

# EVALUACIÓN CON PREGUNTAS DE OPCIÓN MÚLTIPLE. DESAFÍOS Y OPORTUNIDADES

Alberto Picón<sup>1</sup>

## RESUMEN

Las pruebas con preguntas de opción múltiple (POM), originalmente aplicadas en 1916 con el nombre popular de *multiple-choice test*, han ido ampliando su espectro de aplicación durante este primer siglo de vida. Estas pruebas o instrumentos para colección de datos tienen dos usos distintivos y comunes: por un lado, una medida regular de los comportamientos de los evaluados; por el otro, la facilidad de recoger elementos para la toma de decisiones en poblaciones numerosas. El diseño, elaboración, aplicación y análisis de las POM se basan en el mismo marco teórico de cualquier diseño de investigación: la definición de ciertas variables que se operativizan a través de los ítems y que deben satisfacer el rigor científico de medir con eficacia lo que se intenta medir, y hacerlo con resultados confiables; en suma, validez y confiabilidad. Dependiendo del nivel educativo en que se aplican, las POM pueden dar cuenta del grado en que un estudiante ha desarrollado las competencias propuestas en los objetivos de la educación, por ejemplo, en cada nivel de la dimensión cognitiva, en términos de Bloom. Actualmente las POM pueden ubicarse en la intersección de la evaluación de aprendizajes, la psicopedagogía y la neurociencia.

**Palabras clave:** *pruebas de múltiple opción, numerosidad en educación superior, evaluación de aprendizajes, ciencias económicas, ciencias de la administración.*

La Facultad de Ciencias Económicas y de Administración está implementando el Plan de Estudios 2012 en un contexto de incremento sostenido de la matrícula. En este marco, los equipos docentes se encuentran ante la necesidad de revisar los contenidos y propósitos de formación, las estrategias de enseñanza, los materiales educativos utilizados y los sistemas de evaluación empleados, intentando contemplar, al mismo tiempo, las características actuales de los estudiantes a su ingreso. En este contexto de alta numerosidad estudiantil una de las formas de evaluación a la cual los equipos docentes están recurriendo cada vez con mayor frecuencia son las pruebas con preguntas de opción múltiple (POM). Las POM facilitan la obtención de resultados objetivos y estandarizados; pero presentan el desafío de trabajar en su diseño desde varias perspectivas. En lo pedagógico y didáctico, los equipos docentes deberían manejar con claridad cuáles son las potencialidades y limitaciones de este instrumento para evaluar los aprendizajes esperados y efectivos de los estudiantes en cada Unidad Curricular (UC).

## Crisis y oportunidad

Puede rastrearse el origen de las preguntas con respuestas de opción múltiple en la historia de los censos. Los Estados Unidos de América (EUA) los han realizado desde 1790<sup>2</sup>. Y puede señalarse la obra *Educational Psychology* (Thorndike, 1903) como el catalizador que reunió los estudios sobre el desempeño de los alumnos en edad escolar y aquella técnica. Se considera que el trabajo de Frederick Kelly, en la selección de reclutas para las fuerzas armadas, el primer hito en el uso de los test de opción múltiple para medir la capacidad de las personas y seleccionarlas en orden a su preparación o formación posterior. Un hito en el desarrollo de la tecnología educativa y de la selección de personal. En el primer caso, para medir conocimientos y predecir resultados futuros; en el segundo, para ubicar a cada

---

<sup>1</sup> UDELAR-FCEA-Unidad de Apoyo a la Enseñanza

<sup>2</sup> <http://www.census.gov/>

empleado en el puesto adecuado a su capacidad. Y todo, porque había una Gran Guerra que demandaba tropas y quien las liderara en el campo de batalla: soldados y oficiales.

A pesar de que el propósito de las POM es presentar a los participantes un método equivalente y objetivo para medir su desempeño en un área dada, es posible que la prueba contenga sesgos ideológicos a partir de las creencias del diseñador, por eso es conveniente, y actualmente se realizan de ese modo, que la propuesta sea el producto de un equipo de trabajo. Entre 1917 y 1918 alrededor de 1,7 millones de hombres fueron examinados mediante pruebas estandarizadas diseñadas para seleccionar futuros oficiales o soldados de línea en los EUA. Sin embargo, el test contenía defectos en su diseño, pues partía de una perspectiva racista y la prueba confirmaba las expectativas de los evaluadores blancos sobre el comportamiento de los afroamericanos (Watters, 2015).

