

CALIBRACIÓN ONLINE EN TEST ADAPTATIVOS INFORMATIZADOS.

**Andrés Antivilo Bruna
Julio Olea Díaz
Francisco J. Abad**

***Universidad Autónoma de Madrid
Facultad de Psicología***

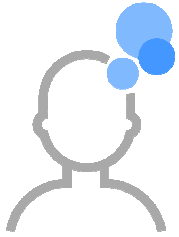
CALIBRACIÓN ONLINE EN TEST ADAPTATIVOS INFORMATIZADOS.

 ANTECEDENTES

 MÉTODO

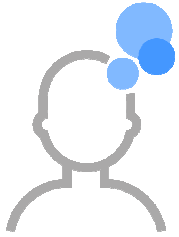
 RESULTADOS

 CONCLUSIÓN



Antecedentes

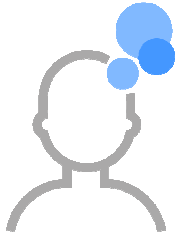
- ❑ Durante la última década se ha observado un crecimiento exponencial del uso de los TAIs (Guo y Wang, 2003; Wainer, 2000).
- ❑ El uso excesivo de ciertos ítems puede comprometer la seguridad de la evaluación (Van der Linden, 2006).
- ❑ *Deterioro (Drift)*: Las propiedades de los ítems pueden cambiar con el tiempo.
- ❑ ¿Cómo llevarla a cabo la actualización del Banco de ítems?
- ❑ *Calibración Online (asumiendo parámetros fijos)*.



Métodos de calibración asumiendo parámetros fijos

- El conocimiento de los parámetros de los ítems operativos es el que permite establecer el de los ítems pretest.

- 1. No prior weight updating – One EM cycle (NWU-OEM).
- 2. No prior weight updating – Multiple EM cycles (NWU-MEM).
- 3. One prior weight updating – One EM cycle (OWU-OEM).
- 4. One prior weight updating – Multiple EM cycles (OWU-MEM).



Estudios Previos

- ❑ **Ban et al. (2001):** Métodos NWU + Stocking A y B + BILOG.

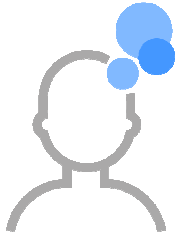
Resultado Principal: NWU-MEM y Método B de Stocking.

- ❑ **Ban et al. (2002):** Métodos NWU + Stocking B.

Resultado Principal: NWU-MEM y Método B de Stocking.

- ❑ **Kim (2006):** Cinco métodos APF + BILOG MG + PARSCALE.

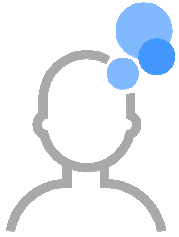
Resultado Principal: Métodos MEM con $N(0,1)$.



Método

- ❑ **Objetivo:** Determinar qué diseño de calibración online basado en la Estimación máximo verosímil marginal vía algoritmo *Expectation-Maximization (EM)* entrega las estimaciones más precisas de ítems pretest.

- ❑ **Generación de Datos:**
 1. Los *ítems operativos* del TAI fueron obtenidos del banco de eCAT.
 2. Los *ítems operativos* del Test Convencional fueron simulados.
 3. Para todos los *ítems pretest* se estableció que $a = 1.20$ y $c = 0.20$. Para el *Parámetro b* se asignaron valores equidistantes en 0.5 puntos, siendo la media de la distribución cero.



Método

❑ Factores manipulados:

1. Tipo de Test (TAI o Test Convencional; 30 ítems).
2. Tamaño Muestral (500, 1000 ó 2000).
3. Ratio ítems pretest/ ítems operativos (número de ítems pretest: 3, 5 ó 7).

10 réplicas para cada condición.

❑ Criterio: RMSE.

- ❑ Se elaboró un *algoritmo adaptativo* que permitió simular la aplicación del TAI.

Resultados TAI

METODO	RMSE-A		RMSE-B		RMSE-C	
	N de ítems pretest (<i>Ratio</i>)					
	3 (.10)	7 (.23)	3 (.10)	7 (.23)	3 (.10)	7 (.23)
<i>N = 500</i>						
<i>NWU-OEM</i>	.22	.28	.10	.14	.02	.03
NWU-MEM	.05	.15	.10	.09	.02	.03
<i>OWU-OEM</i>	.21	.35	.11	.14	.03	.02
OWU-MEM	.06	.16	.10	.09	.03	.02
<i>N = 2000</i>						
<i>NWU-OEM</i>	.22	.26	.03	.13	.04	.03
NWU-MEM	.04	.07	.02	.07	.02	.02
<i>OWU-OEM</i>	.22	.26	.01	.13	.03	.02
OWU-MEM	.03	.09	.03	.07	.01	.02

