

SERIE DE DOCUMENTOS TÉCNICOS / 1

Aprender 2017

NOTAS TÉCNICAS

SERIE DE DOCUMENTOS TÉCNICOS / 1

Aprender 2017

NOTAS

TÉCNICAS

AUTORIDADES

Presidente

Ing. Mauricio Macri

Ministro de Educación

Dr. Alejandro Finocchiaro

Jefe de Gabinete del Ministerio de Educación

Cdor. Javier Mezzamico

Secretaria de Evaluación Educativa

Prof. Elena Duro

Secretario de Gestión Educativa

Lic. Manuel Vidal

Secretaria de Políticas Universitarias

Mg. Danya Tavela

Secretaria de Innovación y Calidad

Sra. María de las Mercedes Miguel

Las metodologías utilizadas en el área de la investigación educativa comprenden un conjunto de estrategias, recursos y procedimientos diversos, que buscan adecuarse y analizar en mayor profundidad los distintos aspectos del fenómeno de estudio.

A continuación se definen algunos conceptos metodológicos claves para el entendimiento de este informe. Estas notas pretenden esclarecer y acercar algunos criterios fundamentales en el ámbito de la evaluación educativa, de manera breve y accesible.

TEORÍA DE RESPUESTA AL ÍTEM

Las preguntas (ítems) de las pruebas Aprender 2017 se analizaron en base al modelo de la Teoría de Respuesta al Ítem (TRI), en la que se basa gran parte de las evaluaciones estandarizadas internacionales, así como las ONE (2005-2015) y Aprender 2016. Sus supuestos básicos en el caso del modelo de dos parámetros son:

- a) Para cada disciplina evaluada en cierto año, el estudiante posee una habilidad, rasgo latente (no observado) o competencia, que puede asociarse con un número real θ .
- b) La probabilidad de responder correctamente un cierto ítem es una función creciente y continua del valor θ .
- c) La unidimensionalidad: una sola habilidad explica la probabilidad de respuesta a cada ítem en cierta disciplina.

En el modelo logístico a dos parámetros la probabilidad de respuesta correcta se modeliza mediante:

$$\Pr(I_i = 1 | \theta) = \frac{\exp(D \cdot a_i \cdot (\theta - b_i))}{1 + \exp(D \cdot a_i \cdot (\theta - b_i))}$$

donde D es un escalar que normalmente se fija en 1,76;

a_i es la discriminación del ítem i (el ítem funcionará bien si este parámetro es mayor a 0);

y b_i es la dificultad del ítem.

En ONE 2013 se aplicó un modelo logístico a un parámetro; en Aprender 2016 y 2017 se optó por un modelo logístico de dos parámetros, que aporta más flexibilidad al permitir discriminaciones distintas entre los ítems. Esto implica que, para cada ítem, se estimó no sólo su dificultad sino también su capacidad de discriminación.

NIVELES DE DESEMPEÑO

Una vez asignado el puntaje de los estudiantes se los clasifica en cuatro niveles de desempeño. Este proceso se realizó en Aprender 2016 y 2017 mediante el método llamado Bookmark utilizado en la mayoría de los países de la región. El proceso de establecimiento de los puntos de corte que determinarán los niveles de desempeño, se basa, para cada año y disciplina, en el juicio experto de un grupo de docentes con representatividad federal que se reúnen para dialogar, debatir y llegar a consensos con el objetivo de establecer los puntos de corte entre los distintos niveles de desempeño de los estudiantes que participaron de la evaluación nacional Aprender.

El método Bookmark, aplicado en diversas rondas de trabajo, permite determinar aquellos contenidos, habilidades y capacidades que los estudiantes de cada nivel de desempeño deben dominar en las disciplinas y años de escolarización evaluados.

De este modo, el establecimiento de estándares del método Bookmark comprende cuatro tareas básicas: 1) selección de la cantidad de niveles de desempeño que se desean establecer; 2) elección de los nombres o etiquetas de cada nivel; 3) redacción de los descriptores de cada nivel; y 4) establecimiento de los puntajes de corte correspondientes.

Los niveles de desempeño involucran una etiqueta y una serie de descriptores que permiten conocer en mayor detalle cuáles son los contenidos y capacidades que efectivamente los estudiantes de un determinado nivel dominan. En nuestro país, a partir de las pruebas Aprender 2016, se han establecido cuatro niveles de desempeño: Por debajo del nivel básico, Básico, Satisfactorio y Avanzado.

Estos cuatro niveles fueron definidos a partir del trabajo realizado por más de 200 docentes en los talleres federales Bookmark realizados durante febrero de 2017. Durante estas jornadas se determinaron dos puntos de corte que determinaron tres niveles de desempeño (Bajo, Satisfactorio, Avanzado) para las evaluaciones nacionales de Matemática y Lengua de 5°/6° año y 2°/3° de nivel secundario y de 6° y 3° grado de nivel primario y también para Ciencias Naturales y Ciencias Sociales de 5°/6° año de nivel secundario.

