

un título ingenioso y bonito.

Constanza Segovia Gonzalez

Francisca Vilca Sánchez

Abstract:

Las 120 palabras qls de abstract adipiscing elit. Donec efficitur posuere maximus. Nam rutrum eget leo in ultrices. Vestibulum tristique est vel tortor lacinia fermentum. Curabitur bibendum lacus eu neque fringilla, non volutpat orci sagittis. Pellentesque sit amet libero tempus, ultrices nunc et, consectetur dolor. Nulla sit amet nisi malesuada, pellentesque quam vel, lobortis eros. Ut facilisis aliquet lectus, eu semper purus rutrum non. Aenean at pulvinar mauris, sit amet molestie orci. Etiam sed quam risus. Cras luctus blandit pulvinar. Sed ac purus vitae sapien euismod lacinia. Interdum et malesuada fames ac ante ipsum primis in faucibus. Nullam ac nibh vel tellus maximus blandit. Nam vel nisl eros. Cras eget eros non orci gravida dictum vitae ut lectus. Integer rutrum.

Introducción

“La educación es el motor para construir cualquier sociedad” (Umar, 2019), si tenemos una buena formación durante nuestra vida, las oportunidades de tener éxito aumentan considerablemente. Sin embargo, resulta complejo cuantificar la calidad de la educación que se entrega o se recibe. Para las naciones es un desafío constante poder catalogar lo que es una buena o mala educación. Actualmente, una buena forma para medir la calidad de la educación es a través de pruebas cognitivas, pero investigaciones recientes demuestran que las habilidades no-cognitivas tienen un impacto mayor o igual que las cognitivas (Duckworth & Yeager, 1992).

En Chile, uno de los organismos encargados de mejorar el sistema educacional es la Agencia de Calidad de la Educación (ACE), cuyo principal objetivo es “colaborar con las comunidades educativas para el aseguramiento del aprendizaje oportuno y el desarrollo integral de los niños, niñas, adolescentes y adultos” (Agencia de Calidad de la Educación, 2023). Además de revisar los logros de los aprendizajes estándares, uno de sus enfoques se basa en evaluar otros Indicadores de Calidad o Indicadores de Desarrollo Personal y Social (IDPS), para ampliar la concepción que se tiene sobre la calidad de educación, evaluando Autoestima Académica y Motivación Escolar, Participación y Formación Ciudadana, Clima de Convivencia Escolar entre otros. Para lograr medir estos indicadores la ACE, realiza cuestionarios de autoreporte,

donde estudiantes, docentes, padres y apoderados, dan su visión sobre como funciona el establecimiento educacional en los distintos aspectos, generando así un puntaje que permita compararlos.

En base a lo anterior, surge el problema de que los puntajes solo permiten comparar a los establecimientos educacionales en el mismo año y no a través del tiempo, lo que le impide a la ACE comprobar si es que hay una mejora con respecto al año anterior. Es por ello que el objetivo de esta investigación será intentar encontrar un mecanismo para hacer comparables los puntajes obtenidos en pruebas no cognitivas en específico, en el área de Clima Convivencia Escolar y hacer que sean equivalentes los puntajes de otro año. Para ello se utilizarán datos IDPS del año 2017 para realizar un estudio de simulación que compruebe la extensión del método Haebara para obtener las constantes de equating de modelos multidimensionales para respuestas politómicas con estructura bi-factor, para así comparar puntajes en la misma escala. Por esta razón, esta investigación toma relevancia, pues da validez a la afirmación sobre si hubo una mejora con respecto al año anterior del constructo de Clima de Convivencia Escolar.

Esta investigación se organizará en seis secciones. La segunda, consiste en dar un contexto sobre la naturaleza de los datos IDPS que se usarán de modo práctico. En la tercera sección, se explicará la metodología usada para el mecanismo de obtención de los puntajes bifactor para ítems politómicos y la extensión del método de Haebara que se usará para encontrar las constantes de equating que serán usadas para hacer la equivalencia con el año que sigue. A continuación, en la cuarta sección, se presentará el mecanismo del estudio de simulación para comprobar que la extensión planteada con el método Haebara es correcta. En la quinta parte, se mostrarán los resultados de la simulación comprobando el funcionamiento del método. Finalmente, en la sexta y última sección, se entregán las conclusiones sobre los resultados obtenidos y algunos comentarios sobre que se podría hacer con esta investigación.

