



SERIE DE DOCUMENTOS TÉCNICOS / 3

APRENDER 2018 BOOKMARK

Secretaría de
Evaluación Educativa



Ministerio de Educación,
Cultura, Ciencia y Tecnología
Presidencia de la Nación

SERIE DE DOCUMENTOS TÉCNICOS / 3

APRENDER 2018 BOOKMARK

AUTORIDADES

Presidente

Mauricio Macri

Jefe de Gabinete de Ministros

Marcos Peña

Ministro de Educación, Cultura, Ciencia y Tecnología de la Nación

Alejandro Finocchiaro

Secretario de Gobierno de Cultura

Pablo Avelluto

Secretario de Gobierno de Ciencia, Tecnología e Innovación Productiva

Lino Barañao

Titular de la Unidad de Coordinación General del Ministerio de Educación, Cultura, Ciencia y Tecnología

Manuel Vidal

Secretaria de Evaluación Educativa

Elena Duro

Secretario de Gestión Educativa

Oscar Ghillione

Secretario de Políticas Universitarias

Pablo Domenichini

Secretaria de Innovación y Calidad Educativa

Mercedes Miguel

COORDINACIÓN

Prof. Elena Duro

EQUIPO A CARGO DE LA ELABORACIÓN DEL DOCUMENTO

María Aranguren

Augusto Horszowski

INTRODUCCIÓN

Las evaluaciones de aprendizajes tienen por objetivo conocer los contenidos y capacidades dominados por los estudiantes de un determinado año y área.

Normalmente, para poder brindar información útil acerca de los resultados obtenidos por los estudiantes, se suele clasificar su desempeño en distintos niveles de acuerdo al grado de dominio de los contenidos y capacidades evaluadas.

Existen diferentes procedimientos utilizados para determinar los niveles de desempeño de las evaluaciones de aprendizaje. Las pruebas Aprender 2018 continúan un camino ya iniciado por los Operativos Nacionales de Evaluación (ONE) y por las pruebas Aprender 2016 y 2017. En este sentido, la determinación de los niveles de desempeño y de los puntos de corte para las pruebas administradas, ha sido realizada siguiendo los procedimientos indicados por el método Bookmark, utilizado en ocasiones precedentes en nuestro país y en la región y recomendado por la literatura internacional debido a sus características y beneficios.

A continuación, se presentan algunas definiciones acerca de las tareas implicadas en el establecimiento de estándares, las particularidades principales del método Bookmark y cómo ha sido implementado en nuestro país para la determinación de los puntos de corte de las pruebas de Lengua y Matemática de 6° grado de nivel primario de Aprender 2018.

ESTABLECIMIENTO DE ESTÁNDARES

En el ámbito educativo, el establecimiento de estándares se refiere a la selección de los criterios que serán utilizados para valorar las puntuaciones de una prueba e involucra una serie de procedimientos que responden a la necesidad de poder comunicar de una manera eficiente y articulada los resultados de las pruebas de aprendizajes.

Actualmente, la gran mayoría de las pruebas educativas se definen como pruebas criterioales. Esto significa que el conjunto de ítems incluido en las pruebas representa un dominio o un criterio claramente definido (Hambleton & Swaminathan, 1978). Esta característica permite que los resultados obtenidos por los estudiantes sean valorados de acuerdo a ciertos estándares o criterios absolutos referidos al dominio que se está evaluando. De esta manera, se puede decir que un estudiante ha logrado adquirir en el proceso de aprendizaje determinados contenidos y capacidades delimitados en un dominio específico.

Los estándares con los que se evalúan los logros de los estudiantes se traducen en lo que comúnmente conocemos como niveles de desempeño. De ahí que el establecimiento de estándares suponga cuatro tareas básicas:

1- selección de la cantidad de niveles de desempeño que se desean establecer, 2- selección de las etiquetas de cada nivel, 3- determinación de los puntajes de corte correspondientes y 4- redacción de los descriptores de cada nivel.

Los niveles de desempeño definen cuáles son las capacidades y contenidos que dominan los estudiantes de un determinado nivel. De acuerdo a las necesidades de la evaluación se pueden delimitar dos, tres, cuatro, cinco o seis niveles de desempeño, aunque se recomienda utilizar entre tres y cuatro (Perie, 2008). Una vez seleccionada la cantidad de niveles de desempeño será necesario definir sus etiquetas que permitirán sintetizar en una palabra una serie de ideas, facilitando de esta manera la comunicación (Herrera Ortiz, Benavides Posadas & Monroy Cazorla, 2009).

