



Plan de análisis del piloto ERCE 2019 Cuestionarios de Factores Asociados Jorge Manzi MIDE UC

Objetivos de la presentación

1. Presentar las actividades involucradas en el análisis del comportamiento empírico de los ítems de los cuestionarios de factores asociados y módulo de innovación, en la fase piloto para el ERCE, orientados a identificar los factores asociados a los logros de aprendizaje de estudiantes de 3º y 6º grado en Latinoamérica y el Caribe.
2. Presentar las propiedades a estudiar, la metodología con que se revisarán y los criterios para considerar que el comportamiento de los ítems que componen los cuestionarios piloto de factores asociados y módulo de innovación, son aceptables.

- **Marco general de trabajo**

Las propiedades por estudiar están orientadas a asegurar que el comportamiento de los ítems es apropiado para su uso en la aplicación definitiva y se enmarcan en los tres pilares fundamentales de la medición establecidos por los Estándares de AERA, APA & NCME (2014): **confiabilidad, validez e imparcialidad.**

Así, los procedimientos aquí descritos buscan producir evidencias de la confiabilidad de los puntajes totales, evidencias de validez como soporte para las interpretaciones de los índices generados, así como evidencias de la comparabilidad de los puntajes entre diferentes grupos.

Marco general de trabajo

- Entre los cuestionarios de Factores Asociados de Estudiantes de Tercer Grado, y Sexto Grado, junto al cuestionario de Familias, Docentes y directores, se cuenta con una serie de ítems agrupados en escalas.
- Cada uno de estos requiere ser sometido a una serie de evaluaciones que permitirán contar con evidencias respecto a la pertinencia y calidad de las respuestas a estos ítems. Así, se garantizará que los puntajes generados puedan ser **interpretados en una misma dirección** (unidimensionalidad), con **niveles aceptables de precisión** (confiabilidad de los puntajes generados), y con **razonable comparabilidad entre los países** participantes.
- Para evaluar la pertinencia de los ítems a los puntajes totales esperados, se llevarán a cabo diferentes rutinas de análisis, las cuales se identificarán a continuación.

Análisis de respuestas

Respuestas perdidas y casos válidos

- Se estima la cantidad de **respuestas omitidas por ítem a nivel regional, para cada grado, y para cada país.**
- Este análisis se ejecuta analizando los **patrones de omisión por ítem** y los **patrones de omisión por evaluado**. La información levantada a partir de este análisis permite ir tomando decisiones asociadas a, por ejemplo, la extensión de los cuestionarios, la eliminación de ítems o la exclusión de examinados que hayan mostrado una alta tasa de omisión.

Análisis de respuestas

Distribución de respuestas

- Se generan **tablas de descriptivos generales**, que incluyen medias, desviación estándar, percentiles e histogramas. Estas tablas permiten identificar distribuciones de respuestas asimétricas (en que la mediana se separa de forma importante del punto medio de las respuestas), y si algunas categorías de respuesta no están siendo empleadas por los participantes en cada ítem.

Modelo de Medición

Dimensionalidad

- Se evalúa la cantidad de factores latentes requeridos para recuperar la matriz de covarianza de las respuestas a los ítems analizados. Para estos fines se emplea el **análisis paralelo**, en que se contrasta la cantidad factores extraíbles de la matriz de respuesta con la cantidad factores extraíbles de un conjunto de matrices aleatorias de similares características.
- También se emplea una rutina de **Análisis Factorial Exploratorio**. Se retiene la cantidad de factores, según la solución que presente de forma conjunta los siguientes indicadores de ajuste: CFI y TLI mayores .95, y el indicador RMSEA menor a .06, así como indicadores con cargas factoriales iguales o superiores a .4.

Modelo de Medición

Ajuste global

- Seguido de la definición de la cantidad de factores esperados, se ajusta un **Modelo Factorial Confirmatorio**, en que se evalúan las cargas factoriales de los indicadores presentes en el modelo, esperando que sean superiores a .4.
- Ítems que presentan cargas factoriales menores a .4, se consideran como de baja relación con el factor latente asociado.