Posteriormente al establecimiento de los puntos de corte por parte de los docentes y tomando como insumo principal sus aportes, se subdividió el nivel Bajo en dos niveles (Por debajo del nivel básico y Básico) para poder analizar en mayor profundidad los contenidos y capacidades que dominaban los estudiantes de este nivel.

En el nivel Por debajo del nivel básico se incluyeron a los estudiantes cuyo puntaje se distanciaba más de un 25% del punto de corte entre el nivel Bajo y el nivel Satisfactorio. En tanto, en el nivel Básico se incluyeron aquellos estudiantes cuyos puntajes no superaban esa distancia. Esta subdivisión y el criterio utilizado es similar a los ajustes que suelen realizarse en las evaluaciones estandarizadas a nivel regional e internacional.

Aquellos con un puntaje cercano a Satisfactorio y aquellos cuya puntuación está alejada de este nivel.

Si θ_1 es el punto de corte que marca el inicio de los puntajes correspondientes al nivel Satisfactorio y θ_2 es el punto de corte que marca el inicio de los puntajes correspon-

dientes al nivel Avanzado, entonces las categorías con los niveles de desempeño quedan definidas del siguiente modo:

POR DEBAJO DEL NIVEL BÁSICO: $\theta \leq 0.75 \cdot \theta_1$

BÁSICO: $0.75 \cdot \theta_1 < \theta \leq \theta_1$

SATISFACTORIO: $\theta_1 < \theta \leq \theta_2$

AVANZADO: $\theta_2 < \theta$

Una vez establecidos los cuatro niveles de desempeño para las pruebas Aprender 2016 se elaboraron los descriptores de cada nivel en los cuales se detallan los saberes y capacidades dominados por los estudiantes de cada nivel. Los contenidos de aprendizajes detallados para cada descriptor de nivel siguen los lineamientos definidos en los Núcleos de Aprendizaje (NAP) acordados a nivel nacional por las 24 jurisdicciones en el marco del Consejo Federal de Educación.

Para Aprender 2017 secundaria, se utilizaron los criterios definidos en Aprender 2016 ya que las materias evaluadas coincidieron en ambos años (Matemática y Lengua). En el caso de primaria, se realizaron nuevos talleres federales Bookmark para las asignaturas Ciencias Sociales y Ciencias Naturales debido a que no habían sido evaluadas el año anterior y por lo tanto era necesario establecer los puntos de corte en dichas áreas. Se realizaron en febrero de 2018 y participaron 40 docentes de todo el país.

ETAPAS DEL PROCESAMIENTO

Para la obtención de los puntos de puntajes y puntos de corte de las pruebas Aprender 2017, se procesaron los instrumentos de evaluaciones estandarizados siguiendo las siguientes fases de trabajo:

- Depuración de las tablas relevadas, tanto de las pruebas como de las correspondientes a las variables de contexto.
- Ajuste preliminar del modelo y estimación de los parámetros de los ítems.
- Análisis del funcionamiento de los ítems: discriminación, funcionamiento diferencial, entre otros. Evaluación del ajuste del modelo.
- Estimación definitiva del modelo habiendo eliminado ítems con funcionamiento deficiente.
- Cálculo del puntaje estandarizado de los estudiantes, para cada año y disciplina.

En el caso de Ciencias Naturales y Ciencias Sociales de 6° grado de nivel primario, se establecieron los puntos de corte que determinaron los tramos de logro en cada una de las disciplinas en base a los resultados de los talleres Bookmark 2018. En el caso de secundaria, para poder aplicar los puntos de corte establecidos durante 2017, los puntajes obtenidos en las pruebas Aprender 2017 para Matemática y Lengua 5°/6° año de nivel secundario se llevaron a la escala 2016 (escalamiento).

TRATAMIENTO DE LA NO RESPUESTA

La no respuesta comprende tanto el caso de la no concurrencia de estudiantes a la prueba como el de estudiantes que concurren pero no responden todas las preguntas. Existen dos tipos de no respuesta:

- Estudiantes participantes de la prueba, pero que algunas de sus respuestas figuran en la base en missing: preguntas no respondidas, o que su respuesta fue mal leída por ejemplo. Denominamos a ésta no respuesta parcial.
- Estudiantes no participantes de la prueba, o que fueron por algún motivo eliminados del proceso de puntuación. Denominamos a esta no respuesta no respuesta total.

Para el primer caso, se realiza (con algunas limitaciones que veremos) el procedimiento de imputación. En el segundo caso, se aplica el procedimiento de ponderación.

