

# Análisis del impacto del uso de tres opciones de los exámenes de selección múltiple en la cantidad de alumnos aprobados y distractores no funcionantes y en los índices de dificultad y discriminación

\*VECE DE REYNAGA M.M, \*\*D'URSO M., \*\*\*TEJERIZO FE M.M

\*Facultad de Medicina U.N. de Tucumán, Argentina; \*\*Facultad de Medicina U.N. de Tucumán, Argentina;

\*\*\*Facultad de Medicina, U.N. de Tucumán, Argentina

✉ vecereynaga@gmail.com, durso.marcela@gmail.com, mtejerizofe@hotmail.com

## RESUMEN

**Metodología:** Se trabajó con 7 evaluaciones del 2012-2013 (4 opciones), 9 del 2014-2015 (3 opciones) y el software donde se registraron alumnos aprobados, distractores no funcionantes (DNF) e índices de dificultad (p) y discriminación (ID) de cada ítem.

**Objetivos:** analizar exámenes 2012-2013 y comparar con 2014-2015, para comprobar si al reducir el número de opciones hubo cambios en: 1) porcentaje de alumnos aprobados; 2) cantidad de DNF; 3) porcentaje de ítems con dificultad media (DM) y de discriminación buenos y muy buenos (B-MB) relacionados con el número de DNF.

**Resultados:** 1) Alumnos aprobados: no hay diferencias significativas ( $p=0,082$ ). 2) Diferencia significativa ( $p < 0,0001$ ) en el incremento del promedio de ítems con 0 DNF. 3) A menor número de DNF en cada ítem, mayor porcentaje de ítems con DM y B-MB. En ambos períodos hay diferencia significativa ( $p < 0,0001$ ).

**Conclusiones:** comparando los resultados de evaluaciones con ítems de 3 opciones en relación con los de 4, se logró demostrar que: el número de alumnos aprobados no varía; se produjo un incremento significativo en el número de ítems con 0 DNF, la mayoría de los cuales van acompañados de buenos p e ID.

**Palabras clave:** distractores no funcionantes, exámenes de selección múltiple, índices de dificultad, índice de discriminación

## 10 - EVALUACIÓN DEL APRENDIZAJE

### ANTECEDENTES

La evaluación de los instrumentos de examen es una parte importante de la tarea educativa. Este concepto fue claramente descrito hace algunos años en la publicación de Scriven: *“los evaluadores tienen la obligación profesional de que las evaluaciones propuestas o finalizadas estén sujetas a una evaluación competente”*<sup>1</sup>.

El pedagogo José Gimeno Sacristán, en su libro *“Teoría de la Enseñanza y Desarrollo del Currículum”*<sup>2</sup>, propone considerar a la evaluación como una *“comprobación de la validez de las estrategias didácticas”*, es decir, el momento de buscar los datos que nos ayuden a decidir si la estrategia metodológica desarrollada fue o no adecuada, o en qué medida lo fue, para guiar un proceso de enseñanza que desemboque en la obtención de los resultados de aprendizaje previamente propuestos. Desde este enfoque, uno de los grandes valores de la evaluación es ser un instrumento de investigación en la didáctica.

La Cátedra de Microbiología de la Facultad de Medicina de la UNT implementó en el año 2007 un proyecto de innovación pedagógica en el cursado intensivo de la materia, que proponía como estrategia de enseñanza-aprendizaje el *“aprendizaje basado en problemas”* y como método de evaluación el examen escrito estructurado de selección múltiple con 100 ítems de 4 opciones, en reemplazo de la evaluación oral.