Es conocido que un evaluador puede sesgar la medición de diferentes aspectos, a partir de los resultados en uno de ellos (Thorndike, 1920), lo que se conoce como efecto de halo, efecto *Pigmalion* y otros nombres bastante populares; así como que dos docentes corrigiendo un mismo grupo de pruebas obtienen resultados diferentes, ya sea porque aplican otra perspectiva o por favoritismo (Kamenetz, 2015, pág. 50). Mientras que la diferencia de criterio puede reducirse mediante pautas de corrección, que honestamente aplica el equipo de corrección, el sesgo es subjetivo. Hay quienes abogan por la eliminación del docente en el proceso de evaluación, mediante la aplicación de POM. Según el pensamiento de Kelly (1916), enseñanza efectiva significaba resultados uniformes, un pensamiento influido grandemente por la línea de producción promovida por Ford: “apreciando un producto confiable, uniforme y fácilmente replicado (...) por encima del ingenio, la creatividad, la individualidad, la idiosincrasia, el juicio crítico, y la variabilidad” (Davidson, 2011, pág. 105ss). Según Davidson, Kelly no pretendía que las pruebas estandarizadas se expandieran como lo hicieron. En realidad, este tipo de prueba tenía como objetivo medir el pensamiento de orden inferior (“lower order thinking”) en las masas. Las POM “se han convertido en el patrón oro” en los EUA desde las pruebas escolares hasta las que se realizan en escuelas profesionales de posgrado (Davidson, 2011). Es fácil de aplicar y corregir, especialmente indicado donde hay gran cantidad de personas, y su carácter estándar expone a todos a las mismas alternativas.

## **Luces y sombras de las POM**

Las POM son una herramienta válida y confiable para medir competencias en contextos determinados. Es decir, no sirven para todo. La educación básica tiene objetivos diversos, algunos generales y otros individuales. Mientras los objetivos generales intentan promover ciertos aprendizajes concretos, que define cada sociedad desde una visión política, social y económica, por ejemplo; por otra parte, la educación básica intenta promover el potencial de cada alumno, que es único, aunque recurrente en varios o muchos otros estudiantes. En el primer caso, es oportuno utilizar POM para medir el aprovechamiento de un estudiante; en el segundo, sencillamente no tiene sentido.

Cuando pensamos en las carreras profesionales, la formación está directamente relacionada con un perfil de egreso concreto, específico, en el que la última tendencia, desde Bologna, es la identificación de competencias profesionales que se espera adquieran todos los integrantes de una comunidad profesional, del gremio. Hay un “libro blanco” para cada profesión y que atraviesa las universidades de la Unión Europea (UE). Ese mismo tipo de alineación ocurre en el MERCOSUR, cuando acordamos reconocimiento mutuo de estudios. Y esa misma correlación nos lleva a evaluar y acreditar las carreras e instituciones de educación superior que las imparten. Así, entonces, cuando un individuo obtiene su título profesional, se espera

que sea capaz de abordar los temas afines a su carrera con la misma solvencia o al menos en un entorno de competencia aceptado tanto por el colectivo científico como por la sociedad. La sociedad cuenta con la garantía de un producto satisfactorio, el servicio profesional de que se trate, sin importar cuál de ellos lo realice. La evaluación durante la formación debe ser necesariamente idéntica para todos, es decir estándar. Este es un espacio apropiado para aplicar POM.

La numerosidad es el otro ingrediente relevante para tomar decisiones sobre POM. Especialmente, cuando se trata de clasificar un número importante de individuos, y sobre todo, para seleccionar. Este criterio, el de la selección, no es incompatible con la aspiración individual de seguir una carrera determinada y alcanzar la titulación. Esta última discusión está centrada en el modelo en enseñanza y aprendizaje, y en el tipo y calidad de oportunidades de aprendizaje que se ofrecen a los candidatos, no con los niveles de conocimientos y experticia que se requieren para asegurar la competencia profesional de uno y de todos los que ostentan el título.