IMPUTACIÓN

Imputar significa asignar una respuesta a una pregunta que no fue respondida. Esta respuesta debe ser una respuesta válida, es decir, estar dentro del rango de opciones admitidas.

9

¿Para qué se imputa?

Para mejorar las estimaciones y cálculos realizados con los datos.

Al administrar una prueba e presenta habitualmente el caso de que haya 'ítems' no respondidos (ítems en blanco).

Frente a la situación de un ítem en blanco existen tres alternativas:

- 1 Dejar la respuesta en blanco: esto significaría favorecer a los estudiantes que respondieron menos ejercicios de la prueba (al no contabilizar las respuestas a estos ítems en blanco, que por haber sido dejados en blanco se puede suponer que el estudiante no sabía la respuesta).
- 2 Imputar la respuesta faltante como ítem respondido incorrectamente: esto implicaría penalizar a los alumnos de manera excesiva, ya que existe la probabilidad de que hubieran respondido correctamente el ítem.
- 3 Calcular la probabilidad de que el estudiante haya respondido ese ítem en forma correcta o incorrecta, en función de las respuestas dadas por él en otros ítems, su sexo, sector de gestión, repitencia, etc.. Esta fue la alternativa seleccionada para imputar en Aprender 2016.

La imputación es una práctica habitual en la práctica estadística al trabajar con datos

incompletos (ítems dejados en blanco).

¿Qué se imputa en Aprender 2017?

Al igual que en Aprender 2016, en el dispositivo de 2017 se imputarán sólo algunas respuestas faltantes de la prueba (correspondientes a ejercicios de la misma); lo que se imputa o adjudica es un valor que corresponde a ítem respondido correctamente/no correctamente. No se imputaron preguntas de cuestionarios de contexto.

El procedimiento de imputación se realizó solo en aquellos casos en los que los estudiantes respondieron por lo menos el 50% de los ejercicios de la evaluación.

Los ejercicios imputados fueron solo aquellos que quedaron sin respuesta de forma intercalada, en el marco de otros ejercicios que sí fueron respondidos. Si quedaron sin respuesta los últimos cinco ejercicios de la prueba, los mismos no fueron imputados debido a que estos ítems finales, de haber sido dejados en blanco puede asumirse la existencia de un efecto cansancio.

Por ejemplo, si un estudiante respondió, de un total de 10 ejercicios, los primeros cinco y dejó en blanco los últimos cinco; estos últimos no fueron imputados. Si, por el contrario, el estudiante dejó sin responder los ejercicios 2 y 5 pero respondió desde el ejercicio 6 hasta el final, las respuestas de los ejercicios 2 y 5 fueron imputadas.

Para adjudicar a una pregunta no respondida el valor de ítem respondido correctamente / incorrectamente (es decir imputarlo) se toman en cuenta las respuestas dadas por el alumno en los otros ejercicios, su sexo, el sector de gestión y la repitencia. En función de esto, se calcula mediante un modelo (logit encadenado) la probabilidad que tiene ese estudiante de haber contestado correctamente o incorrectamente el ítem faltante. Es esta probabilidad la que determina el valor final que se adjudica a la pregunta en blanco de ese estudiante.

PONDERACIÓN

¿Qué es ponderar?

Cuando la distribución de la frecuencia de una variable no coincide con su distribución en la población, se utiliza la ponderación para equilibrar el peso relativo de los casos. También se utiliza habitualmente la ponderación en las muestras para expandir la cantidad de casos y reflejar la cantidad total del universo.

¿Cómo y para qué se pondera?

Como se detalló anteriormente, cuando se plantea una evaluación como Aprender, se establece un universo de estudio que está compuesto por la totalidad de estudiantes que fueron convocados a responder la prueba (ya sea en el caso de los censos o de las muestras). Cuando la prueba se aplica puede suceder que no participe el 100% de los estudiantes a los que debería aplicarse la prueba. Cuando esto sucede, se genera un

desbalance provocado por la no participación del total del universo. Si se elaboraran porcentajes o promedios con estos datos, los resultados de la investigación no serían representativos del universo de estudio. Es por esta razón que se realizan los procesos de ponderación que tienen el objetivo de corregir estos posibles sesgos o desbalances.

La ponderación se aplica siempre sobre los estudiantes que participaron y sus respuestas. Y se utiliza al momento de procesar estadísticamente la base. Cuando se pondera sólo asignamos pesos relativos a los datos existentes, no se asignan valores a los estudiantes faltantes.

La ponderación en Aprender 2017 permite que a la hora de hacer los análisis se respeten los pesos poblacionales relativos de cada jurisdicción, sector de gestión y ámbito. En el caso de PBA, se incluyó, además, partido como variable para el cálculo de las ponderaciones a los efectos que se represente correctamente tanto el Gran Buenos Aires como el interior de la Provincia de Buenos Aires.