Clásicamente se señala que la mayor desventaja de estas pruebas es que evalúan conocimiento de tipo memorístico, más que razonamiento elaborado y la diferencia entre *“reconocer”* la respuesta correcta en lugar de recuperarla de la memoria. Esto ocurre especialmente cuando son construidas en formato de estímulo pobre en contexto, sin embargo, es posible diseñar ítems con descripciones ricas del mismo que simulan casos reales y evalúan más adecuadamente las competencias metodológicas. Teniendo en cuenta la complejidad de este tipo de pruebas, en cuanto a su elaboración, redacción de ítems, necesidad de re-

cursos humanos calificados y tiempo necesario para realizarlas, se decidió convocar a todos los docentes de la cátedra para su construcción. Se elaboró un banco de 1000 ítems (aproximadamente 100 por cada unidad temática) que fueron revisados por los profesores. A partir de este banco se seleccionaron los que se incluyeron en cada examen.

Desde entonces se empezaron a desarrollar, de manera sistemática, mecanismos de evaluación de calidad de los exámenes para poder abordar procesos de transformación y mejora.

Durante los años **2007** y **2008** se realizó la evaluación de los exámenes, teniendo en cuenta las características psicométricas: confiabilidad de la prueba, validez de los ítems y sus índices de dificultad ( $p$ )<sup>3</sup> y discriminación (ID)<sup>4</sup>.

En **2009** se agregó el análisis de distribución de los distractores no funcionantes (DNF)<sup>5</sup> (aquellos que son respondidos por menos del 5% del total de alumnos que rinden la prueba). Es importante destacar que tanto  $p$ , como ID y el análisis de la distribución de los DNF, sólo pueden realizarse luego de tomar la evaluación y que son siempre relativos al grupo de sujetos que rinden.

A partir del **2010** se decidió aplicar antes de cada evaluación, el índice de calidad desarrollado por el Dr. Alberto Galofré<sup>6</sup> (ICG) teniendo en cuenta las recomendaciones para la construcción de los ítems, lo que permitió mejorar o eliminar aquellos que poseían defectos.

Al analizar los resultados se observó un nivel de calidad de los ítem y de los exámenes satisfactorio (ICG de 3 o mayor) y una clara tendencia al aumento de ítems con 0 DNF. A pesar de estas mejoras, menos del 30% tenían 0 DNF. Se llegó a la conclusión de que esto merecía una revisión de la plausibilidad de los distractores y cómo plantean actualmente diferentes autores<sup>5, 14</sup>, considerar la posibilidad de utilizar sólo 2 igualmente atractivos.

Los resultados obtenidos, motivaron a los docentes de la cátedra a reducir el número de opciones en los exámenes a partir del año 2014.

Las preguntas que surgieron fueron: ¿al disminuir las opciones, se incrementan las posibilidades de acertar y como consecuencia la cantidad de alumnos aprobados? ¿Cómo afecta este cambio en los DNF y los  $p$  e ID de los ítems?

De allí surgieron los objetivos del presente trabajo: analizar los exámenes de los años 2012-2013 (4 opciones) y compararlos con los de los años 2014-2015 (3 opciones), para comprobar si hubo cambios en: 1) el porcentaje de alumnos aprobados; 2) la cantidad de DNF y 3) el porcentaje de ítems con dificultad media (DM) y de discriminación buenos y muy buenos (B-MB) relacionados con el número de DNF.

## DESCRIPCIÓN

Se realizó un trabajo descriptivo, de corte transversal, cuantitativo, de evaluación de exámenes de opción múltiple. Se trabajó con 7 evaluaciones de los años 2012-2013 (4 opciones), 9 evaluaciones de los años 2014-2015 (3 opciones) y el software<sup>8</sup> de recolección de datos. El método de corrección de los exámenes finales se realizó de forma electrónica. Los datos enviados por la dirección de tecnologías de información y comunicación se transcribieron al software que calcula el porcentaje de alumnos aprobados; el  $p$  e ID y el porcentaje de elección de cada distractor el cual permite catalogar como DNF al que fue elegido por menos del 5% de los examinados. Si un distractor no cumplía con este requisito significa que no funcionó como tal<sup>9, 10, 11, 12</sup>. El  $p$  expresa la proporción de sujetos que respondió correctamente un ítem de una prueba.

El ID sirve para diferenciar entre los alumnos que saben y los que no.

