ **La evaluación del grado de dificultad de la prueba depende del nivel del grupo que rinda la prueba.**

Si el grupo examinado tiene un nivel alto, la puntuación individual de los examinados será alta y el promedio general del grupo, también. Al contrario, si el grupo examinado tiene un nivel bajo, la puntuación individual de los examinados será baja y el promedio general del grupo, también. Es decir, una misma prueba se considera fácil para un grupo de nivel alto y difícil para uno de nivel bajo. Esto se denomina dependencia grupal (group dependence) del puntaje de la prueba o dependencia de la muestra (sample dependence).

✓ **El puntaje individual de los examinados depende del grado de dificultad de los ítems de la prueba.**

Si los ítems de la prueba son fáciles en general, el punto obtenido será alto, y al contrario, será bajo si son difíciles. Esto se denomina dependencia de los ítems (item dependence).

Cuando los grupos examinados y los ítems de las pruebas son distintos, sus puntajes no se pueden comparar directamente.

**Para poder compararlos, es necesario separar mutuamente la información que depende de los examinados y la que depende de las pruebas.**

De estos ejemplos, se desprende que la evaluación del grado de dificultad de la prueba depende del nivel del grupo que rinda la prueba. Si el grupo examinado tiene un nivel alto, la puntuación individual de los examinados será alta y el promedio general del grupo, también. Al contrario, si el grupo examinado tiene un nivel bajo, la puntuación individual de los examinados será baja y el promedio general del grupo, también. Es decir, una misma prueba se considera fácil para un grupo de nivel alto y difícil para uno de nivel bajo. Esto se denomina dependencia grupal (group dependence) del puntaje de la prueba o dependencia de la muestra (sample dependence). Además, el puntaje individual de los examinados depende del grado de dificultad de los ítems de la prueba. Si los ítems de la prueba son fáciles en general, el punto obtenido será alto, y al contrario, será bajo si son difíciles. Esto se denomina dependencia de los ítems (item dependence). A causa de esta «dependencia grupal del puntaje de la prueba» y «dependencia de los ítems», cuando los grupos examinados y los ítems de las pruebas son distintos, sus puntajes no se pueden comparar directamente.

Para poder compararlos, es necesario separar mutuamente la información que depende de los examinados y la que depende de las pruebas.

## 2. Qué es la teoría de respuesta al ítem (TRI)

**La teoría de respuesta al ítem permite expresar por separado el grado de dificultad de los ítems incluidos en la prueba y la capacidad de los examinados por dicha prueba.**

En la teoría clásica de los tests (TCT), cuando la puntuación de las pruebas se expresa como puntaje bruto no es posible distinguir si el promedio alto ante una prueba se debe a la facilidad de las preguntas o a la excelencia de los examinados. En la TCT resulta imposible separar el grado de dificultad de los ítems de la habilidad de los examinados. Esto sí puede hacerse en la TRI, al usar la **curva característica del ítem (CCI) (item characteristic curve: ICC)**.

Así,

- ✓ **Se puede distinguir entre preguntas malas y buenas, calculando la dificultad y la discriminación para cada tipo de pregunta.**  
Al indicar con valores si cada pregunta mide adecuadamente o no las habilidad de los estudiantes, permite distinguir entre preguntas buenas y malas. Es decir, permite presentar en la prueba definitiva únicamente las preguntas estadísticamente demostradas como buenas.
- ✓ **Permite comparar el puntaje de los alumnos, aunque las preguntas planteadas sean distintas.**  
Al crear un banco de ítems con preguntas de niveles ya igualados, se pueden comparar las capacidades sin importar de los ítems usados en las pruebas. Es decir, se pueden crear varios juegos de preguntas del mismo nivel.
- ✓ **Resulta posible realizar una «prueba adaptativa» que consista en ítems acordes al nivel de habilidad de los examinados.**