Es evidente que medir pruebas inéditas, es decir, las que muestran el comportamiento del estudiante a partir de su propia creación, sea respondiendo preguntas directas, discutiendo situaciones problema a los que se le expone, o ensayando el abordaje de temas más o menos extensos, es una actividad agotadora. El desgaste que experimenta un corrector solamente podría atenuarse con un régimen de trabajo que, o bien, extendería el período de corrección más allá de lo aceptable, o bien, demandaría un número de correctores entre los cuales la pauta podría desvirtuarse, aun sin que mediara la voluntad de infringirla. Una POM parece la solución: el trabajo extenso ocurre durante el diseño previo a su aplicación, que por otra parte, afectará a todos los estudiantes por igual.

El lado oscuro de las POM es que hace falta tener claro qué se quiere medir y asegurarse de que el instrumento, o mejor dicho, el conjunto de ítems que lo componen, den cuenta efectivamente del estado y comportamiento que se pretende observar. A mediados del siglo XX un equipo de investigadores en los EUA identificó una serie de objetivos de la educación en todos los niveles, que han sido la base de trabajos de investigación y de aplicación de pruebas desde entonces. Se conocen como taxonomía de Bloom.

## **Evaluación y pruebas**

El proyecto de Benjamín Bloom y su equipo investigador, dedicado a conocer qué se evalúa en las pruebas del sistema educativo, puso de manifiesto los objetivos de la educación desde primaria hasta la universidad (K-12+College), sin importar qué tipo de contenidos, qué asignaturas, qué docentes (Bloom B. S., 1956)<sup>3</sup>.

Bloom, junto a un equipo de académicos y científicos en todo EUA, sintetizó las demandas de desempeño, que subyacen en las pruebas de prácticamente todas las asignaturas, desde infantil hasta el *college*. Este conjunto de habilidades y destrezas<sup>4</sup> - “abilities and skills” en sus términos (pág. 38ss) - es el que se conoce como taxonomía de Bloom en el dominio cognitivo. Los autores plantean que aunque los objetivos, y los procedimientos y materias de prueba, se desagreguen al extremo, siempre es posible reducir un número relativamente limitado de tipos de conducta del alumno (Bloom, Engelhart, Hail, Furst, & Krathwohl, 1971, pág. 24). Es

---

<sup>3</sup> Existe versión en español de 1971 editada por Marfil, Alcoy (Valencia).

<sup>4</sup> La discusión sobre el uso de estos términos u otros es incesante pues la definición de uno refiere al otro u otros por igual: habilidad, destreza, talento, capacidad, aptitud, inteligencia y así sucesivamente. En esta sección utilizamos destrezas que parece más abarcador.

decir, en todas las propuestas de evaluación se requiere utilizar las destrezas de orden inferior, según la taxonomía: recordar, comprender, aplicar, o sus sinónimos y afines; y, las destrezas de orden superior: analizar, sintetizar, evaluar, o sus sinónimos y afines. Churches (2009) propone para el siglo XXI: analizar, evaluar, crear. (Picón Martínez, 2013, págs. 113-114).

La secuencia creciente de estas habilidades y destrezas explica la lógica de las pruebas de evaluación que sugiere ir avanzando de lo más fácil a lo más difícil. Una simplificación mal entendida: es más fácil recordar un nombre que una lista de nombres, pero en ambos casos se utiliza la habilidad de recordar, nada más. Podemos decir que un estudiante tiene más o menos memoria que otro, pero no podemos decir nada de las otras destrezas que lo convierten en un individuo autónomo para hacer. Otra cosa es proponer una prueba donde recordar aparezca primero que analizar, o que cuando un ítem requiere varias destrezas sea de orden inferior, superior o de ambas, está ubicado en una posición donde los que le preceden no requieren más destrezas y los que le siguen no requieren menos. Pero el final, la prueba completa, para el caso de la educación superior, las involucra todas: memoria, comprensión, aplicación, análisis, síntesis, evaluación.

El origen de las pruebas de evaluación de nuestros días ha sido la necesidad de sustituir el mecanismo clásico que podríamos llamar tutoría. En el modelo formativo de la antigüedad clásica, el sabio toma a su cargo a un pupilo que considera prometedor y que lo acompaña, discute, reflexiona y practica, hasta que lo acredita y lo envía por su propio camino. Discípulo del maestro es su credencial. Las necesidades de la sociedad y el interés por aprender han reducido esa práctica, que no ha desaparecido, por supuesto. Pero la educación, en general, y la educación profesional, en particular, no cuentan con los recursos para proveer al cien por ciento de los candidatos con ese modelo de aprendizaje y acreditación. Es muy costoso y demanda mucho tiempo: piénsese en un cirujano especializado en un tipo de intervención. ¿Cuántas tiene que realizar para acreditarse? ¿Cuál es el costo de cada una?