Para realizar el análisis estadístico de los resultados se empleó el Test chi cuadrado y el test  $t$ .

## RESULTADOS

No hay diferencias significativas en la proporción de aprobados entre un período y otro.

Se realizó un análisis de la varianza (ANOVA) con comparaciones múltiples de Bonferroni, para comparar los valores medios porcentuales de DM y B-MB entre los distractores para cada período. De las comparaciones múltiples se observa que:

- En ambos períodos, el valor medio porcentual de ítems con DM, es significativamente mayor en aquellos con 0 DNF respecto a los que tienen 1 y 2 DNF. También existe una diferencia significativa entre los valores medios porcentuales correspondientes a los ítems con 1 y 2 DNF.
- Durante el período 2012-2013, en los ítems B-MB no se encontraron diferencias significativas en el valor medio porcentual, entre los con 0 y 1 DNF. Al analizar la varianza y realizar las comparaciones múltiples de Bonferroni, se encontraron diferencias significativas entre 0 y 2 DNF y entre 1 y 2.
- En el período 2014-2015, el valor medio porcentual de ítems B-MB es significativamente mayor en

**Tabla 1.** Porcentaje de alumnos aprobados en los diferentes períodos

Períodos	Porcentaje
2012-2013 (n = 422)	75% (316)
2014-2015 (n = 502)	75% (377)

Test chi cuadrado  $p = 0.939$ .

aquellos con 0 DNF respecto a los que tienen 1 y 2 DNF. También existe una diferencia significativa entre los valores medios porcentuales correspondientes a los ítems con 1 y 2 DNF.

## CONCLUSIONES Y DISCUSIÓN

Comparando los resultados de evaluaciones con ítems de 3 opciones en relación con los de 4, se logró demostrar que el disminuir el número de opciones, no influye en la cantidad de alumnos aprobados (Tabla 1 y que se produce un incremento significativo en el número de ítems con 0 DNF (Tabla 2). Éstos son los que tienen mayor porcentaje de DM y B-MB (Tabla 3), lo que ratificó la decisión de la cátedra de continuar con evaluaciones con ítems de 3 opciones.

Este número de opciones debería aplicarse de manera sistemática a todos los ítems de opción múltiple que componen el banco de preguntas de una cátedra, teniendo en cuenta que la mejora de los mismos no está relacionada con la cantidad total de distractores, sino con la calidad.

Además, de los resultados obtenidos, es importante resaltar como beneficios el hecho de que reducen el tiempo de elaboración al docente y consumen menos tiempo al alumno para responder.

La revisión de la plausibilidad de los distractores y la utilización de sólo 2 igualmente atractivos, permitió obtener exámenes con el 72% de los ítems con 0 DNF. En un trabajo publicado por Marie Tarrant<sup>12</sup> obtuvieron un 13,8% con 0 DNF (4 opciones).

**Tabla 2.** Distribución promedio del porcentaje de ítems según la cantidad de DNF por períodos

DNF	0	1	2
Años 2012- 2013	26	42	28
Años 2014- 2015	72	26	2
Test t	p < 0,0001	p < 0,0001	p < 0,0001

**Tabla 3.** Análisis de asociación entre número de DNF y porcentaje de DM y B-MB en cada período

DNF	2012-2013 Período 1 (P1)		2014-2015 Período 2 (P2)	
	DM (%)	B-MB	DM	B-MB
0				
nP1= 26	72.3	45.3	62.4	41.1
nP2= 73				
1				
nP1= 42	55.5	50.4	26.4	32.5
nP2= 25				
2				
nP1= 28	34.9	29.5	0	0
nP2= 2				
ANOVA	p < 0.0001	p < 0.0001	p < 0.0001	p < 0.0001

El análisis de los DNF debe realizarse siempre antes de comunicar los resultados de un examen, ya que permite analizar la validez de los ítems y decidir si alguno debe ser anulado, lo cual lleva implícito la modificación del puntaje máximo de la prueba.