A partir de aquí, voy a explicar la teoría de respuesta al ítem. La teoría de respuesta al ítem permite superar las limitaciones ya mencionadas de la teoría clásica de los tests. Es decir, la teoría de respuesta al ítem permite expresar por separado el grado de dificultad de los ítems incluidos en una prueba y la capacidad de los examinados con dicha prueba. En la teoría clásica de los tests (TCT), cuando la puntuación de una prueba se expresa como puntaje bruto no es posible distinguir si el promedio alto ante una prueba se debe a la facilidad de las preguntas o a la excelencia de los examinados. En la teoría clásica de los tests resulta imposible separar el grado de dificultad de los ítems de la capacidad de los examinados. Sin embargo, en TRI, sí se puede hacerlo al usar la curva característica de un ítem (CCI). Posteriormente explicaremos en detalle cómo interpretar la curva característica de un ítem, porque es sumamente importante. Ahora explicaremos qué se puede hacer con la teoría de respuesta al ítem. En primer lugar, resulta posible distinguir entre preguntas malas y buenas, calculando la dificultad y la discriminación para cada tipo de pregunta. Al indicar con valores si cada pregunta mide adecuadamente o no las capacidades de los estudiantes, permite distinguir entre preguntas buenas y malas. Es decir, se pueden presentar en la prueba definitiva únicamente



las preguntas estadísticamente demostradas como buenas.

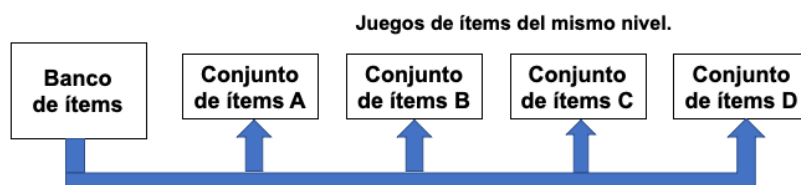
En segundo lugar, resulta posible comparar el puntaje de los alumnos, aunque las preguntas planteadas sean distintas. Como expliqué anteriormente, la «discriminación» y la «dificultad» se puede expresar en valores, separando la capacidad de los examinados. Por eso, se puede comparar la capacidad de los examinados aunque se les presenten distintos ítems con la misma «discriminación» y «dificultad». Es decir, si se preparan varios ítems de la misma «discriminación» y «dificultad» se pueden crear varios juegos de preguntas del mismo nivel.

En tercer lugar, resulta posible realizar una «prueba adaptativa» consistente en ítems acordes al nivel de capacidad del grupo específico de sujeto, dado que se conoce de antemano la dificultad de los ítems.

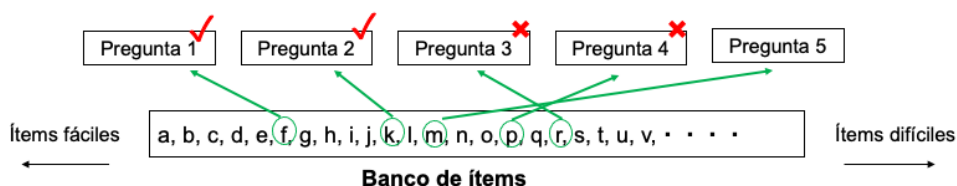
En las siguientes diapositivas explicaré en detalle el segundo y tercer punto.

### 3. Métodos de pruebas con la TRI

#### ➤ Realización de varias pruebas del mismo nivel



#### ➤ Computer adaptive test (Prueba adaptativa computarizada)

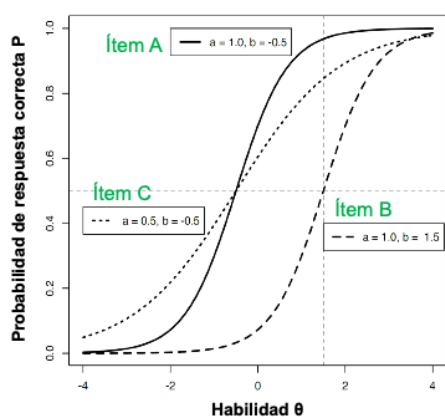


Este diagrama es un esquema que representa la realización de varias pruebas del mismo nivel, mencionada en el segundo punto de la diapositiva anterior. A partir de los valores de « discriminación» y « dificultad» calculados previamente mediante la TRI resulta posible crear juegos de ítems diferentes pero del mismo nivel.

