

Equiparación de puntuaciones. Procedimientos para la equivalencia en la calificación de pruebas estándar en diferentes grupos.

Diana Rodríguez Valbuena

Estudio cofinanciado por la Universidad Nacional como proyecto ganador 19432 de 2014 dentro de la Convocatoria nacional para el fortalecimiento de la investigación, la creación y la innovación en posgrados de la Universidad Nacional de Colombia 2013-2015; y por el ICFES mediante contrato de financiamiento – Recuperación contingente no 59 de 2014, derivado de la convocatoria a grupos de investigación para presentar propuestas de trabajos con datos derivados de pruebas de evaluación de la calidad de la educación.

Antecedentes

- Los procedimientos de equiparación de puntuaciones se utilizan para hacer intercambiables los resultados de dos pruebas que miden el mismo constructo, para que estén puedan ser comparadas de manera adecuada.
- “La equiparación de puntuaciones es necesaria en cualquier programa de evaluación que produzca diferentes formas de prueba continuamente” (Holland, 2007)
- Se consideran diferentes las pruebas presentadas por personas con y sin limitación visual debido al canal de entrada de la información

- En este caso es importante considerar si el presentar la misma prueba por medios diferentes es equivalente para las personas con limitación visual.
- Un estudio del grupo de investigación (Lancheros, 2013) simuló condiciones similares a las de las pruebas SABER, en el que se evidenció las bondades de los procedimientos de equiparación Tucker y Media Sigma, evaluadas mediante la magnitud y la varianza del error de equiparación de cada uno de ellos, a través de las réplicas.
- Suh, Mroch, Kane y Ripkey (2009) sugieren que a nivel experimental es necesario dar cuenta del funcionamiento de los procedimientos de equiparación en diferentes tamaños de muestra y con distintos tipos de test de anclaje.

Objetivos

- Evaluar el funcionamiento de dos procedimientos de equiparación (Tucker y Media-sigma) con los datos obtenidos en la aplicación de la subprueba de lenguaje de la prueba SABER 11, aplicada el segundo semestre de 2013.
- Evaluar el funcionamiento de dos procedimientos de equiparación, cuando se eliminan de la calificación total los ítems que presentan funcionamiento diferencial.
- Determinar si es adecuado el uso de alguno de los métodos de equiparación para la calificación de las pruebas SABER, teniendo en cuenta el nivel de error que se identifique en cada una de las aplicaciones de los procedimientos.

Metodología

- Dos formas de prueba, compuestas por 24 ítems cada una.
- Depuración de base de datos.
- 5 ítems de anclaje por cada forma de prueba, elegidos con base en el parámetro de dificultad
- 25 muestreos de 1150 personas sin limitación visual

Prueba	Forma 1	Forma 2
Personas con limitación visual	46	58
Personas sin limitación visual	230400	288053
Total	230446	288111

- Se realiza el procedimiento bajo un diseño de grupos no equivalentes con test de anclaje.
- Se eligen 5 ítems de anclaje con base en la dificultad en cada grupo.
- En el método lineal de Tucker se realiza una regresión de los puntajes de las dos pruebas, con respecto al funcionamiento del test de anclaje en los dos grupos
- En el método Media-sigma se realiza la transformación al sustituir las medias y desviaciones estandar del parametro estimado de dificultad de los items por la estimación del anclaje.

Resultados

Ítems de anclaje

	Dificultad SLV	Dificultad LV
Ítem 10	0,10	0,24
Ítem 16	-1,13	-1,01
Ítem 18	0,01	0,14
Ítem 32	-0,26	-0,15
Ítem 36	0,55	0,66

Ítems con Funcionamiento diferencial

	Prueba t	Valor p	Grupo desfavorecido
Ítem 2	2,46	0,01	SLV
Ítem 23	-2,22	0,02	LV
Ítem 31	2,19	0,03	SLV

Tablas de puntajes transformados. Método Tucker (Forma 1 completa y modificada)

Puntaje prueba SLV	Estimado Tucker	Error Tucker	Estimado Tucker sin DIF	Error Tucker sin DIF
0	4,57	0,07	4,17	0,07
1	5,17	0,05	4,81	0,05
2	5,76	0,04	5,49	0,03
3	6,36	0,03	6,16	0,02
4	6,96	0,03	6,84	0,02
5	7,55	0,04	7,51	0,01
6	8,15	0,05	8,19	0,01
7	8,74	0,07	8,86	0,01
8	9,34	0,09	9,54	0,01
9	9,94	0,13	10,21	0,01
10	10,53	0,16	10,89	0,02
11	11,13	0,21	11,56	0,03