Datos IDPS

Los datos que recolecta la encuesta “Otros Indicadores de Calidad o Indicadores de Desarrollo Personal y Social (IDPS), son un conjunto de índices que entregan información relacionada con el desarrollo personal y social de los estudiantes de un establecimiento” (Agencia de Calidad de la Educación, 2023). Esto lo realiza midiendo diferentes indicadores como “*Autoestima Académica y Motivación Escolar*”, “*Clima de Convivencia Escolar*”, “*Participación y Formación Ciudad*”, “*Asistencia Escolar*”, entre otros.

En esta investigación el foco estará en el Clima de Convivencia Escolar. Este es un constructo que es evaluado mediante cuestionarios de autoreporte donde estudiantes, docentes, padres y apoderados evalúan al establecimiento con respecto a tres dimensiones, el Ambiente de Respeto, el Ambiente Organizado y el Ambiente Seguro. Como el objetivo es lograr equiparar puntajes obtenidos en pruebas no-cognitivas, para ejemplificar se usará la respuesta de 9679 docentes para 16 ítems politómicos en el año 2017, específicamente en la dimensión de Ambiente Seguro.

El cuestionario de autoreporte usado para medir la dimensión de Ambiente Seguro en su mayoría es evaluado con escalas, como una escala Likert (Likert, 1932) o una escala de frecuencia. La escala de Likert permite capturar realmente la percepción de los individuos con respecto a sus vivencias, por ejemplo, “Totalmente en desacuerdo”, “En desacuerdo”, “Indiferente”, “De acuerdo”, “Totalmente de acuerdo”, etc. A su vez, las escalas de frecuencia se utilizan para medir la frecuencia de algún evento de interés, por ejemplo, “Muy frecuentemente”, “Frecuentemente”, “Ocasionalmente”, “Raramente”, “Nunca”, etc. Como los 16 ítems de este cuestionario son de alguna de esta categoría, se puede afirmar que se está trabajando con respuesta politómicas, lo que obliga a buscar un modelo que sea capaz de ajustarse a estos datos.

Metodología

Modelo MGRM-bifactor:

El primer problema que se enfrenta al querer realizar una comparación entre los puntajes del Clima de Convivencia Escolar para distintos años (equating) es la naturaleza de los datos que se posee, ya que provienen de una escala de Likert o de una escala de frecuencia, por lo que no existe una respuesta correcta, sino que diversas categorías. Una posible solución para encontrar un modelo que se ajuste a los datos es a través de la Teoría de Respuesta al Ítem (IRT, por sus siglas en inglés), por lo que el problema se reduce a encontrar el mejor modelo que ajuste datos politómicos. Según la investigación de González & Robledo (2023), el mejor candidato para ajustar estos datos es un Modelo de Respuesta Graduada Multidimensional con estructura bifactor (MGRM-bifactor), ya que este presenta las mejores medidas con respecto a la bondad de ajuste, pues no hay solo una dimensión para la variable latente θ , sino que para este caso en particular hay tres para el constructo a evaluar (parte bifactor del modelo) y múltiples alternativas de respuestas aceptables (parte MGRM).

El modelo MGRM-bifactor es un buen candidato para ajustar los datos IDPS, pues hace que la interpretación sea más simple. La respuesta para este modelo será una combinación entre el *factor general* y los *factores específicos*. Ahora bien, la probabilidad de responder una categoría específica k para un modelo MGRM-bifactor viene dada por:

$$P_r(X_j = k | \theta^G, \theta^S, a_j^G, \mathbf{a}_j^S, d_{j,k}) = P_{j,k} - P_{j,k+1} \quad (1)$$

donde,

$$P_{j,k} = P_r(X_j \geq k | \theta^G, \theta^S, a_j^G, \mathbf{a}_j^S, d_{j,k}) = \frac{\exp\{a_j^G \theta^G + \sum_{i=1}^S \mathbb{1}_{\{j \in i\}} \mathbf{a}_j^S{}' \theta^S + d_{j,k}\}}{1 + \exp\{a_j^G \theta^G + \sum_{i=1}^S \mathbb{1}_{\{j \in i\}} \mathbf{a}_j^S{}' \theta^S + d_{j,k}\}} \quad (2)$$

aquí θ^G es el factor general, es decir, la variable latente que explica todos los ítemes, $\theta^S = (\theta_1^S, \theta_2^S, \dots, \theta_S^S)$ son los factores específicos, asociado cada uno a uno a un factor. También, esta a_j^G y \mathbf{a}_j^S que son los parámetros de discriminación asociados a θ^G y θ^S , respectivamente y finalmente se tiene $d_{j,k}$ que es el parámetro de locación para el ítem j y la categoría k . Por otro lado, se tiene la función indicadora $\mathbb{1}_{\{j \in i\}}$ que tomará el valor 1 si el ítem j está en el factor i , ($i = 1, \dots, S$) y 0 si no lo está.