Esta figura corresponde a la «prueba adaptativa consistente en ítems acordes al nivel de capacidad de los examinados» mencionado en el punto 3 de la diapositiva anterior. A partir de la « discriminación» y « dificultad» calculadas previamente los ítems se ordenan desde los más fáciles a los más difíciles. En la primera pregunta se plantea el ítem F y si es respondido correctamente, para la pregunta 2 se plantea el ítem k que es algo más difícil. Si este también es respondido correctamente, para la pregunta 3 se plantea el ítem r que es a su vez algo más difícil. Si es respondido equivocadamente, se plantea el ítem p que es un poco más fácil. Si este también es respondido equivocadamente, se plantea el ítem m que es a su vez un poco más fácil. Entonces así se puede determinar dónde se encuentra más o menos la capacidad del estudiante.

Es el mismo concepto de un examen visual, en el que se determina la capacidad visual probando con figuras más grandes o más pequeñas. Las figuras grandes a pequeñas que se usan en el examen visual corresponden a un valor definido para expresar la capacidad visual según las figuras que nos sea posible distinguir. De esa forma es posible cuantificar la capacidad visual. Del mismo modo, cada ítem de las pruebas que aplican la TRI poseen valores de «discriminación» y «dificultad». En las siguientes diapositivas lo explicaremos en detalle.

#### 4. Curva característica del ítem (CCI) (item characteristic curve: ICC)



Ítem	Discriminación	Dificultad
A	1,0	-0,5
B	1,0	1,5
C	0,5	-0,5

Toma la capacidad  $\theta$  del examinado en el eje horizontal, y la probabilidad de respuesta correcta al ítem en el eje vertical. De esta forma, se puede definir un grado de dificultad de los ítems acorde a la capacidad de los examinados.

##### Dificultad de los ítems (b)

Representa la dificultad de las preguntas.  $\theta$  es el valor de la capacidad de los examinados con una probabilidad de respuesta correcta de 0,5. Por ejemplo, en el ítem A, cuando la probabilidad de respuesta correcta  $P(\theta)$  es 0,5, el valor de capacidad es -0,5; es decir, el grado de dificultad es -0,5.

##### Discriminación de los ítems (a)

Es la habilidad para distinguir la capacidad de los examinados. Es un valor característico (0,2 a 2 aprox.) proporcional a la inclinación de la tangente de la CCI en una probabilidad de respuesta correcta de 0,5.

Por ejemplo, en el ítem B, dado que la inclinación de la tangente cuando la probabilidad de respuesta correcta  $P(\theta)$  es 0,5 la discriminación del ítem es 1,0. La probabilidad cambia considerablemente.

✓ Por ejemplo, la probabilidad de respuesta correcta para el ítem A donde  $\theta = -2$  y  $\theta = 1,5$  es de 0 % y 100 % aprox. respectivamente, por lo que podemos decir que se discrimina la capacidad de casi todos los examinados.



Pensemos:

- ✓ Las similitudes y diferencias entre los ítems A y B, además de sus motivos.
- ✓ Las similitudes y diferencias entre los ítems A y C, además de sus motivos.

Ahora explicaré la curva característica del ítem, que es la clave de la TRI. Primero, veamos esta figura. El eje horizontal es la aptitud de los examinados. Se expresa en valores entre -4 y 4. El eje vertical es la probabilidad asociada de contestar correctamente al ítem.

Por ejemplo, la curva del ítem B muestra que el porcentaje de respuestas correctas para los examinados con una habilidad de 1,5 es de 0,5. Esto significa que la mitad de los candidatos con habilidad 1,5 pueden responder correctamente al ítem. A medida que la habilidad se desplaza hacia la derecha, la probabilidad de respuestas correctas también aumenta.

