

El rol del azar en las pruebas de opción múltiple

The role of chance in multiple choice tests

César Medina^{1*} y Sandra Velazco¹

¹Facultad de Ciencias Exactas y Tecnología, Universidad Nacional de Tucumán, Av. Independencia 1800, CP 4000, San Miguel de Tucumán, Tucumán. Argentina.

*E-mail: ces.medina@yandex.com

Recibido el 28 de diciembre de 2020 | Aceptado el 3 de mayo de 2021

Resumen

Se presenta una síntesis sobre algunas cuestiones importantes de la evaluación como una etapa fundamental del proceso educativo, destacando su rol instrumental que admite distintas finalidades y modalidades. Se discuten algunos aspectos metodológicos relacionados con el diseño y la elaboración de las pruebas de opción múltiple, poniendo el énfasis en aquellos relacionados con la necesidad de reducir la probabilidad de que los alumnos aprueben por azar. Al respecto, se discuten con detalle matemático dos estrategias: a) la puntuación negativa de las respuestas incorrectas y b) el diseño de las pruebas considerando tres variables críticas: la cantidad de ítems, la cantidad de opciones y el grado de exigencia. Se destaca que un incremento en la cantidad de ítems, por sí solo, no solo no reduce la probabilidad de acertar por azar, sino que puede aumentarla si las otras dos variables no están adecuadamente diseñadas.

Palabras clave: Evaluación; Opción múltiple; Aprobar por Azar; Distribución Binomial.

Abstract

A summary is presented about some important issues of evaluation as a fundamental stage of the educational process, highlighting its instrumental role which fits different purposes and versions. Some methodological aspects are discussed related to design and development of multiple-choice tests, emphasizing those connected with the need to reduce the probability of students passing by chance. In this respect, two strategies are discussed with mathematical detail: a) Negative scoring for wrong answers and b) Test design considering three crucial variables: the number of items, the number of options and the degree of demand. It is highlighted that increasing the number of items solely, not only does not reduce the probability of passing by chance, but even may raise it if the other two variables are not properly designed.

Keywords: Evaluation; Multiple Choice; Pass by Chance; Binomial Distribution.

I. INTRODUCCIÓN

La educación, como todas las áreas del conocimiento y la actividad humana, siempre se ha servido de los adelantos tecnológicos de cada época; pero nunca ha experimentado cambios metodológicos tan radicales como los ocurridos en los últimos veinticinco años, gracias al desarrollo de las Tecnologías de la Información y la Comunicación (TIC).

Actualmente, la mayoría de los centros educativos de todo el mundo, y en particular las universidades, ofrecen Entornos Virtuales de Aprendizaje (EVA), es decir, plataformas digitales que permiten la implementación de todo tipo de cursos (Belloch, s/f; López Meneses y Miranda Velasco, 2007; López Rayón Parra, Ledesma Saucedo y Escalera Escajeda, 2009; Onrubia, 2005).

Tales plataformas suelen estar administradas por un plantel técnico mediante un Sistema de Gestión de Aprendizaje (SGA)¹, los cuales cuentan con una gran cantidad de recursos, tales como bases de datos y otros tipos de almacenamiento, espacios de consulta y ayuda, acceso diferenciado para estudiantes y docentes, diversas posibilidades de comunicación sincrónica y asincrónica, distintas modalidades de evaluación, y registro y seguimiento de los estudiantes, entre otras herramientas (Entorno virtual de aprendizaje, en Wikipedia, 2020).

Desde principios de 2020, en los países afectados por la pandemia de covid-19 se establecieron medidas de aislamiento social que imposibilitaron todo tipo de educación presencial, y el uso de los EVA creció de manera exponencial. Muchos docentes que hasta entonces habían utilizado estos sistemas solo de manera ocasional se vieron en la urgente necesidad de adquirir conocimientos técnicos y didácticos para adaptar sus cursos a la modalidad virtual. En un cambio tan profundo, complejo y repentino, cabe esperar confusiones conceptuales y desaciertos metodológicos; muchos de los cuales pueden ser sutiles y pasar inadvertidos para la mayoría de los actores del proceso educativo, pero repercuten en la calidad y eficacia del mismo. En este trabajo se presentan propuestas tendientes a la superación de este tipo de dificultades.

La sección II presenta una breve síntesis sobre diversas cuestiones importantes de la evaluación, destacando las principales modalidades que se puede implementar eficientemente mediante los SGA. Esta síntesis está planteada como marco contextual para la sección III, donde se discuten aspectos metodológicos sobre el diseño y elaboración de las pruebas de opción múltiple, enfocados en la necesidad de disminuir la probabilidad de que los alumnos acierten respuestas por azar. Con este objetivo se analizan dos estrategias matemáticas y se discute su aplicabilidad.

II. LA EVALUACIÓN: RESUMEN CONTEXTUAL

A. Finalidad, alcance y relevancia de la evaluación

Diversas propuestas educativas desarrolladas en las últimas décadas presentan la evaluación como un instrumento que admite distintas finalidades. En primera instancia, la evaluación es una herramienta de diagnóstico y seguimiento continuo del progreso de los estudiantes y el proceso educativo. Permite detectar a tiempo debilidades y fortalezas de los modos de aprender y enseñar, posibilitando al docente la toma de decisiones oportunas para brindar apoyo a los estudiantes, plantear nuevas estrategias y realizar las modificaciones y ajustes necesarios en la metodología, el programa y los objetivos del curso (Cano, 2008; Millán Núñez-Cortés, 2010; Ruiz Betancourt, Rodríguez Guaraca, Gallegos Arias y Villacis Vallejo, 2018).

Desde una perspectiva más tradicional e institucional, la evaluación es un proceso mediante el cual el docente reconoce, valora y califica el desempeño de un estudiante, su adquisición y construcción de conocimientos, y su desarrollo de capacidades y competencias (Cano, 2008; Centro de Estudios Científicos y Tecnológicos N° 13, 2010; Stufflebeam y Shinkfield, 1987). La calificación, como dictamen, decide si el estudiante puede avanzar en su plan de estudios o debe realizar nuevamente la evaluación, o una tarea o actividad, o todo el curso. En ciertos casos, el dictamen puede significar la continuidad o el abandono de la carrera; y en la evaluación final de esta, el dictamen certifica al evaluado como un nuevo profesional idóneo en su incumbencia, con toda la responsabilidad social que esto implica tanto para el evaluado como para el evaluador. En otros casos, dependiendo del área de conocimiento y currículo de cada universidad, la titulación no es tan decisiva en la carrera como puede serlo la especialización o el posgrado, y el acceso a estas instancias también suele decidirse mediante una evaluación (Aspa Marco y Rodríguez de Castro, 2010; Millán Núñez-Cortés, 2010).