Resultados TAI

METODO	RMSE-A		RMSE-B		RMSE-C	
	N de ítems pretest (<i>Ratio</i>)					
	3 (.10)	7 (.23)	3 (.10)	7 (.23)	3 (.10)	7 (.23)
<i>N = 500</i>						
<i>NWU-OEM</i>	.22	.28	.10	.14	.02	.03
NWU-MEM	.05	.15	.10	.09	.02	.03
<i>OWU-OEM</i>	.21	.35	.11	.14	.03	.02
OWU-MEM	.06	.16	.10	.09	.03	.02
<i>N = 2000</i>						
<i>NWU-OEM</i>	.22	.26	.03	.13	.04	.03
NWU-MEM	.04	.07	.02	.07	.02	.02
<i>OWU-OEM</i>	.22	.26	.01	.13	.03	.02
OWU-MEM	.03	.09	.03	.07	.01	.02

Resultados TAI

	RMSE-A		RMSE-B		RMSE-C	
	N de ítems pretest (<i>Ratio</i>)					
	3 (.10)	7 (.23)	3 (.10)	7 (.23)	3 (.10)	7 (.23)
METODO						
<i>N = 500</i>						
<i>NWU-OEM</i>	.22	.28	.10	.14	.02	.03
NWU-MEM	.05	.15	.10	.09	.02	.03
<i>OWU-OEM</i>	.21	.35	.11	.14	.03	.02
OWU-MEM	.06	.16	.10	.09	.03	.02
<i>N = 2000</i>						
<i>NWU-OEM</i>	.22	.26	.03	.13	.04	.03
NWU-MEM	.04	.07	.02	.07	.02	.02
<i>OWU-OEM</i>	.22	.26	.01	.13	.03	.02
OWU-MEM	.03	.09	.03	.07	.01	.02

Resultados Test Convencional

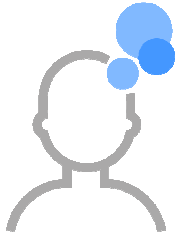
	RMSE-A		RMSE-B		RMSE-C	
	N de ítems pretest (<i>Ratio</i>)					
	3 (.10)	7 (.23)	3 (.10)	7 (.23)	3 (.10)	7 (.23)
METODO						
<i>N = 500</i>						
<i>NWU-OEM</i>	.26	.33	.09	.16	.04	.04
NWU-MEM	.12	.17	.11	.13	.02	.03
<i>OWU-OEM</i>	.27	.34	.11	.16	.04	.04
OWU-MEM	.12	.18	.12	.14	.02	.03
<i>N = 2000</i>						
<i>NWU-OEM</i>	.23	.25	.13	.18	.04	.03
NWU-MEM	.07	.13	.08	.16	.03	.02
<i>OWU-OEM</i>	.24	.25	.15	.18	.04	.03
OWU-MEM	.08	.13	.10	.16	.03	.02

Resultados Test Convencional

	RMSE-A		RMSE-B		RMSE-C	
	N de ítems pretest (<i>Ratio</i>)					
	3 (.10)	7 (.23)	3 (.10)	7 (.23)	3 (.10)	7 (.23)
METODO						
<i>N = 500</i>						
<i>NWU-OEM</i>	.26	.33	.09	.16	.04	.04
NWU-MEM	.12	.17	.11	.13	.02	.03
<i>OWU-OEM</i>	.27	.34	.11	.16	.04	.04
OWU-MEM	.12	.18	.12	.14	.02	.03
<i>N = 2000</i>						
<i>NWU-OEM</i>	.23	.25	.13	.18	.04	.03
NWU-MEM	.07	.13	.08	.16	.03	.02
<i>OWU-OEM</i>	.24	.25	.15	.18	.04	.03
OWU-MEM	.08	.13	.10	.16	.03	.02

Resultados Test Convencional

	RMSE-A		RMSE-B		RMSE-C	
	N de ítems pretest (<i>Ratio</i>)					
	3 (.10)	7 (.23)	3 (.10)	7 (.23)	3 (.10)	7 (.23)
METODO						
<i>N = 500</i>						
NWU-OEM	.26	.33	.09	.16	.04	.04
NWU-MEM	.12	.17	.11	.13	.02	.03
OWU-OEM	.27	.34	.11	.16	.04	.04
OWU-MEM	.12	.18	.12	.14	.02	.03
<i>N = 2000</i>						
NWU-OEM	.23	.25	.13	.18	.04	.03
NWU-MEM	.07	.13	.08	.16	.03	.02
OWU-OEM	.24	.25	.15	.18	.04	.03
OWU-MEM	.08	.13	.10	.16	.03	.02



Conclusión

- ❑ Los métodos de calibración online resultan más precisos en el TAI que en el Test convencional.
- ❑ Los resultados informados deben ser interpretados con cautela, ya que el número de réplicas utilizado es limitado.
- ❑ Futuros estudios sobre diseños de calibración online debieran evaluar la importancia de incluir información referente a las posibles distribuciones previas, tanto de los ítems (operativos y pretest) como de los examinados.
- ❑ Explorar el posible efecto de incluir ítems pretest con parámetros extremos o de incluir ítems desajustados al modelo 3PL.

CALIBRACIÓN ONLINE EN TEST ADAPTATIVOS INFORMATIZADOS

**Andrés Antivilo Bruna
Julio Olea Díaz
Francisco J. Abad**

MUCHAS GRACIAS