Existen diferentes nominaciones para los niveles de desempeño que varían de acuerdo a los usos y costumbres de cada contexto cultural (e.g.: Básico, Competente, Avanzado; Elemental, Satisfactoria, Sobresaliente; Limitado, Básico, Acelerado, Avanzado; Lejos de Suficiente, Por debajo de suficiente, Suficiente, Competente, Avanzado; No alcanza los objetivos, Alcanza los objetivos, Alcanza los objetivos satisfactoriamente.). Acorde a ello, las etiquetas son seleccionadas de modo que representen mejor las categorías que se postulan.

Los puntos de corte implican la operacionalización de esos criterios o conceptos en un valor numérico. Por ejemplo, se podría suponer que un estudiante de nivel Satisfactorio logra un dominio adecuado o intermedio del conjunto de contenidos y capacidades cognitivas evaluados en una determinada área y año, y que el punto de corte entre el nivel Satisfactorio y el nivel Avanzado equivale a un valor theta (θ) de 350.

Para determinar los puntos de corte se utilizan diferentes procedimientos estadísticos y/o metodológicos como por ejemplo, el método Bookmark (Lewis, Mitzel, & Green, 1996; Lewis, Mitzel, Green, & Patz, 1999), el método de Angoff (1971), el método Bow (Kahl, Crockett, DePascale, & Rindfleish, 1994), el método de Juicio Analítico (Plake & Hambleton, 2001), entre otros. Cada metodología sugiere una serie de procedimientos y actividades para poder colocar los puntos de corte.

Por último, una vez seleccionada la cantidad de niveles de desempeño y las etiquetas de cada nivel, y establecidos los puntos de corte, se procederá a redactar los descriptores. Éstos refieren a las capacidades y contenidos dominados por los estudiantes de cada nivel de desempeño. En otras palabras, los descriptores establecen el significado y alcance de los puntos de corte. De esta manera, ayudan a los maestros, directivos, padres y actores políticos a interpretar qué es lo que los estudiantes de cada nivel conocen y pueden hacer, y potencialmente, qué no conocen y no pueden hacer.

Para ser precisos, no existen métodos para seleccionar la cantidad de niveles de desempeño o las etiquetas de los niveles, ni tampoco para describir los contenidos y capacidades dominados por los estudiantes de cada nivel. En cambio, existen diferentes propuestas metodológicas para poder determinar los puntos de corte de las pruebas. Cualquier procedimiento utilizado para el establecimiento de los puntos de corte requiere de las otras tres tareas y, en conjunto, las mismas refieren al establecimiento de estándares.

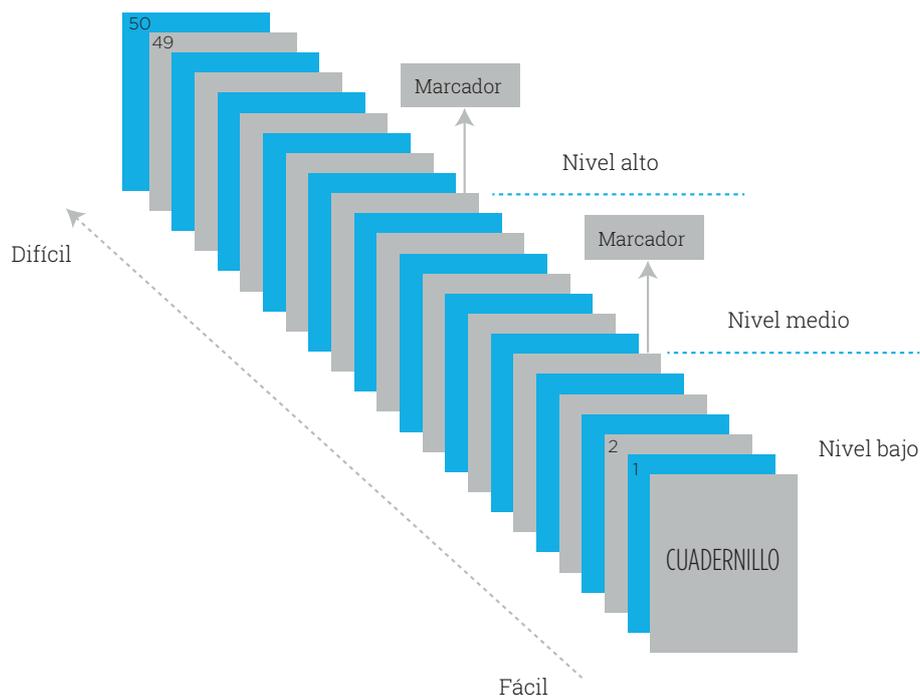
MÉTODO BOOKMARK

Tal como se ha mencionado anteriormente, existen diferentes procedimientos para poder determinar los puntos de corte. En nuestro país, desde los ONE 2010-2013 hasta el programa Aprender 2016 y 2017, se ha utilizado el método Bookmark. Este método permite delimitar y expresar en un valor numérico aquellos contenidos y capacidades que los estudiantes de cada nivel de desempeño deben dominar en cada disciplina y año de escolarización a partir del juicio realizado por un conjunto de expertos.