Estas razones manifiestan el carácter crucial de la evaluación, como instrumento didáctico y como base de dictamen, para todos los actores del proceso educativo y para la sociedad donde habrán de desempeñarse los profesionales. Por ello es muy importante el desarrollo de criterios de control del diseño, elaboración y aplicación de la evaluación tendientes a mejorar progresivamente su calidad, en términos de objetividad, validez y confiabilidad (estos conceptos se discuten brevemente en la subsección D).

B. Modalidades de evaluación

Los SGA ofrecen diversas modalidades de evaluación que pueden implementarse de un modo simple y versátil. Entre estas modalidades cabe destacar dos muy utilizadas: las pruebas de desarrollo, y las pruebas objetivas.

Las primeras constituyen una modalidad bastante tradicional y su versión digital no presenta mayores cambios respecto de la convencional de enunciados impresos. Consisten en cuestionarios, actividades o tareas definidas de un modo relativamente amplio, en los cuales el estudiante puede desarrollar un discurso más o menos elaborado, con

¹ Referidos en algunos contextos por su sigla en inglés, LMS (*Learn Management System*).

discusiones, justificaciones, argumentaciones, mediciones, gráficas, cálculos, etc. Estas pruebas pueden presentarse en forma sincrónica o asincrónica, en los plazos fijados por el docente, y suelen usarse para calificar al alumno al final de un curso o etapa del mismo. Permiten valorar destrezas cognitivas como la capacidad de organizar, crear, discurrir, sintetizar, etc. En algunos casos, su calificación se presta a cierta subjetividad del docente, según sea que los criterios valorativos y el estilo discursivo del evaluado concuerden, o no, con sus preferencias.

C. Pruebas objetivas

En contraposición a las pruebas de desarrollo, se encuentran las pruebas objetivas, cuyas versiones digitales ofrecen considerables ventajas respecto de los enunciados impresos. En estas, el alumno no desarrolla ningún discurso ni presenta materiales de su elaboración. Suelen tener el formato de un cuestionario con preguntas breves y directas, y se espera que el alumno las responda de forma categórica, con poca o ninguna justificación o argumentación. Según su modalidad, se puede ofrecer al alumno una lista de respuestas a la pregunta formulada, para que él elija la correcta o la más pertinente (pruebas de opción múltiple), o bien se le puede plantear el ordenamiento de un conjunto de elementos, en base a una dada consigna, o la asociación de conceptos listados en dos o más columnas, o la identificación de un elemento en una gráfica, o el llenado de espacios vacíos en un texto, etc.

Contrariamente a lo que muchos suponen, estas pruebas no solo sirven para evaluar el aprendizaje memorístico. Si están bien diseñadas, pueden exigir al estudiante un grado muy alto de atención, reflexión y razonamiento; y son útiles para valorar distintas destrezas cognitivas tales como la comprensión, el análisis, la capacidad de plantear y resolver problemas específicos, de identificar cuestiones críticas, de tomar decisiones, etc. (Educrea, s/f).

A los fines de asignar una calificación, estos cuestionarios suelen ser más extensos, se administran en forma sincrónica, en un horario limitado para su resolución y se admite un solo intento de presentación. Si, por el contrario, se usan para diagnóstico, seguimiento o autoevaluación, suelen ser más reducidos, administrarse en forma asincrónica, habilitarse durante lapsos más prolongados, y se admite un número mayor de intentos. En cualquier caso, la realimentación y la comunicación de los resultados puede hacerse en forma inmediata o diferida.

Los SGA guardan automáticamente las pruebas objetivas en sus bases de datos, permitiendo un rápido acceso a cada uno de sus ítems (preguntas, gráficas, etc.), lo cual facilita la tarea de perfeccionarlas y adaptarlas a nuevos cursos y situaciones. Estas pruebas son especialmente convenientes para evaluar a grupos numerosos porque los resultados son calculados por los SGA en forma inmediata, mediante un algoritmo digital basado en las puntuaciones previamente definidas por el evaluador. Esto representa un importante ahorro de tiempo y trabajo, al evitar la tarea de lectura y corrección. Por contrapartida, estas pruebas exigen, en su diseño y elaboración, mucho más trabajo, reflexión y cuidado que las pruebas de desarrollo.

Las pruebas objetivas son muy utilizadas en centros educativos de todo el mundo, en particular en universidades y en instancias institucionales donde se desea evaluar a grupos numerosos y heterogéneos, por ejemplo, aspirantes al ingreso o a becas, posgrados, especialidades, etc. No solo gozan de creciente aceptación entre los evaluadores, sino también entre los evaluados, quienes las consideran, en general, un instrumento que favorece la inclusión social y permite un alto grado de objetividad y equidad (Millán Núñez-Cortés, 2010; Noriega Echevarría, 2010).

D. Calidad de las pruebas objetivas

La calidad de las pruebas objetivas se estima técnicamente en términos de tres condiciones básicas que deben satisfacer: objetividad, validez y confiabilidad.

La objetividad se satisface en la medida en que el instrumento se ajuste al objeto que se desea evaluar, y en la medida en que exista un consenso intersubjetivo de los especialistas sobre dicho objeto que, en términos genéricos, será algún tipo de aprendizaje bien definido. Es imposible alcanzar una objetividad absoluta, pero se puede aspirar a un alto grado de objetividad si se pone especial esmero en despojarse de preconceptos y prejuicios, y en aproximarse al objeto siguiendo pautas y criterios racionales y científicos (Tristán López y Pedraza Corpus, 2017).

La validez, por su parte, se satisface en la medida en que el instrumento cumple apropiadamente con su finalidad, es decir, cuando sus resultados reflejan una medida de lo que realmente se busca evaluar. Por último, la confiabilidad se satisface en la medida en que el instrumento evaluativo genera resultados reproducibles, de manera relativamente estable, cuando se aplica en condiciones comparables entre sí. Algunos autores definen y discuten distintas clases de validez y confiabilidad, pero ese grado de detalle excede el marco de este resumen contextual (Macías Calvillo, 2011; Mousalli-Kayat, 2017; Tristán López y Pedraza Corpus, 2017).