Para el ejemplo práctico de esta investigación se fija la cantidad de factores específicos en $S = 3$, pues como se mencionó anteriormente es el que mejor se ajusta a los datos según González & Robledo (2023) y la cantidad de ítemes en $j = 16$, ya que así se construye el cuestionario para la dimensión de Ambiente Seguro. Entonces, en esta parte del procedimiento lo que se hace es mediante la función `bfactor()` del paquete `mirt()` estimar los modelos para los datos IDPS de la dimensión Ambiente Seguro para el año 2017 y 2018. Aquí se obtienen los coeficientes θ^G , θ^S , a_j^G , \mathbf{a}_j^S y $d_{j,k}$ para los 16 ítemes comunes de cada año, esto con el fin de posteriormente de hacer la equivalencia entre estos puntajes.

Método de Haebara para un modelo MGRM-bifactor

Ahora que se logra obtener una forma de calcular los puntajes para el Ambiente Seguro, se tiene el segundo problema y el objetivo principal de esta investigación que es hacer comparaciones entre distintos grupos, en particular para estos datos, para saber si mejoró o no el Ambiente Seguro de los establecimientos educacionales, con respecto de un año a otro. En esta investigación se usará una extensión del método Haebara (Haebara, 1980) para ítemes politómicos, con el fin de hacer equivalentes las escalas y así comparar los puntajes de dos años consecutivos.

La idea fundamental del método Haebara se basa en usar los puntajes de un año y modificarlos mediante transformaciones lineales a la nueva escala, para hacer esto es necesario encontrar las constantes que permitan realizar esta transformación. Esta idea fue planteada por Haebara (1980), sin embargo, esta es solo formulada para ítemes dicotómicos. Por lo que, con el paso del tiempo se han presentado modificaciones a la idea original, como es el caso de Kim (2019) donde usando un modelo bifactor-3PL realiza una extensión del método Haebara. En esta investigación, al ser una extensión de lo investigado por González & Robledo (2023) la propuesta es usar el método Haebara para un modelo MGRM-bifactor de 3 dimensiones.

Para el modelo MGRM-bifactor propuesto se establece que la estructura del factor específico será de $S = 3$ dimensiones, cada una denotada como s_1, s_2 y s_3 . Se define $\mathbf{U} = (u_G, u_1, u_2, u_3)$ y $\mathbf{b} = (b_G, b_1, b_2, b_3)$ dos vectores, cada uno de largo $S + 1 = 4$, como las constantes de equating a estimar. Además, se define Y como la escala antigua y X como la nueva escala. Entonces se tiene que la transformación lineal desde la escala X a Y para el modelo MGRM-bifactor propuesto funciona para cada parámetro como:

$$\theta_G^{Yx} = \frac{\theta_X^G - b_G}{u_G}$$

$$\theta_s^{Yx} = \frac{\theta_X^s - b_s}{u_s} \quad s = 1, 2, 3$$

$$a_{Yx}^G = u_G a_X^G$$

$$a_{Yx}^s = u_s a_X^s \quad s = 1, 2, 3$$

$$d_{k,s}^{Yx} = d_k^X + a_X^G b_G + a_X^s b_s$$

El desafío principal al realizar estas transformaciones es estimar las constantes de equating \mathbf{U} y \mathbf{b} . Por lo que la solución propuesta para este problema se reduce a optimizar:

$$\arg \min_{\mathbf{U}, \mathbf{b}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{j \in V} \sum_{k=1}^{m_j} (P_r(X_j = k | \theta^G, \theta^S, a_Y^G, \mathbf{a}_Y^S, d_{j,k}^{Yx}) - P_r(X_j = k | \theta^G, \theta^S, a_{Yx}^G, \mathbf{a}_{Yx}^S, d_{j,k}^{Yx}))^2 d\theta^G \theta^S \quad (3)$$

Donde V es el conjunto con los ítemes comunes y m_j indica la cantidad de categorías por ítem. En esta parte del procedimiento, lo que se hace es determinar las constantes de linking, para ello se usarán funciones que fueron creadas en \mathbf{R} y los parámetros a_j^G , \mathbf{a}_j^S y $d_{j,k}$ de 2017 y 2018. Los resultados obtenidos pueden ser observados en la Tabla 1:

u_G	u_1	u_2	u_3	b_G	b_1	b_2	b_3
0.935	0.977	0.959	1.064	0.166	-0.096	-0.109	-0.099

Tabla 1: Constantes de Linking estimadas con el método Haebara

Este procedimiento se realiza mediante funciones en \mathbf{R} que tienen un gran costo computacional, debido al proceso de optimización de las constantes \mathbf{U} y \mathbf{b} , una vez que se determina el valor de las constantes de linking, es posible hacer la transformación de las escalas logrando así la comparación con puntajes equilibrados. Sin embargo, al ser esta extensión del método Haebara, es necesario someterla a revisión, es por eso que en esta investigación se realizará un estudio de simulación del método.