Los jueces u expertos convocados para la determinación de los puntos de corte son docentes en ejercicio del área y año evaluado. Para lograr este objetivo, los docentes deberán revisar un conjunto limitado de los ítems que componen la prueba, que son presentados en un cuadernillo siguiendo un orden de dificultad creciente (figura 1). Este orden de dificultad se corresponde con la dificultad empírica que tuvo cada ítem para los estudiantes evaluados.

El ordenamiento de los ítems puede realizarse gracias a que los resultados de las pruebas criterioles, como lo son las pruebas Aprender, pueden ser analizados desde la Teoría de Respuesta al Ítem (TRI) o Teoría del Rasgo Latente. El uso de la TRI permite estimar tanto el índice de dificultad de los ítems como el índice de discriminación. El índice de dificultad permite observar qué tan fácil o difícil les resultó ese ítem a los estudiantes. El índice de discriminación permite ver qué tan bien un ítem o un conjunto de ítems discrimina entre estudiantes que tienen un alto y un bajo desempeño en las áreas evaluadas. Por otra parte, la TRI permite calcular un puntaje theta (θ) de cada estudiante, valor que representará la competencia de ese estudiante en base a las respuestas que haya dado al conjunto de ítems que integran la prueba.

La TRI es también una metodología que se adapta a los estudios longitudinales, que apunta a evaluar a lo largo del tiempo la evolución de los desempeños. Para una exposición más detallada sobre la metodología TRI se puede consultar el documento de Notas Metodológicas (Secretaría de Evaluación Educativa [SEE], 2017).



A partir de la revisión del material, los docentes deben identificar aquellos ítems marcadores, es decir, aquellos ítems que separan un nivel de desempeño de otro. La cantidad de marcadores a establecer dependerá de la cantidad de puntos de corte que se deseen determinar. En el caso estándar de tres niveles de desempeño, cada docente deberá identificar dos ítems marcadores, uno para cada punto de corte.

La identificación de los marcadores es llevada a cabo en tres rondas de trabajo (Lin, 2006). Cada una tiene el objetivo de generar un espacio donde puedan debatirse las diferencias y perspectivas de los docentes, ya sea en pequeños grupos de trabajo como en el conjunto. La intención del debate y de las rondas es colaborar al consenso y reducir las diferencias entre los participantes.

Al finalizar la última ronda, se espera que cada docente haya colocado sus marcadores para cada punto de corte y se haya llegado a un acuerdo en relación a los contenidos y capacidades que se espera que los estudiantes de un determinado nivel dominen, y que éstos se vean reflejados en la elección de un ítem marcador.

El ítem marcador no equivale al punto de corte. De ahí que no sea necesario que todos los docentes elijan el mismo ítem marcador. Sin embargo, sí se requiere que haya un acuerdo conceptual en cuáles son los contenidos y capacidades que un estudiante de un determinado nivel de desempeño debe dominar ya que en caso contrario las diferencias entre los marcadores y entre las expectativas de los docentes serán muy grandes. En condiciones normales –en donde el cuadernillo de ítems ordenados refleja todo el continuo de dificultad del dominio–, si un docente supone que el ítem marcador que separa el nivel Bajo del nivel Medio está en el número 3 y otro en el número 27, la amplitud de esta diferencia manifiesta que ambos están utilizando diferentes criterios para determinar cuáles son los contenidos y capacidades que dominan los estudiantes del nivel Bajo (no olvidemos que los ítems se encuentran ordenados según el nivel de dificultad empírico resultante del procesamiento de los datos). No obstante, si un docente coloca su marcador en el ítem 6 y otro en el ítem 7 y no hay una amplia diferencia en lo que a la dificultad del ítem respecta, esto indica que ambas posiciones son perfectamente admisibles.

Ahora bien, cada marcador de cada docente se corresponderá con un valor de habilidad θ (theta). Para la obtención de los puntajes de corte se promedian los distintos valores θ correspondientes a cada uno de los participantes. Una vez obtenidos estos puntos de corte preliminares se procede a calcular la proporción de alumnos en cada uno de los niveles de desempeño.