Aquí explicaré acerca de las dos expresiones que ya he mencionad varias veces, «dificultad» y «discriminación». Estos dos valores son sumamente importantes en la TRI y son los que definen las características de un ítem. El primero es la dificultad del ítem, que, como el término sugiere, es un valor que indica la dificultad de la pregunta. El valor de la habilidad de los examinados con una probabilidad de respuesta correcta de 0,5 corresponde a la dificultad. Por ejemplo, para los ítems A y C, cuando la probabilidad de respuesta correcta es 0,5, el valor de la habilidad es -0,5, por lo que la dificultad del ítem es -0,5. En cuanto al ítem B, el 50% de los

examinados responde correctamente con un valor de habilidad de 1,5. Esto significa que el valor de la habilidad es 1,5 cuando la probabilidad de respuesta correcta es 0,5, por lo que la dificultad del ítem es 1,5.

A continuación explicaré la discriminación del ítem, que consiste en la capacidad para discernir la habilidad de los examinados. No se pueden considerar buenas preguntas aquellas que pueden responder tanto sujetos con bajo nivel de rasgo como sujetos con altos niveles de rasgo, ni tampoco aquellas en que los sujetos que tienen un nivel alto de habilidad se equivoquen y los aquellos que tienen un bajo nivel respondan correctamente. Los ítems buenos son aquellos en que el porcentaje de respuesta correcta sube abruptamente desde un punto de la escala de habilidad. Esa es la discriminación.

La discriminación se representa con la pendiente de la tangente de la curva característica de un ítem en que la probabilidad de respuesta correcta es 0,5. En teoría, dado que se trata de una tangente, sería hasta infinito negativo y positivo, pero como los datos se equiparan y ajustan, el rango está entre 0,2 y 2 aproximadamente.

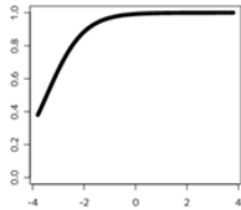
Por ejemplo, entre los ítems A y C, el A tiene una pendiente mayor cuando la probabilidad de respuesta correcta es 0,5, por lo que se puede decir que la discriminación es mayor en el ítem A. De hecho, la discriminación del ítem A es 1,0 y del C, 0,5.

Asimismo, si observamos la curva característica del ítem A, la probabilidad de respuesta correcta donde  $\theta = -2$  y  $\theta = 1,5$  es de 0 % y 100 % aprox. respectivamente, por lo que podemos decir que se discrimina la habilidad de casi todos los examinados. Indica que los examinados de habilidad inferior o igual a -2 no son capaces de responder correctamente, y los de habilidad superior o igual 1,5 pueden casi todos responder correctamente. Entonces, para comprobar la explicación hasta aquí, ahora les daré cerca de un minuto para que piensen acerca de las «similitudes y diferencias entre los ítems A y B, además de sus motivos», así como las «similitudes y diferencias entre los ítems A y C, además de sus motivos».

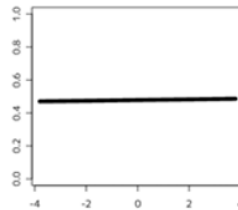


¿Qué particularidades podríamos decir que tienen respectivamente las siguientes curvas características del ítem?

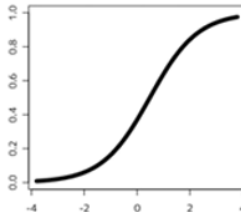
A.



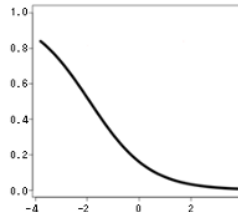
B.



C.



D.



Profundicemos más en las curvas características del ítem. ¿Qué particularidades podríamos decir que tienen respectivamente las siguientes curvas características del ítem de A a D? Tienen 1 minuto para pensarlo.

En A, incluso los sujetos con más bajo nivel de rasgo -4 responden correctamente en un 40 % aproximadamente. Podemos decir que es un ítem sumamente fácil dado que en cerca de la habilidad -1,5 la probabilidad de respuesta correcta es casi del 100 %.

En B, la probabilidad de respuesta correcta no cambia en absoluto en la habilidad entre -4 y 4. Es decir, la probabilidad de respuesta correcta es la misma tanto para sujetos de nivel bajo como sujetos con algo nivel.