Simulación

Como se explicó en la sección anterior, la metodología ocupada es una extensión del Método de Haebara para encontrar las constantes de linking. Para ello en esta investigación se somete a un estudio de simulación cuyo objetivo será probar que para diferentes conjuntos de datos con características similares a los datos del 2018, siempre se llegará a las mismas constantes de equating.

Para comenzar el proceso de simulación, se debe tener en cuenta que dada la complejidad del algoritmo para obtener las constantes de equating, es difícil tener datos simulados con todos los parámetros obtenidos de forma aleatoriamente desde una función conocida. Por lo mismo, que para este proceso de simulación lo que se hará es fijar los datos del 2017 y la aleatoriedad del proceso se hará con datos simulados del 2018, para eso se simulará desde una Normal multivariada de media 0 y varianza I el parámetro de habilidad θ y se dejarán fijos el parámetro de discriminación \mathbf{a} y de locación \mathbf{d} , de esta manera se obtendrán datos simulados de forma aleatoria.

Ahora con los conjuntos de datos simulados junto con el algoritmo para estimar las constantes de equating se obtendrán varios sets de constantes de equating. Si el valor esperado de estas constantes coincide con los resultados de la Tabla 1, se puede afirmar que la extensión del método Haebara propuesta en esta investigación funciona.

Esta descripción puede sonar bastante similar, pues es lo que se conoce en Procesos Estocásticos como Simulación por Monte Carlo, ya que básicamente lo que se hace es simular datos con algún parámetro libre, ajustar un modelo a los datos y repetir el procedimiento múltiples ocasiones para comprobar que el promedio de los coeficientes de los modelos estimados se acerquen a lo que inicialmente fueron los valores para iniciar el procedimiento. Es común en problemas de gran complejidad matemática, recurrir a este tipo de demostraciones para comprobar el funcionamiento de los algoritmos (Bonate, 2001).

Para hacer este procedimiento se dividirá en cuatro pasos la simulación:

- **Paso 1:** Crear n set de datos, usando la función `simdata()` del paquete `mirt`, a la cual se le dan como valores iniciales: el parámetro de discriminación \mathbf{a} y de locación \mathbf{d} obtenidos del modelo ajustado a los datos reales del 2018. El factor aleatorio se dará con el parámetro de habilidad θ , que será simulado desde una Normal multivariada de media 0 y varianza I .
- **Paso 2:** Se estiman los modelos MGRM-bifactor de 3 dimensiones específicas para cada uno de los sets de datos simulados en el **Paso 1**, ya que se necesitan los parámetros de discriminación y locación para generar los nuevos modelos.
- **Paso 3:** Una vez estimados los nuevos parámetros y fijando los parámetros del modelo obtenido con los datos reales del 2017, se realiza el proceso de optimización de la ecuación (3), usando las funciones creadas en R para encontrar las constantes de linking para cada proceso de simulación.

- **Paso 4:** Finalmente con las constantes de linking identificadas para los nuevos modelos, bastará calcular el valor esperado para cada una de las constantes de los n conjuntos de datos simulados. Una vez obtenidos sus valores, si estos coinciden con los resultados obtenidos en la Tabla 1 será posible comprobar que el método funciona.

Debido al costo computacional que genera el proceso de optimización del **Paso 4**, se simularán solo $n = 100$ conjuntos de datos, por lo que se espera que los resultados sean bastante cercanos y no exactamente iguales a lo que se presenta en la Tabla 1.

Resultados de la Simulación

Una vez terminado el proceso de simulación descrito en la sección anterior, se determina que en promedio las constantes de linking para los nuevos conjuntos de datos simulados son las descritas en la Tabla 2:

u_G	u_1	u_2	u_3	b_G	b_1	b_2	b_3
0.939	0.972	0.964	1.064	0.166	-0.108	-0.103	-0.094

Tabla 2: Promedio de las constantes de Linking estimadas desde 100 conjuntos de datos simulados

Se puede ver que la estimación es cercana a las constantes de equating reales que se consiguieron en la Tabla 1, resultado positivo, respecto al objetivo de esta investigación. Además, el error de estimación de las constantes es aproximadamente de un 0.05% en promedio, por lo que es posible afirmar que la extensión del método de Haebara para datos politómicos funciona.