Además, la propia universidad ha desarrollado investigación y ha reconocido que existen modelos más efectivos: le permiten formar, utilizan los recursos disponibles sin desperdicio, y producen profesionales competentes. Estos modelos son adecuados para algunas carreras y para alguna parte de todas las carreras. De la antigua entrevista personal con el tribunal, bolillero mediante, hemos llegado a las POM.

Construir una POM no es difícil. Lo que verdaderamente lleva tiempo, análisis y discusión – e investigación – es andar el camino inverso desde el perfil de egreso y las competencias profesionales hasta identificar cuáles son los comportamientos que dan cuenta de un nivel de adquisición adecuado y más atrás, cuáles son los contenidos que podrán generar esos comportamientos.

## **Lenguaje superficial y lenguaje profundo de las POM**

Una evaluación basada en POM incluye una cantidad de ítems con una estructura esencialmente similar. Un enunciado que se completa con dos o más opciones alternativas que pueden o no ser excluyentes. El caso mínimo es una propuesta dicotómica con opciones: verdadero y falso. Los otros incluyen de tres a cinco, una es la respuesta y las demás son distractores. Los distractores deben tener apariencia de ser respuestas posibles, de lo contrario no cumplen su función, como se muestra en la figura 1. La prueba general debe estar construida de manera que pueda descartarse el azar para alcanzar el nivel mínimo de suficiencia.

Ítem bien construido	Ítem mal construido
El altímetro indica la distancia vertical desde el instrumento hasta el a. piso. <b>b. nivel de presión 1013.2 hPa.</b> c. nivel medio del mar.	El altímetro indica la distancia vertical desde el instrumento hasta el a. nivel de la presión en la superficie. b. nivel de presión 1013.2 hPa. c. nivel de la presión a nivel del mar.

Figura 1. Ejemplo de POM

El ejemplo de ítem mal construido contiene los dos errores más evidentes: i) texto innecesario que ocupa tiempo excesivo sin aportar evidencia para el evaluador; y, ii) la posibilidad de deducir la respuesta. Efectivamente, en este último caso el estudiante que tiene el conocimiento básico de que el altímetro funciona midiendo la presión circundante puede deducir que un instrumento instalado en una aeronave no puede medir el valor de la presión atmosférica más allá de su entorno, y por lo tanto descartará las opciones a. y c.; pero no puede concluirse que conoce las implicancias de la respuesta, opción b. como sí se puede concluir del ítem correcto. Como se ve en ese ejemplo, todo el texto que puede aplicarse a las tres opciones se incluye en el enunciado, lo cual sugiere que todos los ítems deberían construirse reduciendo la extensión de las opciones a las mínimas palabras posibles, preferentemente una (Rodríguez Ayán, 2014, págs. 3-4).

El azar se puede amortiguar mediante una distribución equilibrada de respuestas; para tres opciones sería conveniente que haya un tercio de respuestas en cada una: un tercio a., un tercio b., un tercio c. Algunas pruebas, que ponen el nivel de suficiencia igual o superior al 50% de respuestas correctas, descartan la cantidad de respuestas correctas que equivalen a un tercio del total de POM, de manera que la aprobación requiere un 66,66% de éxito. Existen muchos otros procedimientos para medir el resultado de una prueba con POM y contrarrestar el azar, por ejemplo, penalizando los errores o aportando puntos por no responder al azar, es decir, por dejar el ítem “en blanco” (íd).

Estos elementos más o menos evidentes pueden considerarse el lenguaje superficial de las POM, pero: qué significan las respuestas, conduce a lo que denomino lenguaje profundo. En este sentido, voy a limitar el desglose a: i) la asignación de puntos o determinación del nivel de suficiencia; ii) el significado de algunos errores; y iii) la evidencia combinada de POM.

En general las pruebas con POM se miden linealmente: todas los ítems tiene el mismo valor en relación a la calificación y un número  $n$  de respuestas correctas asegura el nivel de suficiencia. Se pueden hacer ajustes por errores, bonificaciones por evitar el azar o contabilizaciones como el caso de los dos tercios mencionadas más arriba. Cuando reflexionamos sobre esta asignación lineal y la relacionamos con las destrezas de nivel inferior y superior de la taxonomía de Bloom, nos damos cuenta de que en los hechos es poco probable que todas las POM tengan el mismo nivel de dificultad. Por lo tanto, que se les asigne el mismo valor no da verdadera cuenta del comportamiento que se pretende medir. Especialmente, porque los recursos y la intensidad con que los aplica el estudiante en cada caso son diferentes. No es lo mismo tener memoria que tener criterio para evaluar una situación y aportar una respuesta que requiere considerar varios datos.