DETERMINACIÓN DE PUNTOS DE CORTE APRENDER 2018

A partir de las pruebas Aprender 2016, en nuestro país se han establecido cuatro niveles de desempeño: Por debajo del nivel básico, Básico, Satisfactorio y Avanzado (Aranguren & Hoszowski, 2017).

Estos cuatro niveles fueron establecidos a partir del trabajo realizado por más de 200 docentes en los Talleres Federales Bookmark realizados durante el febrero de 2017. Los docentes convocados fueron seleccionados aleatoriamente garantizando que todas las jurisdicciones, sectores (estatal y privado) y ámbitos (urbano y rural) estuviesen representados de forma que los resultados obtenidos en dichas jornadas reflejaran adecuadamente a todo el territorio federal.

Durante los Talleres Federales Bookmark 2017 se determinaron tres niveles de desempeño (Bajo, Satisfactorio y Avanzado) para las evaluaciones nacionales de Matemática y Lengua de 5°/6° año y 2°/3° de nivel secundario y de 6° grado y 3° grado de nivel primario y también para Ciencias Naturales y Ciencias Sociales de 5°/6° año de secundaria.

En Aprender 2016, posteriormente al establecimiento de los puntos de corte por parte de los docentes y tomando como insumo principal sus aportes, se decidió subdividir 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 el nivel Básico se consideraron 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.

Una vez establecidos los cuatro niveles de desempeño para las pruebas Aprender 2016 se elaboraron los descriptores en los cuales se detalla 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 23 provincias y la Ciudad Autónoma de Buenos Aires en el marco del Consejo Federal de Educación.

Las pruebas Aprender 2018 se administraron en las áreas de Lengua y Matemática para los estudiantes de 6° grado de Nivel Primario. En estas evaluaciones se aplicaron los mismos puntos de corte establecidos durante 2017 para las pruebas Aprender 2016. Este procedimiento permite comparar los resultados de las pruebas Aprender 2016 con los resultados de las pruebas Aprender 2018. En otras palabras, permite: a) analizar el aumento o disminución en los puntajes obtenidos por los estudiantes y; b) estimar si hay cambios en la cantidad y/o porcentajes de estudiantes en cada uno de los niveles de desempeño establecidos. Tal como indica la metodología empleada con estas pruebas, para poder aplicar los puntos de corte establecidos durante el 2017, los puntajes obtenidos en las pruebas Aprender 2018 para Matemática y Lengua 6° año de primaria se llevaron a la escala 2016 (escalamiento).

Durante los talleres Bookmark realizados en el mes de febrero del 2017 participaron alrededor de 40 docentes de las áreas de Matemática y Lengua de 6° año del nivel primario, que representaron las distintas jurisdicciones, sectores y ámbitos y se establecieron los dos puntos de corte para los tres niveles Bajo, Satisfactorio y Avanzado.

Teniendo en consideración que las pruebas Aprender 2018 de Matemática y Lengua, son pruebas que incluyen nuevos ítems (distintos a los de las pruebas de Aprender 2016) y, a fin de cumplir con el objetivo de dar información detallada acerca de las capacidades y contenidos dominados por los estudiantes de cada nivel de desempeño, se debieron cotejar que los descriptores redactados en los informes y documentos referidos a la prueba Aprender 2016 fueran similares a los de la prueba Aprender 2018, y en caso de haber algún cambio, realizar los ajustes necesarios. Esto es así porque, al contar con nuevos ítems en la prueba, es posible enriquecer tanto la descripción de las capacidades propias de cada nivel de desempeño como la descripción de los contenidos dominados en cada nivel de desempeño. No es esperable, sin embargo, que las capacidades y contenidos dominados

por los estudiantes de cada nivel varíen cualitativamente, pero sí puede haber pequeños cambios en relación a lo que está evaluando cada ítem en particular. Para permitir la comparabilidad, los puntos de corte se mantienen tal como fueron definidos en el taller Bookmark de febrero 2017.

CONCLUSIONES

El establecimiento de estándares es un proceso complejo que implica diferentes etapas e involucra a una pluralidad de actores y disciplinas. Las conclusiones obtenidas en este proceso sirven para conocer la calidad de los aprendizajes de los estudiantes y para planificar las estrategias que sirvan a la mejora de las prácticas de enseñanza. De ahí que muchas veces los resultados de estos procesos incidan sobre la sociedad en su conjunto.

Para la determinación de los niveles de desempeño y puntos de corte de las pruebas Aprender 2018 se utilizaron los puntos de corte obtenidos durante los talleres Bookmark realizados en febrero del 2017, garantizando la validez, la confiabilidad y la comparabilidad de los resultados.