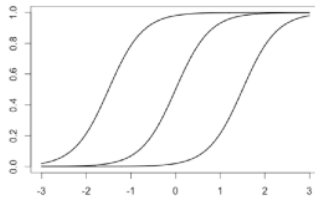
En D, el porcentaje de respuesta correcta de los sujetos con nivel bajo de rasgo es alto, y al revés, el de los sujetos con nivel alto es bajo.

En C, el porcentaje de respuesta correcta sube abruptamente en torno a la habilidad 0,5. Este es el tipo de curva ideal, con el que podemos decir que el ítem es una buena pregunta.

## El modelo logístico de la CCI

### Modelo logístico de 1 parámetro (1PLM)

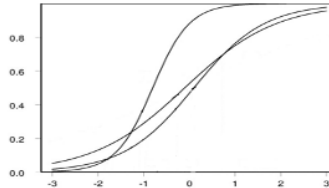
$$P_j(\theta) = \frac{1}{1 + \exp(-1.7a(\theta - b_j))}$$



Modelación de la dificultad únicamente. Se define el nivel de dificultad para cada ítem, pero el parámetro de discriminación se considera equivalente para todos los ítems.

### Modelo logístico de 2 parámetros (2PLM)

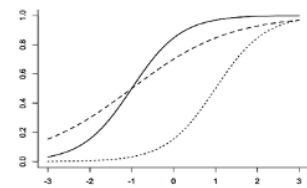
$$P_j(\theta) = \frac{1}{1 + \exp(-1.7a_j(\theta - b_j))}$$



Modelo que representa la probabilidad de respuesta correcta de los ítems de una prueba mediante la dificultad y la discriminación de los ítems.

### Modelo logístico de 3 parámetros (3PLM)

$$P_j(\theta) = c_j + \frac{1 - c_j}{1 + \exp(-1.7a_j(\theta - b_j))}$$



Modelo que aparte de la dificultad y la discriminación de los ítems, incorpora un parámetro de probabilidad de acertar al azar mediante adivinación.

- ✓ El 3PLM es el que entrega mayor información. El 1PLM es el más fácil de interpretar.
- ✓ Con respecto a la «cantidad mínima de muestras» necesaria, se considera que se puede realizar un cálculo adecuado con entre 100 y 200 para el 1PLM, 200 a 400 para el 2PLM y 1000 a 2000 para el 3PLM.

Ahora explicaré brevemente acerca de los modelos de curvas características del ítem. Hay 3 modelos de curva característica del ítem, el de un parámetro (1PLM), el de dos parámetros (2PLM) y el de tres parámetros (3PLM).

Las tres fórmulas tienen diferencias y aspectos en común entre sí. Los aspectos distintos corresponden a las diferencias entre cada modelo. exp representa la función exponencial. Veamos lo que sigue a exp. En el 1PLM y el 2PLM la representación de a es diferente. En el 3PLM, se incorpora el aspecto c. A continuación explicaré cada uno de ellos.

El modelo logístico de 1 parámetro (1PLM) es una modelación del índice de dificultad únicamente. a representa la discriminación y b la dificultad. El 1PLM no tiene en cuenta la discriminación a y únicamente calcula el valor del nivel de dificultad b. En la gráfica, la pendiente es igual en todas. Es decir, la discriminación es igual. El valor de habilidad cuando la probabilidad de respuesta correcta es 0,5; es decir el nivel de dificultad, es diferente en cada caso.

