

PROTOCOLO DE CORRECCIÓN EUNACOM

Este documento técnico tiene por objetivo ofrecer un detalle de los procedimientos involucrados en la implementación de la Teoría de Respuesta al Ítem (IRT por sus siglas en inglés) a través del Modelo de Rasch en la corrección de EUNACOM ST. Los pasos descritos a continuación no corresponden exactamente a los utilizados por el software Winsteps, empleado para corregir y puntuar EUNACOM, dado que este emplea un procedimiento iterativo. Sin embargo, de acuerdo a los creadores del software, la forma de cálculo presentada en este documento entregaría resultados bastante cercanos a las estimaciones realizadas utilizando Winsteps.

Este documento está dividido en cuatro secciones: procedimiento de filtraje o eliminación de ítems del examen, estimación de la dificultad de los ítems empleando modelo Rasch, procedimiento de equating implementado en EUNACOM, procedimiento para la estimación de las habilidades individuales y el procedimiento a través del cual el puntaje obtenido como producto del análisis de Rasch se transforma a la métrica utilizada tradicionalmente por EUNACOM.

1. FILTRAJE DE ÍTEMS

El filtraje es un procedimiento a través del cual son eliminados aquellos ítems cuya dificultad o discriminación se aleja de parámetros aceptables para EUNACOM y se realiza empleando Teoría Clásica de la Medición (TCM). Los pasos involucrados en este procedimiento se detallan a continuación.

1.1 Se corrige el total de los exámenes rendidos por el grupo de examinados asignando 1 a cada respuesta correcta y 0 a cada respuesta incorrecta u omitida.

1.2 Se estima la dificultad de cada ítem. En la TCM la dificultad de un ítem i (p_i) corresponde a la proporción de personas (p), respecto del total de examinados (N) que contestan correctamente a dicho ítem.

1.3 Se calcula la discriminación de cada uno de los ítems que conforman el examen utilizando correlación biserial (R_{bis}).

1.4 Elimine aquellos ítems que poseen valores de $p_i < 0.2$ y/o $R_{bis} < 0.15$.

1.5 Las estimaciones y procedimientos detallados en los puntos siguientes se realizan con el conjunto de ítems que alcanzaron niveles de dificultad y discriminación aceptables.

2. ESTIMACIÓN DE LA DIFICULTAD DE LOS ÍTEMS UTILIZANDO MODELO DE RASCH

Formalmente, el Modelo de Rasch establece que la probabilidad de que una persona (r) responda correctamente ($x=1$) a un ítem (i), depende de su nivel de habilidad (b_r) y la dificultad del ítem (d_i). Esta probabilidad se expresa mediante la función exponencial de la diferencia entre la habilidad de la persona y la dificultad del ítem ($b_r - d_i$).

Dado que esta función puede tomar valores entre 0 e infinito, se utiliza una razón para obtener valores entre 0 y 1.

$$P\{x_{ri} = 1 | b_r, d_i\} = \frac{\exp(b_r - d_i)}{1 + \exp(b_r - d_i)}$$

La dificultad de cada ítem utilizando este modelo se estima a través de los siguientes pasos:

2.1 Se calcula el logit (x_i) de la proporción de incorrectas ($1 - p_i$):

$$x_i = \ln \left[\frac{1 - p_i}{p_i} \right]$$

2.2 Luego se realiza una estimación inicial de la dificultad de cada ítem (d_i^0) a partir de x_i y del promedio de dificultad de los K ítems no filtrados del examen (x).

$$d_i^0 = x_i - x.$$

2.3 Se calcula la proporción p_r como el total de respuestas correctas que obtuvo cada persona respecto de los K ítems.

2.4 Se estima el logit (y_r), que corresponde a la estimación inicial de habilidad de la personas (b_r^0).

$$b_r^0 = y_r = \ln \left[\frac{p_r}{1 - p_r} \right]$$

2.5 Se calcula la varianza de las dificultades de los ítems (U) y la varianza de las habilidades de las personas (V).

$$U = \frac{\sum_i^K x_i^2 - Kx^2}{K - 1}$$

$$V = \frac{\sum_r^N y_r^2 - Ny^2}{N - 1}$$

donde y corresponde al promedio de habilidad de las personas.

2.6 Se estima el factor de expansión para los ítems (Y)¹. Este factor elimina el efecto muestral de las personas que rinden el examen sobre las dificultades estimadas de los ítems.

$$Y = \left[\frac{\left(1 + \frac{V}{2,89}\right)}{\left(1 - \frac{UV}{8,35}\right)} \right]^{\frac{1}{2}}$$

2.7 Se realiza una nueva estimación de dificultad para cada ítem (d_i^1).

$$d_i^1 = Y d_i^0$$

3. PROCEDIMIENTO DE EQUATING

Para asegurar comparabilidad entre distintas versiones del examen es necesario llevar a cabo un procedimiento de equating. El diseño de equating elegido para EUNACOM es el de Ítems Comunes a Conjunto Calibrado. Para ejecutar este procedimiento, EUNACOM construyó banco de ítems previamente calibrados utilizando Modelo de Rasch.