Con el compromiso de continuar con la tarea iniciada a partir de las pruebas Aprender 2016, la información acerca de los niveles de desempeño de los estudiantes pretende dar a conocer los contenidos y capacidades que dominan los estudiantes de los niveles de desempeño: Por debajo del básico, Básico, Satisfactorio y Avanzado en las pruebas de Matemática y Lengua de 6° año del nivel primario en Aprender 2018.

Esperamos que este documento sirva para brindar evidencias acerca de la validez y confiabilidad de los resultados encontrados.

REFERENCIAS

Aranguren, M. & Hoszowski, A. (2017). Aprender 2016. Establecimiento de puntos de corte. Serie de Documentos Técnicos/5. Ministerio de Educación y Deportes de la Nación. Obtenido el 23 de febrero de 2018 en: <https://www.argentina.gob.ar/sites/default/files/manual-bookmark-595bd361cf4e7.pdf>

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), "Educational measurement" (pp. 508-600). Washington, DC: American Council on Education

Hambleton, R.K. & Swaminathan, H. (1978). Criterion-Referenced Testing and Measurement: A review of technical issues and developments. "Review of Educational Research", 40, 1-47.

Herrera Ortiz, M., Benavides Posadas, D. & Monroy Cazorla, L. (2009). "Establecimiento de estándares en un examen criterial. Cuaderno Técnico 3". Méjico, D.F.: Centro Nacional de Evaluación para la Educación Superior.

Kahl, S. R., Crockett, T. J., & DePascale, C. A. (1994, June). Using actual student work to determine cut-scores for proficiency levels: New methods for new tests. Paper presented at the National Conference on Large-Scale Assessment, Albuquerque. 13

Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996, June). Standard setting: A bookmark approach. In D. R. Green (Chair), "IRT-based standard setting procedures utilizing behavioral anchoring". Symposium conducted at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.

Lewis, D. M., Mitzel, H. C., Green, D. R. & Patz, R. J. (1999). "The bookmark standard setting procedure". Monterey, CA: McGraw-Hill.

Lin, J. (2006). The Bookmark Procedure for Setting Cut-Scores and Finalizing Performance Standards: Strengths and Weaknesses. "Alberta Journal of Educational Research", 52(1), 36-52

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark procedure: Psychological perspectives. In G. Cizek (Ed.), Setting performance standards: Concepts, methods, and perspectives (pp. 249-281). Mahwah, NJ: Erlbaum.

Perie, M. (2008). A Guide to Understanding and Developing Performance Level Descriptors. "Educational Measurement: Issues and Practice", 27(4), 15-29.

Plake, B., & Hambleton, R. (2001). The analytic judgment method for setting standards on complex performance assessments. In G. Cizek (Ed.), "Setting performance standards: Concepts, methods, and perspectives" (pp. 283-312). Mahwah, NJ: Erlbaum.

Secretaría de Evaluación Educativa (2017). APRENDER 2016. "Notas Metodológicas. Serie de Documentos Técnicos/2". Ministerio de Educación y Deportes de la Nación. Obtenido el 30 de Noviembre de 2017 en: <http://www.bnm.me.gov.ar/gigal/documentos/EL005591.pdf>

DOCENTES QUE PARTICIPARON DE LOS TALLERES FEDERALES BOOKMARK EN LENGUA Y MATEMÁTICA DE 6º GRADO DE PRIMARIA

Alzugaray, Sonia Andrea
Apertile, Ana
Baigorria, Sandra Haydee
Baliani, Liliana
Barón, María Del Luján
Brito, Manuel Alberto
Conte ,Marta
Elias, Guadalupe
García, Romina Valeria
Giampietro, Noemí
Godoy, Analía
Gramajo, Myriam Graciela
Gutierrez, Carmen
Gutierrez, Nancy Del Carmen
Jimenez, Franco Efraín
Juliá, Silvina
Maidana Rodriguez, Javier Abdón
Mancuso, David Orlando
Martinez, Graciela
Micheletti, Marisa Guadalupe
Morello, Paulina
Navarro, Tenca Marilina
Pedemonte, María Verónica
Pérez Castillo, Natalia Romina
Quintana, Mónica Juana
Rodriguez, Roxana Karina
Romero, María Alejandra
Ruiz, Natalia Alejandra
Selva, María Rosa
Trejo, Luis Carlos
Trigo, Sara Maria Lía
Vecchioni, Ileana
Vecchioni, María Laura
Vega, Sandra Beatriz
Villarruel, Sandra Analía