E. Recomendaciones sobre el diseño y elaboración de una prueba de opción múltiple

Entre las pruebas objetivas, la versión más usada es la de opción múltiple (también llamada de *elección* o *selección múltiple*, *multirrespuesta*, o por su denominación en inglés, *multiple choice*). Estas pruebas consisten en una serie de ítems (*reactivos*) independientes que pueden estar precedidos de un encabezamiento o comentario preliminar.

Cada ítem consta de un enunciado (*tallo* o *base*), generalmente planteado en forma de pregunta, seguido de una lista de respuestas posibles (*alternativas*, *opciones*), de las cuales, por lo general, una sola es correcta. Los SGA aclaran automáticamente al evaluado, en cada ítem, si en la lista hay una sola respuesta correcta, o más de una. Las opciones incorrectas se llaman *distractores*.

Existen recomendaciones bastante consensuadas por especialistas, técnicos y docentes en general sobre el diseño y elaboración de estas pruebas (García-Garro, Ramos-Ortega, Díaz de León-Ponce y Olvera-Chávez, 2007; Macías Calvillo, 2011; Mousalli-Kayat, 2017; Ruiz Betancourt et al., 2018; Tristán López y Pedraza Corpus, 2017; Prueba objetiva, s/f; Vázquez, Murillo-Cabezas, Gómez, Martín, Chaves y Peinado, 2008). A continuación, se presenta una lista que, aunque no es exhaustiva, puede dar lineamientos generales útiles. Nuestras observaciones están entre corchetes

- No hacer preguntas capciosas. [El objetivo no es confundir al alumno, ni con el enunciado ni con los distractores, estos simplemente están destinados a evaluar su comprensión o su capacidad de discriminar entre la opción correcta y las incorrectas.]

- No hacer preguntas en forma negativa. [Esto resulta confuso y ambiguo.]

- No hacer preguntas triviales. [Las preguntas deben orientarse hacia los objetivos de la evaluación.]

- No hacer preguntas basadas en opiniones ni apreciaciones subjetivas.

- Cada enunciado debe tener sentido completo, independiente de las alternativas. [Excepto, claro está, cuando se pide completar una frase o un espacio vacío.]

- Mantener la misma coherencia gramatical entre el enunciado y cada una de las opciones.

- Tanto el enunciado como las alternativas deben redactarse de forma clara y concisa, evitando verbosidad innecesaria, información irrelevante o párrafos demasiado extensos.

- Redactar todas las alternativas con una extensión y una precisión relativamente uniforme.

- Las alternativas deben ser concisas y taxativas. [Consideramos esto imprescindible para no confundir al alumno. Si la respuesta correcta es solo una, no puede haber “*términos medios*”, ni respuestas “*parcialmente correctas*”.]

- Evitar dar indicios sobre cuál es la respuesta correcta o cuál es incorrecta. [Tanto en el contenido como en la redacción del enunciado y las alternativas se debe evitar dar pistas. En particular, conviene revisar todas las alternativas, tanto la correcta como las incorrectas, incluyendo contenido, redacción, signos de puntuación, ortografía, etc. Los docentes suelen revisar mejor las alternativas correctas, y una alternativa con errores, es decir, mal revisada, sugiere al alumno que probablemente es incorrecta.]

- Ordenar aleatoriamente las opciones. [La respuesta correcta debe ubicarse al azar entre las incorrectas. En este sentido, cabe enfatizar que no basta la percepción del docente sobre un ordenamiento al azar. Casi todas las personas tienen una percepción “compensatoria” del azar y, por eso, la mayoría de los docentes tiende a poner la opción correcta en medio de las incorrectas, casi nunca al principio ni al final; o bien siguen patrones, por ejemplo, tratan de que la respuesta correcta no esté en el mismo lugar en dos ítems consecutivos. El verdadero azar no es compensatorio ni sigue patrones, y lo que es peor, el alumno también tiene una percepción compensatoria y también puede seguir y descubrir patrones. Por ello, es recomendable usar la opción de ordenamiento aleatorio del SGA.]

- Usar con precaución la opción “*Ninguna de las opciones anteriores es correcta*”. [Esta opción no debe usarse nunca como “relleno” para simplemente aumentar la cantidad de opciones. Además, si se usa el ordenamiento aleatorio del SGA, debe redactarse de otra manera, porque puede aparecer en cualquier lugar de la lista y la palabra “anteriores” cambiaría de sentido.]

- Los ítems deben ser independientes. [No puede la respuesta de un ítem depender de la de otro, ni puede un dado ítem dar pistas sobre la respuesta correcta o incorrecta de otro.]

- Agrupar las preguntas según su tipo, contenido, orden lógico o cronológico, o grado de dificultad creciente.

- Todos los distractores deben ser plausibles.

- Cada ítem debe medir un solo resultado del aprendizaje.

III. DISCUSIÓN DE ALGUNOS ERRORES METODOLÓGICOS DE LAS PRUEBAS DE OPCIÓN MÚLTIPLE

A. Planteo general y comparativo

Los errores metodológicos de diseño y elaboración de una prueba de opción múltiple menoscaban su validez y confiabilidad. En la versión más corriente de este tipo de pruebas, el alumno no debe ni puede justificar sus respuestas, lo cual impide al evaluador conocer por qué razón eligió una dada alternativa, ya sea correcta o incorrecta. Luego, es dable encontrar dos sesgos que se debe evitar: (i) que el alumno que sabe elija una respuesta incorrecta; (ii) que el alumno que no sabe elija una respuesta correcta.

El primer caso ocurre por confusión del alumno, y suele suceder cuando el docente elabora la prueba sin respetar las recomendaciones de la sección anterior. El segundo caso ocurre cuando el estudiante “adivina” la respuesta.

Ante una lista de alternativas, un alumno que carece de elementos de juicio para decidir cuál es la correcta, elegirá, en general, aquella que le “suene” más plausible. Esta elección puede tomarle un tiempo variable, pero suele ser el desenlace de un proceso cognitivo bastante complejo, durante el cual el alumno busca y procesa pistas e información, intenta recordar lo estudiado o visto en clases, y razona con lógica o intuición. Sin embargo, la falta de conocimiento, el estrés del examen y el apremio del tiempo tornan este proceso vacilante y errático, y la decisión final, más que una “cosa razonada”, termina siendo una “corazonada” donde el azar juega un rol decisivo.