Tablas de puntajes transformados. Método Tucker (Forma 1 completa y modificada)

Puntaje prueba SLV	Estimado Tucker	Error Tucker	Estimado Tucker sin DIF	Error Tucker sin DIF
12	11,72	0,26	12,24	0,04
13	12,32	0,32	12,91	0,05
14	12,92	0,38	13,59	0,07
15	13,51	0,45	14,27	0,09
16	14,11	0,53	14,94	0,11
17	14,70	0,62	15,62	0,13
18	15,30	0,71	16,29	0,16
19	15,90	0,81	16,97	0,19
20	16,49	0,91	17,64	0,22
21	17,09	1,03	18,32	0,25
22	17,68	1,15	18,99	0,29
23	18,28	1,28	19,67	0,32
24	18,88	1,42	20,34	0,36

Gráfica error por puntaje . Método lineal.

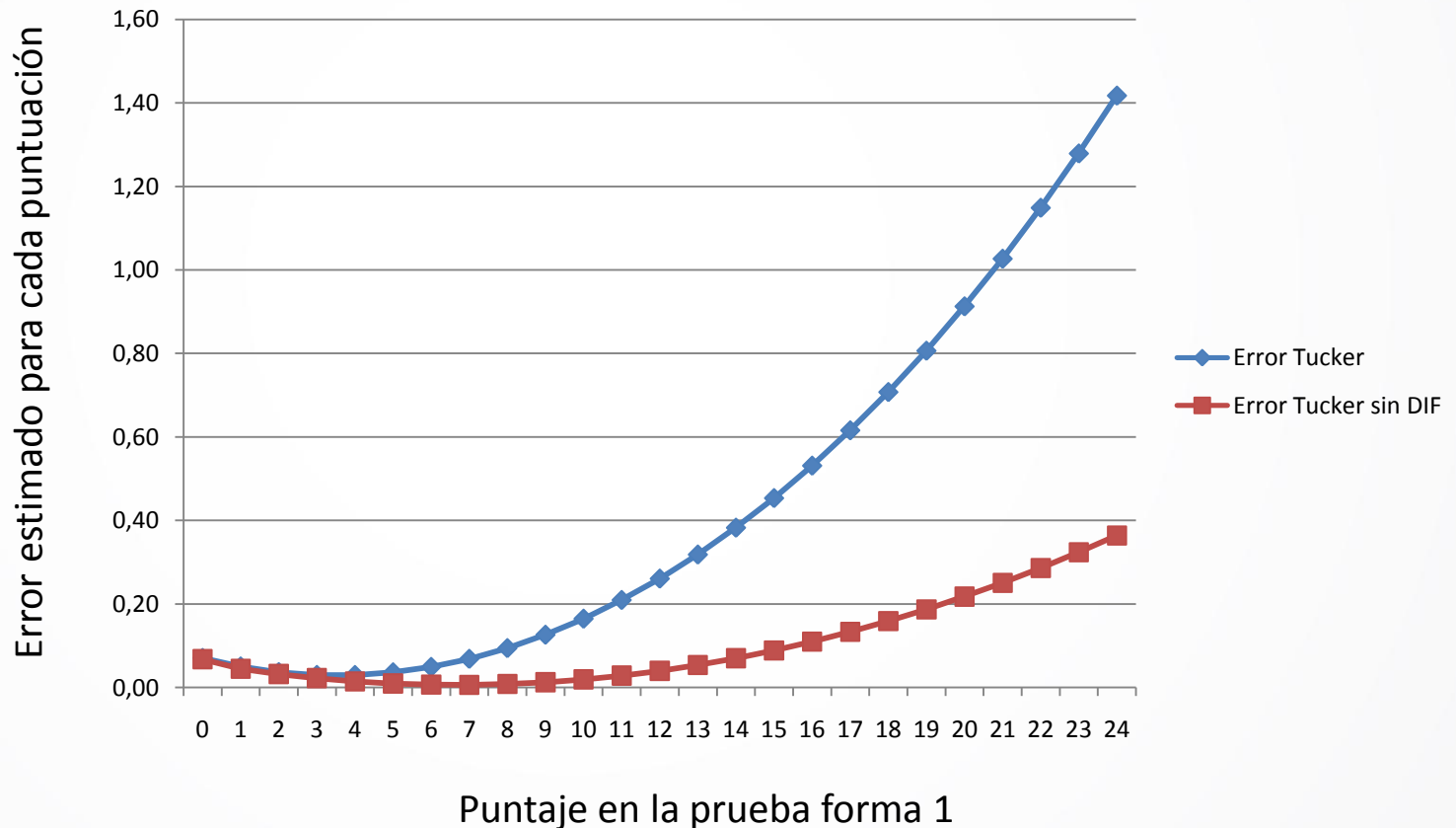


Tabla puntaje transformado

Método Media Sigma- Forma 1 Completa

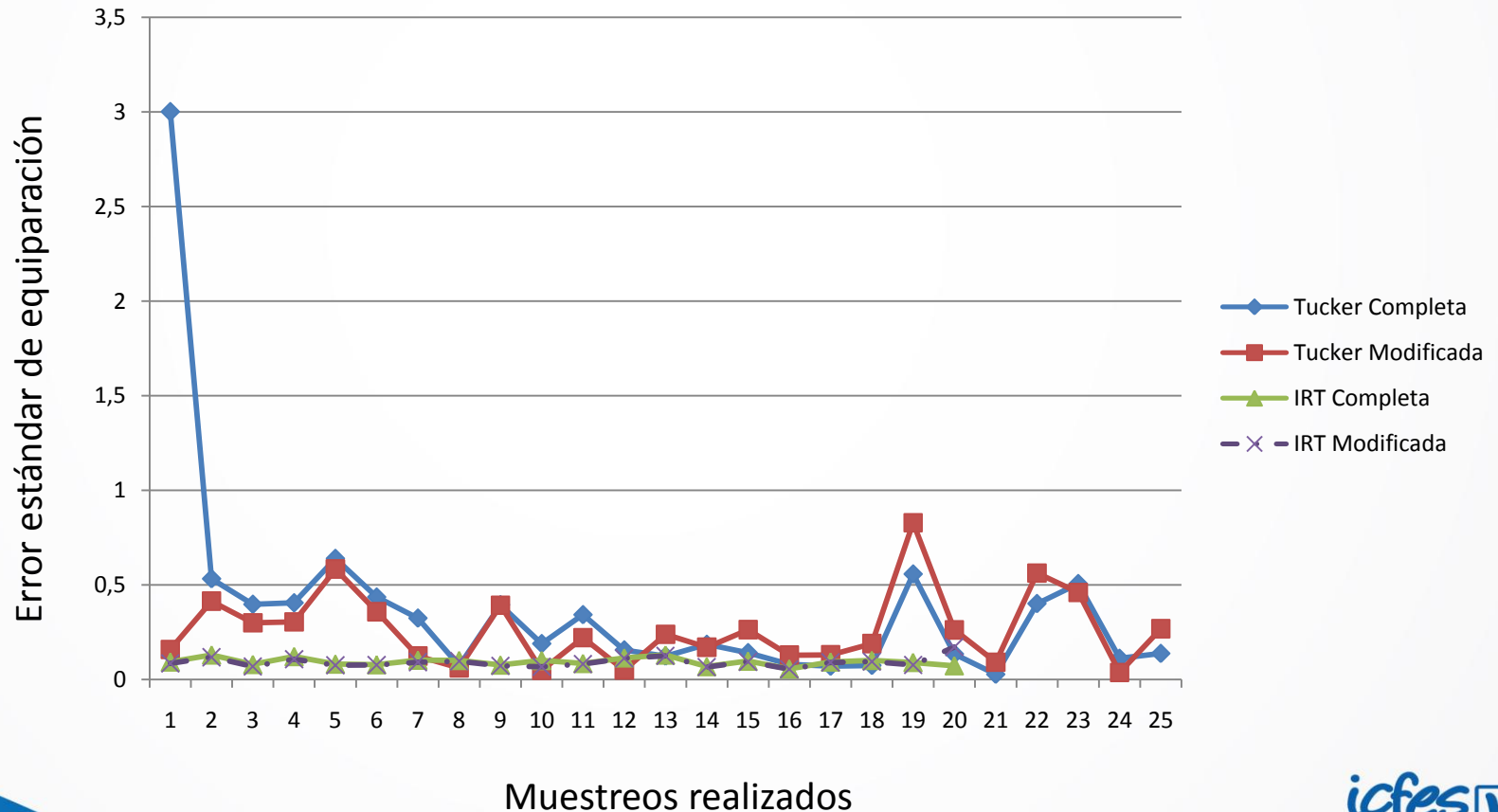
	Completa				Modificada			