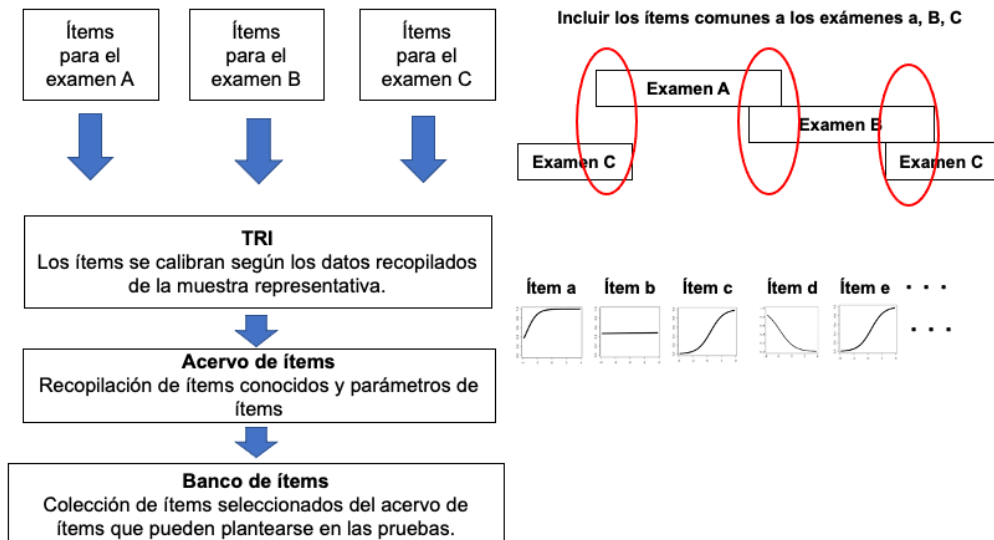
El modelo logístico de 2 parámetros (2PLM) es una modelación que usa el índice de discriminación y el nivel de dificultad. Hay variables tanto para  $a$  como para  $b$ . En las tres gráficas, cuando la probabilidad de respuesta correcta es 0,5 difieren el valor de la habilidad y la pendiente que indica la discriminación.

El modelo logístico de 3 parámetros (3PLM) es una modelación que además del índice de discriminación  $a$  y el nivel de dificultad  $b$ , agrega una nueva variable  $c$ . Esta  $c$  corresponde al valor de responder correctamente al azar cuando el nivel de habilidad es bajo, o cuando el nivel de habilidad es alto pero se equivocan por descuido. En la gráfica se aprecian ítems en que la probabilidad de respuesta correcta es cercana a 0,2 incluso para sujetos de nivel bajo como  $-3$ . Es algo que igual puede pasar, ¿verdad? El 3PLM es el que nos brinda mayor información. El 1PLM es el más fácil de interpretar. Ahora bien, con respecto a la «cantidad mínima de muestras» necesaria, se considera que se puede realizar un cálculo adecuado con entre 100 y 200 para el 1PLM, 200 a 400 para el 2PLM y 1000 a 2000 para el 3PLM.

Finalmente, quizás algunos de ustedes se pregunten de dónde viene ese 1,7 que aparece en las fórmulas. Originalmente, para expresar la curva característica del ítem se usaba la fórmula de integral definida de la distribución normal estándar. Posteriormente se descubrió que usando las fórmulas de este modelo se podía explicar lo mismo, por lo que ahora estas fórmulas de modelo se usan para la curva característica del ítem en general. Ese 1,7 resulta necesario para que el modelo tenga la misma fórmula que la integral definida de la distribución normal estándar.



## 5. Desarrollo de un banco de ítems



Ahora vamos a ver acerca del desarrollo de un banco de ítems, y cómo desarrollar y acumular ítems que contengan los parámetros de discriminación y de dificultad. En primero lugar, es necesario realizar varios tipos de prueba. Aquí lo importante es incluir los ítems en común A, B y C en las pruebas A, B y C respectivas. De esa forma, aunque los sujetos de las pruebas A y B sean totalmente distintos, habrán rendido algunos ítems que serán los mismos. No se puede usar la teoría de respuesta al ítem si no se cuenta con varios ítems superpuestos. La curva característica del ítem se traza poniendo los resultados en la TRI y dándole los valores de la discriminación y de la dificultad para cada ítem. El conjunto de estos se denomina acervo de ítems. Y aquí, se eliminan todas las preguntas inadecuadas. La colección de ítems seleccionados desde el acervo de ítems, que puedan plantearse en las pruebas, se denomina banco de ítems.

## 6. Hacia la adopción de la TRI

### **Desafíos**

#### **Ítems de examen**

- ✓ Se necesita preparar un gran banco de ítems de examen en que se gestione la calidad, el grado de dificultad, etc.
- ✓ Aunque se supone que las preguntas de los exámenes no son públicas y se reutilizan, es posible que se filtren una vez superada una cantidad dada de usos, por lo que es necesario reemplazar o añadir ítems de manera regular.
- ✓ Es decir, la confección de ítems aumenta considerablemente el personal, el tiempo y los gastos.