El efecto del azar en las pruebas de opción múltiple ha sido objeto de diversas investigaciones. Cabe destacar, como referentes importantes, una serie de modelos estadísticos de considerable complejidad desarrollados a mediados del siglo XX, por ejemplo, el de Chernoff (1962), que creemos representativo de los de esa época, por su detalle y amplitud, así como por recoger los aportes y críticas de varios estudios anteriores.

El principal mérito de estos estudios, en nuestra opinión, es el de haber establecido muchos principios teóricos fundamentales y haber relevado las principales variables críticas que conciernen a la calidad de las evaluaciones. Sin embargo, su complejidad matemática los torna poco accesibles a los evaluadores no especializados en estadística, y más allá de esto, su aplicación práctica plantea serias dificultades que han sido señaladas, en algunos casos, por sus propios autores. Entre estas, cabe mencionar las siguientes:

- I) En general, los modelos incluyen varias hipótesis que pueden no cumplirse en muchos casos particulares.
- II) Requieren del cálculo previo de estimadores y parámetros que solo pueden obtenerse desarrollando una o más pruebas piloto. En algunos casos, incluso son necesarias otras tareas preliminares, en particular cuando las poblaciones a evaluar no son homogéneas.
- III) Los parámetros y estimadores están sujetos a una considerable variabilidad de una prueba a otra, sobre todo en poblaciones con algún tipo de sesgo.
- IV) En algunos modelos, los algoritmos se tornan extremadamente complicados a medida que aumenta la cantidad de ítems de la prueba, aun para cantidades típicas o moderadas, lo cual acentúa la dificultad señalada en (III), pues las variables estadísticas son confiables en la medida en que sean calculadas a partir de muestras grandes.

En épocas más recientes, la investigación sobre las pruebas de opción múltiple se ha enfocado más en los aspectos didácticos que en los matemáticos. En nuestra búsqueda bibliográfica de estos últimos, encontramos escasos trabajos, y la mayoría se limita a reafirmar descriptiva o cualitativamente los modelos tradicionales. Cabe mencionar, como excepción, el estudio de Shakil (2009), que ofrece una mayor versatilidad en el diseño de las evaluaciones, basándose en la distribución beta-binomial, bastante más compleja que la convencional distribución binomial.

En términos comparativos, la propuesta de nuestro trabajo es ostensiblemente más simple que las de los modelos citados. Las hipótesis de nuestro estudio son pocas y muy poco restrictivas, y los cálculos que demanda son accesibles a cualquier profesional de ciencias exactas, e incluso a cualquier persona que tenga cierto conocimiento de programas de cómputo con aplicaciones estadísticas. Estos cálculos permiten diseñar las evaluaciones a partir de unas pocas variables críticas y reducir el efecto del azar dentro de márgenes muy amplios, sin recargar la tarea del evaluador ni de los evaluados.

Los métodos que analizamos con esta finalidad son dos: el de la puntuación negativa de las respuestas incorrectas (Prueba objetiva, s/f) y el diseño de las pruebas considerando tres variables críticas: la cantidad de ítems, la cantidad de opciones y el grado de exigencia.

B. La puntuación negativa de las respuestas incorrectas

Para que este método sea preciso y justo, la puntuación negativa debe contrarrestar exactamente la puntuación atribuible al azar. Consideremos un ejemplo:

Ejemplo 1: Se diseña una prueba de opción múltiple que consta de 12 preguntas, en cada una de las cuales hay 3 alternativas y la respuesta correcta acredita un punto. Supongamos, por simplicidad, el caso extremo de un alumno que se presenta a rendir sin ningún conocimiento del temario y contesta totalmente al azar.

La teoría de probabilidades nos indica que, en esas condiciones, ese alumno acertaría, en promedio, una de cada tres respuestas, y su calificación final sería, probablemente, 4 en una escala de 1 a 12. Ahora bien, para contrarrestar exactamente esos 4 puntos inmerecidos, todas las opciones contestadas incorrectamente deberían restar 4 puntos.

Si el alumno ha elegido 4 respuestas correctas (por mero azar), entonces ha elegido 8 respuestas incorrectas que deben restar 4 puntos. Luego, la puntuación de cada respuesta incorrecta, *puntuación* (r_i), debe ser:

$$puntuación(r_i) = -\frac{4}{8} = -\frac{1}{2} \quad (1)$$

Si todas las preguntas tienen la misma cantidad de alternativas, N_{alt} , y una sola respuesta correcta con la misma puntuación, *puntuación*(r_c), entonces todas las preguntas contribuyen de igual manera al resultado final. Luego, sin importar cuántas sean las preguntas, para compensar de un modo preciso y justo el efecto del azar basta que, en cada pregunta, las ($N_{alt} - 1$) alternativas incorrectas compensen la puntuación de la respuesta correcta. Así, en el caso general, la puntuación de cada respuesta incorrecta deberá ser

$$puntuación(r_i) = -\frac{puntuación(r_c)}{N_{alt} - 1} \quad (2)$$

La Ec. (2) tenderá a compensar adecuadamente el efecto del azar en la medida en que la cantidad de preguntas sea grande y la cantidad de alternativas sea uniforme.

Además de estas condiciones, que no parecen demasiado restrictivas, en nuestra experiencia detectamos un par de inconvenientes con este modo de compensar el efecto del azar:

i) Si bien la mayoría de los SGA más conocidos admiten puntuaciones negativas, algunos no admiten cualquier valor y, por tanto, no siempre se puede aplicar la Ec. (2). Además, en preguntas de dos opciones (*verdadero o falso*), algunos no admiten puntuaciones negativas (justamente en el caso en que el efecto del azar puede introducir un sesgo mayor, como veremos más adelante).

ii) Más importante aún: muchos de nuestros colegas y la mayoría de los estudiantes se resisten a aceptar las puntuaciones negativas por considerarlas demasiado severas o injustas. En particular, los estudiantes del ciclo básico, quienes todavía no han estudiado la teoría de probabilidades, interpretan que este método es simplemente un castigo excesivo a su ignorancia.

C. Elevación de la puntuación mínima necesaria para aprobar

Una alternativa mucho más aceptada es elevar la calificación necesaria para aprobar el examen. Al respecto, debe notarse que en las evaluaciones convencionales (escritas o verbales) la escala numérica de calificaciones no guarda una relación lineal con los objetivos ni contenidos, y esto es ampliamente aceptado por toda la comunidad educativa. Es decir, todo alumno sabe que aunque para aprobar un examen sea suficiente una calificación de 4 puntos en una escala del 1 al 10, no obtendrá esos cuatro puntos si solo sabe un 40% de lo que debería saber, o si solo alcanza un 40% de los objetivos, porque difícilmente un docente apruebe a un alumno que ha alcanzado menos de la mitad de las metas del curso.