Análisis de Ítems

Correlaciones Ítem-test

- Se calculan tablas de análisis de ítem, indicando cantidad de observaciones incluyendo correlación ítem test. Estas tablas se producen para el total de la región, y también por país, y para cada factor retenido.

Análisis de Ítems

Análisis de ítems vía IRT

- Se ajustan modelos de crédito parcial sobre el conjunto de ítems que se espera conformen un modelo unidimensional. Con estos análisis se revisará:
 - a) La **distribución de ítems a lo largo de la variable latente esperada** mediante la inspección de Wright maps, los que permiten identificar la posición relativa de los ítems respecto a la distribución de la variable latente esperada, indicando qué porción de esta variable pierde información en caso de excluir ítems del conjunto.
 - a) El **ajuste específico de cada una de las categorías de respuesta al modelo ajustado**, mediante la inspección de residuales (INFIT). Respuestas con residuales mayores a 1.3 o menores a .75 serán consideradas como evidencias de poco ajuste al modelo, siendo candidatas a ser removidas de la escala esperada. Categorías que presentan bajas respuestas en los extremos del espacio de respuesta, podrían ser colapsadas para aminorar este fenómeno.

Análisis de Ítems

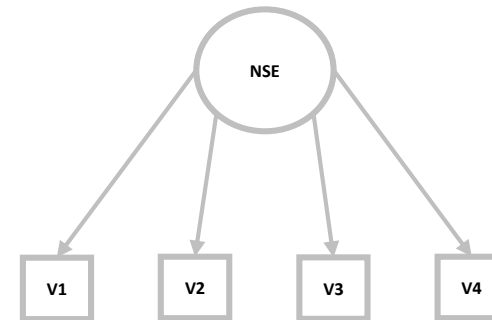
Confiabilidad

- Se generarán tablas de confiabilidad empleando el coeficiente Alfa de Cronbach para el puntaje total. Este índice se calculará para la región y para cada país participante en ERCE. **Confiabilidades menores a .6 serán consideradas como problemáticas** y sugieren someter al conjunto de ítems a revisión para la conformación del ensamblaje final.

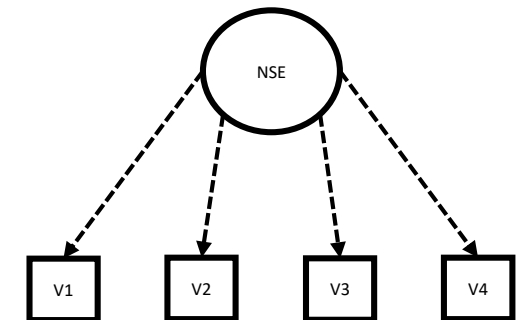
Comparabilidad

Invarianza entre países

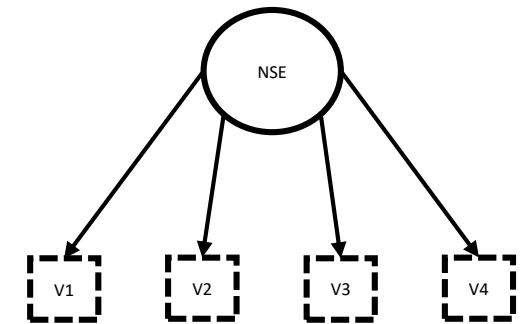
- El conjunto de ítems analizados es sometido a un **análisis factorial confirmatorio multi-grupo**, en el cual los países son especificados como grupo de comparación. Con este análisis se evalúa el modelo general establecido para cada escala, y se ajustan modelos que permiten evaluar distintos niveles de invarianza entre países (configural, métrica y escalar).



Configural: igual relación entre ítems y constructo, pero no permite ningún tipo de comparación significativo.



Métrica: igualdad en cargas factoriales, permite comparación de asociaciones (e.g. coeficientes de regresión entre países).



Escalar: igualdad de interceptos/umbrales, permite comparar promedios entre países.

MUCHAS GRACIAS