Una prueba con POM puede ser una oportunidad para evaluar al mismo tiempo el modelo de enseñanza. En general, algunos contenidos pueden ofrecer dificultades particulares a los estudiantes o que el prejuicio supere a la ciencia. Una opción que contemple el tipo de comportamiento erróneo, puede dar cuenta del impacto del curso para evitarlo; y la persistencia del error, sugerir la necesidad de un abordaje diferente en la siguiente edición.

Aquí no se pone en tela de juicio la calidad del curso sino que se encuentra evidencia de la distancia entre el alumno para el que fue diseñado y el alumno que efectivamente se sentó en el salón de clase.

Existen pruebas psicotécnicas en las que se aplican POM. Estas evaluaciones tienen en cuenta la combinación de respuestas a varios ítems o una misma POM es utilizada en combinaciones diferentes para anticipar comportamientos en diferentes áreas. Este tipo de POM no tiene, por así decirlo, respuestas correctas o incorrectas y en ese sentido no satisfacen los criterios de POM para medir el resultado de un curso curricular, por ejemplo. No obstante, veremos...

Una prueba clásica en el mundo es PISA que se aplica a niños de 15 años incluidos en el sistema educativo. La población seleccionada lo es por su accesibilidad. De hecho no importa en qué curso está matriculado el estudiante (ANEP, 2010; 2014). En realidad mide la calidad del capital humano de los países participantes, si consideramos que desde los 16 los niños pueden tener comportamientos adultos legalmente garantizados: trabajar, casarse, incluso votar o ir a la guerra. Esta prueba está construida con POM y asociadas por secciones. Los resultados no son lineales, sino que los comportamientos se ponderan por el tipo de error y por las combinaciones de respuestas que satisfacen los constructos elaborados por los equipos de investigadores y contratistas que la diseñan. Toda la prueba tiene un marco teórico ad hoc, a partir de cuyas definiciones se elaboran las propuestas y los ítems asociados. De ese modo, un caso o situación problema tiene varias POM en las que se miden diferentes variables. De esos resultados no se sacan conclusiones, solamente se aportan los valores, que cada cual interpreta a su modo. Lo más directo es una lista de precedencia por el número. La hoja que recoge las respuestas podría ser similar a las que utiliza la FCEA y sus UC, pero la disposición de la propuesta, la letra, deja en claro cómo se asocian las preguntas y que evidentemente tienen diferente peso específico.

Una prueba modélica sea con o sin POM debería aportar evidencia de que se han alcanzado los objetivos educativos propuestos y de que quien la aprueba ha desarrollado la parte del perfil de egreso esperada para esa porción de la formación. Dicho en estos términos, el diseño de una prueba implica más que redactar un caso o una batería de preguntas, debería incluir un contraste entre los elementos señalados que es, en general, la parte de investigación educativa que toca a quienes orientan la enseñanza superior, los profesores universitarios.

## **Impacto de las pruebas con POM en los cursos de la FCEA**

En 2014 la Unidad de Apoyo a la Enseñanza (UAE) incorporó tecnología para escanear formularios de recolección de respuestas de POM y un programa propietario para la lectura de los datos. Desde entonces ha colaborado con las UC para diseñar formularios apropiados, según las necesidades de cada una, y el procesamiento luego de las pruebas. La UAE produce un archivo Excel con los datos recogidos, según las orientaciones de cada UC; cada línea corresponde a un formulario, cada uno a un estudiante. La UAE no participa en la elaboración de las propuestas, pero ofrece asesoramiento para la tarea. También se ofrecen cursos en la Red de UAE, generales. La especificidad de cada carrera sugiere que el mejor enfoque es el trabajo directo de la UAE con las UC de la FCEA. Estos recursos están asociados con los demás a disposición, como el Espacio Virtual de Aprendizajes (EVA), los audiovisuales didácticos o *polimedia*, los talleres y cursos para el desarrollo profesional docente, y que en conjunto contribuyen a mejorar la calidad de la enseñanza en contextos de numerosidad, según se entiende este concepto en la Universidad de la República.