En las pruebas virtuales, en cambio, existe una perfecta linealidad entre la escala numérica y lo que el estudiante demuestra saber o ignorar, y este hecho tan simple se debe tener en cuenta cuando se fija una calificación mínima para aprobar.

D. Cantidad de preguntas y cantidad de opciones

En conversaciones informales con docentes de distintas áreas del conocimiento, hemos notado que muchos no tienen una idea muy clara sobre cómo influye la cantidad de preguntas y la cantidad de opciones en la probabilidad de aprobar por azar. Algunos piensan, equivocadamente, que el simple hecho de que la prueba tenga muchas preguntas disminuye el efecto del azar. Otros advierten, acertadamente, que el efecto del azar depende no solo de la cantidad de preguntas sino también de la cantidad de alternativas, pero no saben cómo interactúan estas dos variables.

Si se tiene una prueba de opción múltiple con un dado número de *preguntas independientes*, cada una de las cuales tiene la misma cantidad de alternativas, con una sola correcta, y con la misma puntuación, la probabilidad de aprobar esta prueba solo por azar se puede calcular, con relativa facilidad, con la función de distribución binomial de probabilidad (Devore, 2016). (Nótese que lo resaltado en letra cursiva es una condición necesaria para la aplicabilidad de la distribución binomial).

Si el número de alternativas en cada pregunta es N_{alt} , y una sola de ellas es correcta. La probabilidad p de acertar por azar la respuesta correcta, en cualquier pregunta, es

$$p = \frac{1}{N_{alt}} \quad (3)$$

Ahora bien, si se realizan n preguntas, es claro que la cantidad de preguntas contestadas correctamente por el estudiante puede variar entre 0 y n . Asumiendo que todas las respuestas correctas tienen la misma probabilidad, dada por la Ec. (3), y que las n preguntas son independientes (la respuesta correcta de cualquiera de ellas no depende de las otras), entonces, por ejemplo, la probabilidad de responder correctamente 2 preguntas de un total de 15 preguntas, se calcula como

$$P_{2,15} = \binom{15}{2} p^2 (1-p)^{15-2} \quad (4)$$

donde el primer factor del segundo miembro, llamado *coeficiente binomial*, representa, en este ejemplo, la cantidad de formas en que se puede agrupar o combinar 2 elementos tomados de un conjunto de 15 elementos, sin importar el orden.

En general, la probabilidad de acertar, por azar, k respuestas en un examen de n preguntas será

$$P_{k,n} = \binom{n}{k} p^k (1-p)^{n-k} \quad (5)$$

Luego, la probabilidad de acertar por azar "al menos" k respuestas correctas en un cuestionario de n preguntas, es decir, de acertar k o más respuestas, está dada, como cabe esperar, por la suma desde k hasta n de la Ec. (5):

$$P_{j \geq k, n} = \sum_{j=k}^n \binom{n}{j} p^j (1-p)^{n-j} \quad (6)$$

Si las respuestas incorrectas tienen puntuación nula (no están penalizadas) y cada respuesta correcta vale un punto, la Ec. (6) representa la probabilidad de obtener una calificación de al menos k puntos en una escala de 1 a n . Consideremos un ejemplo:

Ejemplo 2: Se desea calcular la probabilidad de acertar por azar 4 o más respuestas de un cuestionario de 10 preguntas, donde todas tienen 2 opciones (*verdadero o falso*).

En este caso, se tiene: $n=10$, $k=4$ y $p=1/2$; y la Ec. (6) da $P_{j \geq 4, 10} \approx 0.8281$.²

Si el alumno aprueba con una calificación de 4 y todas las preguntas valen un punto, la probabilidad de aprobar por mero azar en una prueba de este tipo es $\sim 83\%$. Esto no debe extrañarnos, porque la probabilidad de acierto por azar (50%) es bastante mayor que la fracción exigida de respuestas correctas (40%).

Contrariamente a lo que algunos suponen, en un caso como este, *la probabilidad de aprobar no disminuye cuando la cantidad de preguntas aumenta*. Por el contrario, si la probabilidad de acierto es mayor que la fracción exigida de respuestas correctas, la probabilidad de aprobar aumenta con el número de preguntas. Esto se debe al principio probabilístico de estabilidad de las frecuencias relativas para números grandes: cuanto mayor sea la cantidad de preguntas, la fracción de aciertos al azar convergerá a la probabilidad p , que en este caso es mayor que la fracción exigida k/n (4/10). Para ilustrarlo veamos la siguiente tabla:

² Los cálculos probabilísticos de este trabajo fueron realizados con la función de distribución binomial "cdfbin" del programa *Scilab*, versión 5.5.1.

TABLA I. Preguntas con 2 opciones ($p = 0.5$), fracción exigida de respuestas correctas $k/n = 0.4$

N.º de preguntas	Probabilidad de aprobar por azar
10	0.828125
20	0.8684120
30	0.8997558
40	0.9230700
50	0.9405398

Si, por el contrario, la fracción exigida de respuestas correctas fuera igual a la probabilidad de acierto, la probabilidad de aprobar converge al valor 0.5, como se observa en la tabla II, donde se han considerado cantidades grandes de preguntas para evidenciar esta convergencia:

TABLA II. Preguntas con 2 opciones ($p = 0.5$), fracción exigida de respuestas correctas $k/n = 0.5$

N.º de preguntas	Probabilidad de aprobar por azar
10	0.6230469
150	0.5325193
300	0.5230138
450	0.5187959
600	0.5162800

Consideremos ahora cómo evoluciona la probabilidad con la cantidad de alternativas, para cantidades moderadas de preguntas y una misma fracción exigida de respuestas correctas, $k/n = 0.4$.