El diseño de equating mencionado implica repetir año a año un conjunto de ítems provenientes del banco (conjunto ítems ancla iniciales o A_0), que permiten llevar cada nuevo examen a la escala de éste. Sin embargo, dado que los ítems repetidos pueden experimentar variaciones (“drift”), se determinó un procedimiento que permite identificar aquel subconjunto de ítems con menor grado de variación (ítems ancla finales o A). Este procedimiento involucra los siguientes pasos:

¹ Teóricamente, tanto los ítems seleccionados para un examen como las personas que lo rinden son parte de una población más grande de ítems y personas. Este factor de expansión permite compensar dicha selección ajustando la estimación por la

distribución muestral de las dificultades de los ítems y de la habilidad personas que formaron parte del examen.

- 3.1 Se calculan los valores de drift iniciales para cada ítem del conjunto A_0 .

$$drift = d_{new} - d_{bank}$$

Donde, d_{new} corresponde a la estimación de dificultad de los ítems ancla obtenida en 2.8 y d_{bank} corresponde a sus dificultades presentes en el banco de ítems.

- 3.2 Si este conjunto A_0 tiene más del 50% de sus ítems con $|drift| > 0.3$, se excluyen a aquellos con mayor $|drift|$ hasta cumplir con este criterio. Así, del conjunto inicial A_0 , se excluyen a ítems, obteniendo un nuevo conjunto $A_1 = A_0 - a$. Si el criterio es satisfecho con el conjunto inicial, $A_1 = A_0$. Este procedimiento se realiza para eliminar sesgos en las estimaciones posteriores².

- 3.3 Para determinar el conjunto final de ítems ancla, se realiza un procedimiento iterativo, calculando la constante de transformación (TC), $bdiff$ y $mOutfit$ en cada iteración j . Tanto $bdiff$ como $mOutfit$ son los indicadores que permiten identificar los ítems con menor drift. Estos valores se calculan como:

$$TC_j = \frac{1}{A_j} \sum_{i=1}^{A_j} d_{bank,i} - d_{new,i}$$

$$bdiff_{j,i} = d_{bank,i} - (d_{new,i} + TC_j)$$

$$mOutfit_j = \frac{1}{A_j} \sum_{i=1}^{A_j} \frac{bdiff_i^2}{SE(d_{new,i})}$$

Donde $SE(d_{new,i})$ corresponde al error de estimación de dificultad de los ítems ancla obtenidas en 2.8³.

- 3.4 Se itera excluyendo del conjunto A_1 aquel ítem con mayor $|bdiff|$ y se repite el paso 3.3 para el nuevo conjunto hasta lograr un $mOutfit \leq 1$. Cada ítem excluido no vuelve a formar parte del conjunto de ítems ancla.

- 3.5 Logrado 3.4, se obtiene el conjunto final de ítems ancla (A), cuya constante de transformación corresponde a la que cumple con el criterio de $mOutfit \leq 1$. Si bien, los ítems excluidos tanto en 3.2 como 3.4 no se consideran para el cálculo de la TC si se deben considerar para el cálculo de puntajes.

- 3.6 Para los ítems del conjunto A se reemplazan las estimaciones $d_{new,i}$ obtenidas en 2.8 por sus valores $d_{bank,i}$. Para el resto de los ítems se calculan sus dificultades finales como

$$d_i = d_i^1 + TC.$$

4. ESTIMACIÓN DE HABILIDADES

Una vez obtenidos los valores finales de dificultad de los ítems del examen se pueden estimar las habilidades finales de los examinados; para esto se siguen los siguientes pasos:

- 4.1 Se calcula nuevamente la varianza de los ítems tal como en el paso 2.6, pero considerando los nuevos valores de dificultad de los ítems estimados en el paso 3.6. A este nuevo valor lo denominaremos U' .

² Este paso se deriva de estudios de simulación de equating realizados por EUNACOM.

³ Este valor es entregado directamente por el software Winsteps, sin embargo, puede obtenerse una aproximación

calculando $SE(d_i) = Y[N/S_i(N - S_i)]^2$, donde S_i corresponde a la cantidad de personas que contestan correctamente el i -ésimo ítem.

4.2 Se calcula el factor de expansión para las personas (X) con el fin de eliminar el efecto muestral de los ítems del examen en las estimaciones de habilidad de las personas.

$$X = \left[\frac{\left(1 + \frac{U'}{2.89}\right)^{\frac{1}{2}}}{\left(1 - \frac{U'V}{8.35}\right)} \right]^{\frac{1}{2}}$$

4.3 Se estiman las habilidades finales de los examinados como

$$b_r = Xb_r^0 = Xy_r.$$

5. ESCALAMIENTO A PUNTAJES FINALES

Las estimaciones de habilidad obtenidas en 4.3 están expresadas en logits (L). Para llevar estas habilidades en logits a puntajes en la escala EUNACOM (E) entre 0 y 100 se realiza una transformación lineal.

Para realizar esta transformación lineal se emplean los siguientes valores: el puntaje mínimo de aprobación (51 puntos, E_1), el puntaje promedio de el examen de diciembre de 2024 (E_2) y sus equivalentes en logits (L_1 y L_2 respectivamente). La fórmula de dicha transformación viene dada por

$$E = (A \times L) + B$$

donde:

$$A = \frac{E_2 - E_1}{L_2 - L_1}$$

$$B = E_1 - (A \times L_1)$$

REFERENCIAS

Kolen, M. J., & Brennan, R. L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices* (2nd ed.). Springer.

Liu, C., Jurich, D., Morrison, C., & Grabovsky, I. (2021). Detection of outliers in anchor items using modified Rasch fit statistics. *Applied Measurement in Education*, 34(4), 327-341.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. MESA Press.