Desde 2014 se han adherido a este modelo de evaluación 26 UC que en conjunto han procesado poco más de 108000 pruebas con POM. La FCEA ofrece anualmente unos 180

cursos, entre obligatorios, optativos y electivos, que corresponden a otras tantas UC. Se espera que al menos 50% de los que cuentan con más de 100 estudiantes matriculados opten por este modelo de pruebas en los próximos dos años y que se generalice para la mayoría en otros dos.

## Comentario final

Las pruebas con POM permiten obtener evidencia de aprendizajes con la misma certeza que otros instrumentos tradicionales como los documentos inéditos, escritos y similares. La probabilidad de aciertos por azar puede reducirse hasta hacerla no significativa. Su aplicación permite abarcar grupos numerosos donde la corrección tradicional podría estar sesgada, de manera que ofrece garantías de igual tratamiento para todos los estudiantes. Los recursos para procesar los resultados aumentan la velocidad para obtener información relevante para todos los actores. La elaboración de estas pruebas demanda cierta formación y no menos análisis que otros modelos de evaluación. A fin de cuentas, lo que se busca es evidencia de que se ha producido la transformación y desarrollo profesional propuestos al comenzar el curso.

## Bibliografía

- ANEP. (2010). *Uruguay en PISA 2009*. Montevideo: ANEP-.
- ANEP. (2014). *Uruguay en PISA 2012*. Montevideo: Programa ANEP-PISA.
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. Philadelphia: David McKay Company.
- Bloom, B., Engelhart, M., Haill, W., Furst, E., & Krathwohl, D. (1971). *Taxonomía de los objetivos de la educación. Clasificación de las Metas Educativas. Tomo 1, ámbito del conocimiento*. Alcoy, Valencia: Marfil.
- Churches, A. (2009). *Bloom's Digital Taxonomy*. En línea: <http://edorigami.wikispaces.com/>. Recuperado el 24 de junio de 2012, de <http://edorigami.wikispaces.com/>
- Davidson, C. (2011). *Now You See It: How the Brain Science of Attention Will Transform the Way We Live, Work, and Learn*. New York: Viking.
- Davidson, C. (23 de 09 de 2011). *Standardized tests for everyone? In the Internet age, that's the wrong answer*. Recuperado el 6 de setiembre de 2016, de [www.washingtonpost.com](http://www.washingtonpost.com): [https://www.washingtonpost.com/opinions/standardized-tests-for-everyone-in-the-internet-age-thats-the-wrong-answer/2011/09/21/gIQA7SZwqK\\_story.html?utm\\_term=.21035ad6888e](https://www.washingtonpost.com/opinions/standardized-tests-for-everyone-in-the-internet-age-thats-the-wrong-answer/2011/09/21/gIQA7SZwqK_story.html?utm_term=.21035ad6888e)
- Davidson, C. (02 de 09 de 2011). *Where Did Standardized Testing Come From Anyway?* Recuperado el 06 de setiembre de 2016, de [www.hastac.org](http://www.hastac.org): <https://www.hastac.org/blogs/cathy-davidson/2011/09/02/where-did-standardized-testing-come-anyway>
- Kamenetz, A. (2015). *The Test: Why our schools are obsessed with standardized testing - but you don't have to be*. Philadelphia: PublicAffairs.
- Kelly, F. J. (feb de 1916). The Kansas Silent Reading Tests. *Journal of Educational Psychology*, 8(2), s/d.
- Picón Martínez, A. (2013). *Perspectivas de la implantación de modelos de evaluación de aprendizajes desde la gestión institucional*. Santiago de Compostela, A Coruña, España: Departamento de Publicaciones - USC. Recuperado el 29 de enero de 2013, de <http://dspace.usc.es/handle/10347/7241>
- Rodríguez Ayán, M. N. (2014). *Diseño y análisis de pruebas de múltiple opción*. Montevideo: UDELAR-FQ.
- Thorndike, E. (1903). *Educational Psychology*. New York: The Science Press.
- Thorndike, E. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4(1), 25-29.

Watters, A. (27 de 01 de 2015). *Multiple Choice and Testing Machines: A History*.  
Recuperado el 6 de set de 2016, de <http://hackededucation.com>:  
<http://hackededucation.com/2015/01/27/multiple-choice-testing-machines>