TABLA III. Preguntas con 3 opciones ($p = 0.3333$), fracción exigida de respuestas correctas $k/n = 0.4$

N.º de preguntas	Probabilidad de aprobar por azar
10	0.4407357
20	0.3385285
30	0.2761356
40	0.2311599
50	0.1964139

Al comparar las tablas I y III, se observa cómo, para una misma fracción de respuestas correctas exigidas, aumentar de 2 a 3 las alternativas de las preguntas cambia el comportamiento cualitativo de la probabilidad de aprobar por azar, que disminuye al aumentar la cantidad de preguntas. Esto es lo que cabe esperar, porque en este último caso, la fracción de respuestas correctas exigidas, $4/10$, es mayor que la probabilidad de acertar cada pregunta, $1/3$.

En la tabla IV, que muestra los resultados de 4 alternativas por pregunta, la probabilidad de aprobar sigue decreciendo con la cantidad de preguntas, como cabe esperar, pero ahora mucho más abruptamente.

TABLA IV. Preguntas con 4 opciones ($p = 0.25$), fracción exigida de respuestas correctas $k/n = 0.4$

N.º de preguntas	Probabilidad de aprobar por azar
10	0.2241249
20	0.1018119
30	0.0506583
40	0.0262449
50	0.0139176

En la tabla V, se consideran 5 alternativas por pregunta, y se observa también un decrecimiento abrupto de la probabilidad de aprobar por azar, a medida que la cantidad de preguntas aumenta.

TABLA V. Preguntas con 5 opciones ($p = 0.2$), fracción exigida de respuestas correctas $k/n = 0.4$

N.º de preguntas	Probabilidad de aprobar por azar
10	0.1208739
20	0.0321427
30	0.0094931
40	0.0029362
50	0.0009324

Todos estos resultados muestran que la probabilidad de aprobar por azar en una prueba de opción múltiple depende no solo de la cantidad de preguntas o ítems que contenga el examen, sino de la fracción de respuestas correctas que se exija para aprobar y de la cantidad de alternativas que incluya cada ítem (esto es lo que determina la probabilidad de acertar por azar en un ítem cualquiera (v. Ec. (3))). Debido a que, por una parte, la fracción exigible de respuestas correctas, o lo que es equivalente, la calificación mínima necesaria para aprobar, no se puede aumentar arbitrariamente ni indefinidamente, y los exámenes demasiado extensos son agotadores tanto para estudiantes como para docentes, la variable más adaptable para moderar la probabilidad de acertar por azar parece ser la cantidad de alternativas por pregunta, pero desde luego, si el conjunto de aprendizajes que se desea evaluar es demasiado o el grado de detalle muy minucioso, puede convenir un balance con más preguntas y menos opciones.

Estos ejemplos y resultados ilustran el efecto del azar y el rol de las variables de las cuales depende, pero pueden parecer poco prácticos, pues solo consideran el caso de que un alumno responda todas las preguntas al azar, lo cual solo ocurriría si se presentara a rendir sin haber logrado ningún aprendizaje.

Lo primero que cabe señalar al respecto es que este caso límite, tomado aquí como referencia, si bien no es habitual, tampoco es imposible. A lo largo de nuestra carrera docente nos ha ocurrido varias veces que un alumno entrega su examen en blanco o con unas pocas respuestas incorrectas; lo cual, en general, no se debe a un afán extremo de especular con el azar ni a una actitud negligente o descarada. Antes bien, estos hechos pueden ser indicios de malos hábitos de estudio, mala administración del tiempo y otras circunstancias personales o sociales cuya discusión está fuera del contexto de este trabajo.

Pero lo que es más importante, la distribución binomial puede aplicarse fácilmente a casos más típicos. Consideremos, por ejemplo, la primera fila de la tabla V, la cual nos dice que en una prueba de 10 preguntas, cada una con 5 alternativas, que requiera una calificación de 4 puntos para aprobar, en promedio aprobará un 12% de los alumnos que se presente sin ningún nivel de aprendizaje. Este mismo resultado nos permite resolver otros casos como el del siguiente ejemplo:

Ejemplo 3: Se diseña una prueba con 12 preguntas, cada una con 5 opciones de las cuales una sola es correcta y vale un punto. Se fija, como exigencia para aprobar, que se conteste correctamente al menos la mitad de las preguntas, es decir, obtener al menos 6 puntos. (Nótese que este examen es más exigente que el planteado en la tabla V, tanto por el hecho de tener más preguntas como por el de exigir una calificación más alta.)

Si la prueba está bien elaborada será válida y confiable, y se puede considerar que el alumno responderá correcta y certeramente en la medida en que haya alcanzado los objetivos del aprendizaje que dicha prueba pretende medir. Bajo esta hipótesis, cabe suponer, por ejemplo, que un alumno que sabe con certeza solo 2 preguntas de las 12 del examen ha alcanzado menos del 17% ($2/12 < 0.17$) de los objetivos del aprendizaje (un nivel bastante pobre). Para este caso particular, la primera fila de la tabla V nos dice que este alumno también tiene una probabilidad de aprobar cercana al 12%. Esto se infiere del hecho de que un alumno que sepa con certeza las respuestas de 2 preguntas, verá que su examen se ha reducido a 10 preguntas que debe contestar al azar, y solo debe sumar 4 puntos más a los 2 puntos que ya tiene asegurados. Esta situación es justamente la que describe la primera fila de la tabla V.

Desde luego, las preguntas no pueden dividirse taxativamente entre las que un dado alumno sabe con certeza y las que contesta completamente al azar; entre estos dos extremos hay toda una gradación de aprendizajes más o menos seguros o inseguros. Pero los saberes inseguros deben preocuparnos mucho menos que el desconocimiento absoluto y, para nuestros fines, al menos en primera aproximación, el tipo de razonamiento aplicado al Ejemplo 3 se podría aplicar a cualquier caso práctico para calcular la probabilidad de aprobar según el nivel de aprendizaje alcanzado por los alumnos. Esto es lo que analizaremos en los resultados presentados en las siguientes tablas.

Consideremos en primer lugar, dos variantes de una prueba de 15 preguntas, cada una con 5 alternativas, pero con diferentes grados de exigencia para aprobar. En el primer caso (tabla VI) se exige una cantidad de 6 respuestas correctas ($k/n = 0.4$), en tanto que en el segundo caso (tabla VII) se exigen 8 respuestas correctas ($k/n \approx 0.533$).

TABLA VI. Probabilidad de aprobar una prueba de 15 preguntas con 5 opciones ($p = 0.2$) y una fracción exigida de respuestas correctas $k/n = 0.4$, según la cantidad de respuestas correctas que el alumno sabe con certeza.