Además, que los ítems no se puedan publicar significa que:

- ✓ No se pueden usar ítems reales de los exámenes en los debates para mejorar la educación.
- ✓ No se puede presentar a los docentes clases y materiales de mejoramiento didáctico a partir de ítems reales de los exámenes.

#### **Aspectos concretos**

- ✓ Requiere que los exámenes se realicen en entornos con PC, redes, etc.

**Exige un nivel de implementación que supera con creces los exámenes e investigaciones de la aptitud académica realizados hasta ahora. Además, para coordinar lo antes mencionado, se requieren ingentes gastos y esfuerzos, incluyendo de personal.**

Ahora abordaremos los desafíos para implementar la TRI y las formas de utilizarla. Lo primero es lo relacionado con los ítems de examen. Se necesita preparar gran cantidad de ítems para exámenes, o sea un banco de ítems, en que se gestione la calidad, el nivel de dificultad, etc. En teoría, los ítems del examen no serán publicadas con la finalidad de reutilización, se deben hacer un control adecuado para evitar las filtraciones. Además, es necesario sustituir o añadir ítems con regularidad, ya que puede filtrarse si se utiliza el mismo ítem más de unas determinadas veces. Esto significa que la mano de obra, el tiempo y los costes necesarios para preparar los ítems aumentan considerablemente. Además, el hecho de que los ítems no se puedan publicar significa que no se pueden usar ítems de exámenes reales para debatir las mejoras educativas. En Japón, la evaluación nacional sobre el rendimiento académico y el estado de aprendizaje que se realiza todos los años publica sus ítems, que son usados por diversas entidades pertinentes en los debates destinados a mejorar la educación, las clases de los docentes y los materiales didácticos. Entre los aspectos físicos, se requiere que los exámenes se realicen en lugar equipado con PC, internet, etc.

Por tales motivos, exige un nivel de implementación que supera con creces los exámenes del rendimiento académico realizados hasta ahora. Además, para contar con lo antes mencionado, se requieren ingentes gastos y esfuerzos, incluyendo de recursos humanos.

Como estudio complementario a la evaluación nacional sobre el rendimiento académico y el estado de aprendizaje realizada cada año en Japón, se realiza aparte la evaluación para el análisis de los cambios a lo largo de tiempo. Como parte de sus ítems no son públicos, se trabajan en los preparativos para incorporar la TRI.

### **Formas de utilizar la TRI**

**Acumulación de datos destinada a desarrollar buenos ítems para las pruebas y para mejorar la calidad de la educación.**

Calcular el grado de dificultad y la capacidad de discriminación de los ítems de un prueba aplicada para luego graficar su CCI, permite distinguir a partir de datos estadísticos, entre preguntas buenas y malas.



- ✓ Se pueden obtener sugerencias para desarrollar preguntas para pruebas futuras.
- ✓ Acumular datos estadísticos sobre cada ítem de las pruebas es sumamente importante para mejorar la calidad de la educación.



Pensemos en cómo podemos aplicar la TRI

A pesar de los enormes obstáculos para implementar la TRI como sistema nacional de evaluación, esta puede aplicarse por separado en diversos usos. Como sabrán, la TRI se puede usar para acumular datos destinados a desarrollar buenos ítems para los exámenes y para mejorar la calidad de la educación. Calcular el nivel de dificultad y la discriminación de los ítems de un prueba aplicada para luego graficar su CCI, permite distinguir, a partir de datos estadísticos, entre ítems buenos y malos. De esta forma, se pueden conseguir sugerencias para desarrollar ítems para pruebas futuras. Además, acumular datos estadísticos sobre cada ítem de las pruebas es sumamente importante para mejorar la calidad de la educación. Pensemos en cómo podemos aplicar la TRI.

**Bibliografia**

加藤健太郎・山田剛史・川端一光 (2014) Rによる項目反応理論, オーム社  
文部科学省(2021) 大学入学共通テストにおけるCBT活用に関する検討状況