Conclusión/ Discusión

Para resumir, el objetivo principal de esta investigación, era encontrar constantes de linking que permitieran comparar puntajes de pruebas no cognitivas a lo largo del tiempo. Para ello se usó la base de datos de los IDPS, específicamente en el área de Clima de Convivencia Escolar donde mediante la extensión del método Haebara para modelos multidimensionales con estructura bi-factor, se quería comprobar su funcionamiento.

El proceso de simulación de esta investigación fue bastante complejo, debido al costo computacional que se necesitaba para encontrar las constantes de equating. Se hicieron diversos intentos, para disminuir el tiempo de ejecución del código, sin presentar ninguna mejora. A pesar de buscar diversas formas de optimizar el algoritmo, no se logró y se tuvo que recurrir

al uso de un servidor externo para llevar a cabo la ejecución del código. Por lo mismo, es que solo se crearon $n = 100$ sets de datos, para el proceso de simulación.

A pesar del costo computacional del proceso, con los resultados presentados en la Tabla 2 se logra respaldar la validez y robustez del método propuesto. Lo que no es solo un cálculo de valores que permite tener puntajes del Ambiente Seguro en la misma escala sino que, permite dar validez a las comparaciones en este constructo, tema importante, ya sea para la ACE o para los propios establecimientos educacionales.

La metodología para obtener constantes de equating basada en curvas características presentada en este trabajo puede ser extendida en varias direcciones. Por ejemplo, existe otro método similar presentado en Stocking & Lord (1938) en donde el proceso de optimización se basa en una suma levemente distinta a la utilizada en el método de Haebara. Otra posible extensión involucra la incorporación de información adicional para obtener las constantes de equating, ya sea mediante el uso de covariables de contexto (en el caso de cuestionarios autoreportados escritos), o por medio de información adicional en forma de “process data” (en el caso de mediciones realizadas en plataformas digitales).

En conclusión, este trabajo destaca la importancia de considerar aspectos no-cognitivos en la evaluación de la calidad de la educación y proporciona una metodología que permite hacer equivalentes las evaluaciones en el área de Clima de Convivencia Escolar que eventualmente se puede extendedr a otras dimensiones. Este trabajo deja en manos de las autoridades educativas datos importantes sobre como están siendo educados sus estudiantes y cuales son los puntos débiles que habría que mejorar para continuar potenciando el sistema educacional en beneficio de todos.

Agradecimientos:

Este trabajo fue parcialmente financiado por el proyecto Fondecyt 1230968.

Referencias

- Agencia de Calidad de la Educación. (2023, 11 julio). *Agencia de Calidad de la Educación*. Agencia - Agencia de Calidad de la Educación. <https://www.agenciaeducacion.cl/>
- Aprendemos Juntos 2030. (2019, 2 diciembre). «*Cuando los migrantes llegamos al mar, la batalla está perdida*». Ousman Umar, emprendedor social [Vídeo]. YouTube. <https://www.youtube.com/watch?v=sXhA8MbaaZA>
- Bonate, P. L. (2001). A brief introduction to Monte Carlo Simulation. *Clinical Pharmacokinetics*, 40(1), 15-22. <https://doi.org/10.2165/00003088-200140010-00002>

- Duckworth, A., & Yeager, D. S. (2015). Measurement matters. *Educational Researcher*, 44(4), 237-251. <https://doi.org/10.3102/0013189x15584327>
- González J. & Robledo H. (2023) The Comparability of School Climate Measures at Schools in Chile.
- Haebara, T. (1980). Equating logistic ability scales by a weighted least square method. *Japanese Psychological Research*, 22(3), 144-149. <https://doi.org/10.4992/psycholres1954.22.144>
- Kim, K. Y. (2019). Two IRT fixed parameter calibration methods for the bifactor model. *Journal of Educational Measurement*, 57(1), 29-50. <https://doi.org/10.1111/jedm.12230>
- Kolen, M. J., & Brennan, R. L. (2014). Test equating, scaling, and linking: Methods and practices: Third edition. *En Springer eBooks*. <https://doi.org/10.1007/978-1-4939-0317-7>
- Liu, Y., Li, Z., & Liu, H. (2019). Reporting valid and reliable overall scores and domain scores using bi-factor model. *Applied Psychological Measurement*, 43(7), 562-576. Retrieved from <https://doi.org/10.1177/0146621618813093> doi : 10.1177/0146621618813093
- R. Philip Chalmers (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, 48(6), 1-29. doi : 10.18637/jss.v048.i06
- Stocking, M., & Lord, F.M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 207-210.