Cant. de resp. conocidas	Nivel de aprendizaje (%)	Probabilidad de aprobar
0	0	0.0610514
1	6.67	0.1298396
2	13.33	0.2526757
3	20	0.4416543
4	26.67	0.6778775
5	33.33	0.8926258

TABLA VII. Probabilidad de aprobar una prueba de 15 preguntas con 5 opciones ($p = 0.2$) y una fracción exigida de respuestas correctas $k/n = 0.533$, según la cantidad de respuestas correctas que el alumno sabe con certeza.

Cant. de resp. conocidas	Nivel de aprendizaje (%)	Probabilidad de aprobar
0	0.	0.0042397
1	6.67	0.0116099
2	13.33	0.0300353
3	20	0.0725555
4	26.67	0.1611392
5	33.33	0.3222005
6	40	0.5637924
7	46.67	0.8322278

Nótese que las 6 primeras filas de las tablas VI y VII tienen la misma cantidad de respuestas conocidas o nivel de aprendizaje, mostrando que para un mismo nivel de aprendizaje, la probabilidad de aprobar disminuye drásticamente en la tabla VII.

Consideremos ahora una prueba con el doble de preguntas que las de las tablas VI y VII, pero con el mismo nivel de exigencia en cada caso, en términos de k/n .

TABLA VIII. Probabilidad de aprobar una prueba de 30 preguntas con 5 opciones y una fracción exigida de respuestas correctas $k/n = 0.4$, según la cantidad de respuestas correctas que el alumno sabe con certeza. (Se han eliminado las cantidades impares de respuestas conocidas, para facilitar la comparación con la tabla VI.)

Cant. de resp. conocidas	Nivel de aprendizaje (%)	Probabilidad de aprobar
0	0	0.0094931
2	6.67	0.0390707
4	13.33	0.1312912
6	20	0.3441076
8	26.67	0.6679586
10	33.33	0.9308247

TABLA IX. Probabilidad de aprobar una prueba de 30 preguntas con 5 opciones y una fracción exigida de respuestas correctas $k/n \approx 0.533$, según la cantidad de respuestas correctas que el alumno sabe con certeza. (Se han eliminado las cantidades impares de respuestas conocidas, para facilitar la comparación con la tabla VII.)

Cant. de resp. conocidas	Nivel de aprendizaje (%)	Probabilidad de aprobar
0	0	0.0000524
2	6.67	0.0003731
4	13.33	0.0023430
6	20	0.0126211
8	26.67	0.0561446
10	33.33	0.1957922
12	40	0.4989745
14	46.67	0.8592625

De la comparación de las tablas VI y VIII, por una parte, y VII y IX, por otra, se observa que para un mismo nivel de aprendizaje, las pruebas que tienen 30 preguntas muestran probabilidades menores que las que tienen 15 preguntas, para niveles de aprendizaje relativamente pobres, es decir, las pruebas más extensas son más difíciles para los alumnos con un bajo nivel de aprendizaje. Pero a medida que el nivel del alumno se aproxima a la exigencia, la probabilidad de aprobar, en ambos tipos de prueba, se torna comparable, y finalmente la probabilidad aumenta en las pruebas de 30 preguntas (comparar las últimas filas de las tablas VI y VIII, y de las tablas VII y IX).

Esto es bastante intuitivo, porque un alumno con un bajo nivel de aprendizaje debería acertar muchas preguntas por azar en las pruebas extensas; pero a medida que el nivel de aprendizaje del alumno se aproxima al nivel exigido son cada vez menos las preguntas que debe responder al azar (o sin tener completa certeza), y las pruebas más extensas ofrecen más posibilidades (más preguntas) para esos pocos aciertos que necesita.

IV. CONCLUSIONES

El uso de entornos virtuales y sistemas de gestión de aprendizaje (EVA y SGA) aumentó en las últimas décadas y más en 2020, debido a la necesidad indefectible de adoptar la modalidad virtual de enseñanza. Aun cuando cabe esperar que la emergencia termine en un plazo razonable, pueden repetirse muchas coyunturas en que la principal modalidad educativa sea la virtual, y en cuanto respecta a la evaluación, la modalidad virtual se ha impuesto desde hace muchos años, en situaciones en que se requiere evaluar a grandes grupos heterogéneos con un alto grado de objetividad, equidad e inclusión social. Este crecimiento de la modalidad virtual es también previsible debido al continuo avance tecnológico y al hecho de que docentes y discentes aprecian las ventajas que ofrece. Las pruebas objetivas, uno de los recursos más usados de la enseñanza virtual, presentan ventajas y desventajas. Deben considerarse como una modalidad válida, pero complementaria de otros tipos de evaluación. Las evaluaciones de desarrollo, por ejemplo, son más aptas para medir ciertos aprendizajes, como creatividad, organización discursiva, etc.

Hasta el presente se han discutido y consensuado importantes normativas respecto del diseño y elaboración de las pruebas objetivas, en particular, de las de opción múltiple, para garantizar su validez y confiabilidad, pero aún quedan aspectos, como el rol del azar, que requieren un análisis minucioso en términos matemáticos. La probabilidad de aprobar por azar puede compensarse de manera eficaz y simple mediante la puntuación negativa de las respuestas incorrectas, pero este método es poco aceptado por docentes y estudiantes, y tiene el inconveniente de que los SGA no ofrecen la versatilidad suficiente para implementarlo apropiadamente.

Otra posibilidad para contrarrestar el efecto del azar es el diseño de las pruebas en base a cálculos probabilísticos, los cuales deben contemplar tres variables fundamentales: el nivel de exigencia, la cantidad de ítems y la cantidad de alternativas por ítem. Esta opción es más aceptada, pero requiere un análisis minucioso que los docentes de áreas del conocimiento ajenas a la matemática no están en condiciones de hacer. El estudio desarrollado en este trabajo plantea algunas cuestiones generales al respecto, pero la implementación general, de manera efectiva y sistemática, necesitaría desarrollos más amplios por equipos técnicos que se dediquen a tipificar modelos de pruebas adaptables a cada necesidad.

Entre los principales resultados presentados en este trabajo, cabe destacar que la cantidad de ítems de las pruebas de opción múltiple, contrariamente a lo que muchos creen, tiene una incidencia relativa sobre el efecto del azar, y cabe prevenir a los docentes sin formación matemática, que si el grado de exigencia es menor o comparable a la probabilidad de acierto de cada ítem (como suele ser el caso de las preguntas de dos opciones) la probabilidad de aprobar por azar aumenta con la cantidad de preguntas. Por otra parte, el hecho de incluir muchas preguntas, si bien puede mejorar la confiabilidad de la evaluación, afecta poco la probabilidad de aprobar de los alumnos con un nivel de aprendizaje cercano al nivel de exigencia, como se deduce de la comparación de los resultados correspondientes a pruebas extensas y cortas (tablas VI, VII, VIII y IX).