SLV	SLV	LV	Estimada	Error	SLV	LV	Estimada	Error
ITEM_1	0,708	0,140	0,844	0,008	0,650	0,130	0,836	0,009
ITEM_2	0,478	1,500	0,612	0,012	_____	1,490	_____	_____
ITEM_5	-1,260	-1,690	-1,134	0,009	-1,315	-1,700	-1,136	0,0092549 1
ITEM_6	1,437	1,180	1,577	0,023	1,379	1,170	1,567	0,023
ITEM_8	1,896	2,480	2,035	0,019	1,836	2,480	2,023	0,019
ITEM_9	-0,606	-1,110	-0,476	0,006	-0,662	-1,120	-0,480	0,005
ITEM_10	0,098	0,240	0,229	0,003	0,042	0,230	0,224	0,003
ITEM_12	0,492	0,040	0,626	0,008	0,433	0,040	0,617	0,008
ITEM_13	0,443	0,240	0,577	0,007	0,385	0,230	0,569	0,007
ITEM_15	0,807	0,550	0,944	0,010	0,749	0,550	0,936	0,010
ITEM_16	-1,134	-1,010	-1,005	0,001	-1,190	-1,010	-1,008	0,001
ITEM_18	0,020	0,140	0,152	0,002	-0,036	0,130	0,148	0,003