## Relevancia y transferibilidad

Redactar distractores de alta calidad es una parte importante del proceso de desarrollo de la prueba.

La elaboración de exámenes de buen nivel es una ardua tarea. Se sugiere tener un banco de ítems en un archivo en el que las preguntas estén clasificadas por temas y según sus propiedades psicométricas<sup>13</sup>. La literatura especializada recomienda construir los exámenes con las preguntas ya probadas y evaluadas<sup>14</sup> y descartar el criterio de que en cada examen las preguntas deben ser nuevas. Reducir a 3 el número de opciones y aplicar el ICG para construir los ítems permite tener un banco de datos con ítems con buenos índices psicométricos y poder elaborar exámenes válidos y confiables. Documentar y difundir resultados, aporta para la discusión y la transformación de las evaluaciones mejorando su calidad. Hay que tener en cuenta que la aprobación de un examen acredita parcialmente para la obtención del título de los alumnos.

“Abordar el problema de la evaluación supone necesariamente tocar todos los problemas fundamentales de la pedagogía. Cuánto más se profundiza en su dominio, más conciencia se tiene del carácter enciclopédico de nuestra ignorancia y más cuestionamos nuestras certidumbres. Cada interrogante planteado lleva a otros interrogantes. Cada árbol se enlaza con otro y el bosque aparece en toda su inmensidad”<sup>16</sup>.

## BIBLIOGRAFÍA

1. Scriven, M. The methodology of evaluation. (1967) En R.W. Tyler, R. M. Gagné y M. Scriven (Eds.). Perspectives of curriculum evaluation, 39-83. Chicago, IL: Rand Mc Nally.
2. Gimeno Sacristán J. (1992). Teoría de la enseñanza y desarrollo del currículo. 4ª Edición: Rei. Buenos Aires. Argentina.
3. Crocker, L. y Algina, J. (1986). Introduction to Classical and Modern Test Theory. Holt, Rinehart and Winston.
4. Ebel, R.L. y Frisbie, D.A. (1986). Essentials of Education Measurement. Englewood Cliffs, NJ: Prentice Hall.
5. Tarrant M, Ware J and Mohammed A. (2009). An assessment of functioning and non- functioning distractors in multiple-choice questions: a descriptive analysis *BMC Medical Education*, 9:40-48. Artículo disponible en: <http://www.biomedcentral.com/1472-6920/9/40>.
6. Galofré A. (2005) Construcción del índice de calidad de las preguntas. <http://www.saidem.org.ar/docs/mbts>. Acceso 7 de febrero 2011.
7. Vece, M. M. y Vece M. B. (2009) “Software de Aplicación: “Evaluando evaluaciones. Cálculo de índices de dificultad”

- tad y de discriminación en evaluaciones estructuradas”. Facultad de Medicina - Facultad de Ciencias Naturales e IML de la Universidad Nacional de Tucumán. Registro dirección nacional derecho de autor 773735. ISBN 978-987-05-67448.
8. Case S, Swanson D. Nacional Board of Medical Examiners. (2006). Cómo elaborar preguntas para evaluaciones escritas en el área de ciencias básicas y clínicas. Tercera edición. Philadelphia.
  9. Guilbert JJ. (1981). Guía pedagógica para el personal de salud. OPS/OMS. Ginebra.
  10. Castro C, Galli A. (2003) Evaluación Educativa. Módulos 1-2-3. AFACIMERA. Buenos Aires. <http://www.saidem.org.ar/docs/mbts.php>. Acceso 19 de abril 2011.
  11. Lafourcade P. (1969) Evaluación de los aprendizajes. Kapelusz. Buenos Aires.
  13. Barbero, M.I. bancos de ítems. (1996). En J. Muñiz (ed.). Psicometría. Madrid. Universitas.
  14. Case S, Swanson D. Nacional Board of Medical Examiners. (2006). Como elaborar preguntas para evaluaciones escritas en el área de ciencias básicas y clínicas. Tercera edición Philadelphia.