Así, en el diseño, a los fines de contrarrestar el azar, las variables más críticas son el nivel de exigencia y la cantidad de alternativas de los ítems, de tal modo que el nivel de exigencia debe ser razonablemente mayor que la probabilidad de acertar cada ítem por azar.

Cualquiera sea la estrategia que se utilice, el rol del azar es un problema que se debe resolver, pues lo contrario implica promover en su carrera a alumnos que no han alcanzado los objetivos necesarios de aprendizaje, y esto puede afectar no solo al sistema educativo sino a toda la sociedad.

AGRADECIMIENTOS

Este trabajo se realizó en el marco del proyecto Piunt E629, subsidiado por la Secretaría de Ciencia, Técnica e Innovación Tecnológica de la Universidad Nacional de Tucumán.

REFERENCIAS

Aspa Marco, F. y Rodríguez de Castro, F. (2010). Evaluación final: ¿sirve el examen MIR? *Educ Med*, 13(1), 73-77.

Belloch, C. (s/f). Entornos Virtuales de Aprendizaje (documento inédito), Unidad de Tecnología Educativa, Universidad de Valencia. Recuperado de <https://www.uv.es/bellochc/pedagogia/EVA3.pdf>.

Cano, M. (2008). La evaluación por competencias en la educación superior. *Revista de Currículo y Formación de Profesorado*, 12(3). Recuperado de <http://www.ugr.es/local/recfpro/rev123COL1.pdf>

Centro de Estudios Científicos y Tecnológicos N° 13. (2010). Manual Cómo Elaborar Pruebas Objetivas (documento inédito). Manual, Instituto Politécnico Nacional, México. Recuperado de <https://www.repositoriodigital.ipn.mx/bitstream/123456789/5792/1/pruebas%5B1%5D.pdf>

Chernoff, H. (1962). The Scoring of Multiple-Choice Questionnaires. *Ann. Math. Statist.* 33(2), 375-393. Recuperado de <https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-33/issue-2/The-Scoring-of-Multiple-Choice-Questionnaires/10.1214/aoms/1177704565.full>

Devore, J. (2016). *Probability and Statistics for Engineering and the Sciences* (9th. Ed.). Boston: Cengage Learning.

Educrea (s/f). Pruebas de Selección Múltiple. <https://educrea.cl/pruebas-de-seleccion-multiple/> Visitado el 24/10/2020.

Entorno virtual de aprendizaje. (2020). En Wikipedia, la enciclopedia libre. https://es.wikipedia.org/w/index.php?title=Entorno_virtual_de_aprendizaje&oldid=135037758 Visitado el 24/10/2020.

García-Garro, A., Ramos-Ortega, G., Díaz de León-Ponce, M. y Olvera-Chávez, A. (2007). Instrumentos de evaluación. *Revista Mexicana de Anestesiología*, 30(3), 158-164.

López Meneses, E. y Miranda Velasco, M. (2007). Influencia de la Tecnología de la Información en el Rol del Profesorado y en los Procesos de Enseñanza-Aprendizaje. *Revista Iberoamericana de Educación a Distancia*, 10(1), 51-60.

López Rayón Parra, A., Ledesma Saucedo, R. y Escalera Escajeda, S. (2009). Ambientes Virtuales de Aprendizaje (documento inédito). Recuperado de www.comunidades.ipn.mx/.../168ambientes%20virtuales%20de%20aprendizaje

Macías Calvillo, E. (2011). Validación y confiabilidad de pruebas de opción múltiple para la evaluación de habilidades (tesis de maestría no publicada), Centro de Investigación en Matemáticas, A. C., Guanajuato. Recuperada de <https://cimat.repositorioinstitucional.mx/jspui/bitstream/1008/245/2/TE%20373.pdf>

Millán Núñez-Cortés, J. (2010). Evaluación del alumno. *Educ Med*, 13(1), 57-60.

Mousalli-Kayat, G. (2017). Los instrumentos de Evaluación en la Investigación Educativa (documento inédito), Universidad de los Andes, Mérida. Recuperado de <https://www.researchgate.net/publication/321397866>

Noriega Echevarría, I. (2010). Evaluación objetiva: el punto de vista de los estudiantes. *Educ Med*, 13(1), 79-82.

Onrubia, J. (2005). Aprender y enseñar en entornos virtuales: actividad conjunta, ayuda pedagógica y construcción del conocimiento. *Revista de Educación a Distancia*, (monográfico II). Recuperado de <http://www.um.es/ead/red/M2/>

Ruiz Betancourt, T. A., Rodríguez Guaraca, N. M., Gallegos Arias, L. Villacis Vallejo, J. M. (2018). Las Pruebas Objetivas como Instrumento de Medición de los Resultados de Aprendizaje. *Revista Atlante: Cuadernos de Educación y Desarrollo*. Recuperado de <https://www.eumed.net/rev/atlante/2018/06/pruebas-resultados-aprendizaje.html>

Shakil, M. (2009). Using Beta-binomial Distribution in Analyzing Some Multiple-Choice Questions of the Final Exam of a Math Course, and its Application in Predicting the Performance of Future Students. *POLYGON - A Web-Based, Multi-Disciplinary Publication, Miami-Dade College, Hialeah Campus*.

Stufflebeam, D. y Shinkfield, A. (1987). *Evaluación Sistemática*. Barcelona: Paidós.

Tristán López, A. y Pedraza Corpus, N. (2017). La Objetividad en las Pruebas Estandarizadas. *Revista iberoamericana de Evaluación Educativa*, 10(1), 11-31.

Prueba objetiva. (s. f.). Universitat Oberta de Catalunya. Assistent de Recursos Metodològics. Recuperado el 24 de octubre de 2020, de http://cv.uoc.edu/UOC/a/moduls/90/90_148_h/lista_recursos.htm

Vázquez, G., Murillo-Cabezas, F., Gómez, J., Martín, C., Chaves, J. y Peinado, J. L. (2008). El examen MIR, su cambio como una opción estratégica. *Educ Med*, 11(4), 203-206.