Tabla puntaje transformado

Método Media Sigma- Forma 1 Completa

	Completa				Modificada			
SLV	SLV	LV	Estimada	Error	SLV	LV	Estimada	Error
ITEM_19	0,014	-0,150	0,145	0,006	-0,043	-0,150	0,140	0,006
ITEM_21	0,029	-0,240	0,159	0,004	-0,028	-0,250	0,153	0,004
ITEM_23	0,488	-0,150	0,623	0,008	0,430	_____	0,615	0,008
ITEM_24	-0,020	-0,240	0,113	0,005	-0,075	-	0,109	0,005
ITEM_25	-0,351	0,340	-0,220	0,007	-0,409	0,330	-0,227	0,015
ITEM_26	-0,661	-1,220	-0,531	0,003	-0,716	-1,230	-0,534	0,005
ITEM_28	-0,628	-0,050	-0,500	0,006	-0,684	-0,060	-0,503	0,006
ITEM_29	-0,177	-0,050	-0,044	0,005	-0,234	-0,060	-0,051	0,017
ITEM_31	-1,733	-1,110	-1,609	0,010	_____	-1,120	_____	_____
ITEM_32	-0,281	-0,150	-0,150	0,005	-0,336	-0,150	-0,153	0,004
ITEM_34	-0,584	-0,340	-0,456	0,007	-0,639	-0,340	-0,459	0,007
ITEM_36	0,526	0,660	0,658	0,002	0,468	0,660	0,650	0,002

Error a lo largo de los muestreos 4 procedimientos en la forma 1



Estadísticos descriptivos de los errores de equiparación

	Método Tucker				Método Media-Sigma			
	Min	Max	Media	Desviación	Min	Max	Media	Desviación
Forma 1	0,028	3,01	0,37	0,57	0,056	0,130	0,094	0,019
F1 mod	0,038	0,82	0,26	0,19	0,055	0,172	0,092	0,026

Conclusiones

- Resultados consistentes con los hallazgos en estudios anteriores, el método Media-Sigma presenta menor cantidad de error.
- Se evidencia que en los puntajes más altos se encuentra mayor error de equiparación
- En la estimación de puntajes equiparados se observa diferencias de hasta cuatro puntos para las personas con limitación visual, lo que evidencia la necesidad de realizar procedimientos que garanticen una comparación adecuada.

Conclusiones

- Debido a que no se cumple el supuesto de unidimensionalidad, se prefiere el método lineal con formas de prueba modificada
- Desafíos metodológicos: Equiparación en pruebas que no cumplen el supuesto de unidimensionalidad.
- Comparar el funcionamiento del método de TRI cuando las pruebas son unidimensionales y cuando no lo son.
- Desarrollo importante para la medición y evaluación en distintas pruebas.

Referencias

- Holland, P. (2007). A Framework and History for Score Linking. En Dorans, Pommerich y Holland (Eds). *Linking and Aligning Scores and Scales* (5-29). Springer
- Lancheros Florian, L. C. (2013). *Métodos de Equiparación de Puntuaciones: Los exámenes de estado en población con y sin limitación visual*. (Tesis inédita de maestría en psicología). Universidad Nacional de Colombia, Bogotá.
- Suh, Y., Mroch, A., Kane M., Ripkey, D. (2009). An Empirical Comparison of Five Linear Equating Methods for the NEAT Design. *Measurement*, 7, 147-173.
doi:10.1080/15366360903418048