

Los que se arriesgan, ¿saben cuánto se arriesgan?

Rosa BERSABÉ MORÁN

Universidad de Málaga

Rosario MARTÍNEZ ARIAS

Universidad Complutense de Madrid

Ricardo TEJEIRO SALGUERO

Universidad de Málaga

Resumen

En una situación real, dos grupos de sujetos que manifestaban diferente nivel de riesgo se compararon en diferentes medidas de precisión en los juicios de probabilidad. Después de contestar un examen de 20 ítems tipo verdadero-falso, se pidió a los alumnos que estimaran la probabilidad de que cada una de las preguntas fuera verdadera. Se formaron dos grupos según el nivel de riesgo de los sujetos. Ambos consiguieron las mismas notas en el examen, pero el grupo de arriesgados contestó un mayor número de preguntas. Los datos obtenidos sugieren que el riesgo afecta a la calibración y al ruido de los juicios probabilísticos. Los sujetos más arriesgados calibraron peor, es decir, las probabilidades que estimaron se desviaban más de las proporciones reales de ítems verdaderos. Por tanto, parece que los que se arriesgan más, saben menos cuánto se arriesgan. La discusión se centra en las implicaciones de estos resultados sobre la conducta en los juegos de azar. Se acompañan gráficos de calibración y covariación.

Palabras clave: juicios de probabilidad, calibración, riesgo, diferencias individuales, juego patológico.

Abstract

In a real-world setting, two groups of subjects showing a different level of risk were compared in several measures of probability judgment accuracy. After answering a 20 true-false items exam, undergraduate students were asked to estimate the subjective probability that each item was true. Subjects were split into two groups of risk. Both achieved the very same grades, but the risky group gave a higher number of answers. The obtained data suggest that risk has an effect on

calibration and noisiness of the probability judgments. The more risky subjects were more poorly calibrated, i.e., the subjective probabilities they estimated deviated further from the actual proportion of true items. Therefore, it seems that those who risk more know less about how much they risk. Discussion focuses on implications of these findings for gambling behavior. Calibration and covariance graphs are reported.

Key words: probability judgments, calibration, risk, individual differences, gambling.

Los juicios de probabilidad son relevantes a la hora de tomar decisiones. Imaginemos, por ejemplo, a un individuo que se estuviese planteando el suscribir o no una póliza de seguros para su vivienda. Para decidirse por una u otra alternativa, sería conveniente que se informara de distintas cuestiones. Por un lado, debería valorar las *consecuencias* (ventajas e inconvenientes) de suscribir o no la póliza. Por otro lado, sería interesante que conociera las *probabilidades* de robos, inundaciones u otra serie de sucesos que dieran lugar a posibles pérdidas materiales en la vivienda. Si se tuviera la absoluta seguridad de que la probabilidad de esos sucesos es nula (quizás, en un mundo de ficción) lo aconsejable sería no suscribir la póliza. La mayoría de las decisiones, sin embargo, no se toman ni teniendo un conocimiento exacto ni una ignorancia total de las probabilidades de los sucesos. En general, nos enfrentamos a situaciones de incertidumbre relativa en las que podemos evaluar de algún modo las verosimilitudes de los distintos sucesos bien porque conozcamos alguna distribución de la probabilidad de éstos o bien porque tengamos alguna creencia sobre las mismas (Martínez Arias, 1991).

El procedimiento más extendido para poner de manifiesto las verosimilitudes de un suceso es el de pedir que se estime la probabilidad (a través de porcentajes) de que ocurra. En este trabajo, se analizan

varias medidas de la precisión de los juicios probabilísticos. De ellas, la calibración ha sido, con diferencia, el aspecto que más atención ha recibido en la literatura sobre el tema, y se refiere a la medida en que los juicios de probabilidad que estima un sujeto sobre un suceso coinciden con la proporción de veces que realmente ocurre.

Numerosas investigaciones han tratado de comprobar si existen diferencias individuales respecto a la calibración de probabilidades (pueden consultarse las siguientes revisiones: Keren, 1991; Lichtenstein, Fischhoff y Phillips, 1982; O'Connor, 1989). Las variables propias del sujeto que se han analizado en dichos estudios son: el sexo; el conocimiento sobre un tema; la experiencia en diferentes ámbitos; diferencias culturales; el estado de ánimo depresivo; y variables de personalidad (autoritarismo, conservadurismo, dogmatismo e intolerancia a la ambigüedad).

En el presente trabajo, se comparan sujetos con el mismo conocimiento de una asignatura (la misma nota), pero diferente nivel de riesgo al contestar las preguntas de un examen. ¿Por qué pensamos que el riesgo puede estar relacionado con la calibración de probabilidades? Algunas investigaciones demuestran que cuando los sujetos apuestan en juegos de azar, en los que se arriesgan a perder lo que han apostado, afloran multitud de sesgos cognitivos

(Bersabé, 1996; Bersabé y Martínez Arias, 1999; Bersabé y Martínez Arias, 2000; Gaboury y Ladouceur, 1989; Griffiths, 1994; Ladouceur y Gaboury, 1988; Wagenaar, 1988; Walker, 1992). Uno de ellos es la ilusión de control que se manifiesta cuando un jugador cree poder controlar el azar con sus destrezas (por ejemplo, eligiendo unos determinados números en la lotería, tirando los dados de una forma especial, etc.). Langer (1975) observó que, en esas situaciones, la expectativa de éxito personal era inapropiadamente mayor de lo que garantizaba la probabilidad objetiva. Así pues, cuando nos enfrentamos a una situación de incertidumbre en la que tenemos algo que perder, nos asaltan los sesgos cognitivos, y se produce una sobreconfianza. Por ello, como hipótesis de partida, suponemos que los más arriesgados, es decir, los que se exponen a perder o fallar más, sobreconfían en mayor medida, con lo cual también calibran peor. En cualquier caso, creemos conveniente analizar, no sólo la calibración, sino todos los componentes de la precisión en los juicios de probabilidad. Con ello, estaremos en mejor disposición de entender cómo están formados los juicios, y de qué modo pueden modificarse.

Método

Participantes

La muestra estaba compuesta por 218 alumnos que cursaban la asignatura de Psicometría en la Facultad de Psicología de la Universidad Complutense de Madrid. Las edades oscilaron entre 19 y 25 años (Media= 19,53; DT=1,07). El grupo de menor riesgo quedó formado por 22 hombres y 87 mujeres; y el de mayor riesgo por 25 hombres y 84 mujeres.

Material

Examen teórico de la asignatura de Psicometría compuesto por 20 ítems de dos alternativas (V-F). Hoja de instrucciones y cuestionario con los mismos 20 ítems del examen en los que se debía estimar la probabilidad de que ese ítem fuera verdadero y falso (información redundante). De los 20 ítems presentados, 11 eran verdaderos y 9 falsos. El examen teórico, junto con otro práctico, determinaba la calificación que obtenían los alumnos en la asignatura.

Procedimiento

En primer lugar, los sujetos realizaron en situación natural el examen teórico tipo test con 20 ítems de dos alternativas (V-F). Rodeaban con un círculo la opción que consideraban correcta, pudiendo dejar preguntas sin contestar. Una vez terminado el examen, se pidió a los alumnos que leyeran la hoja de instrucciones en la que se explicaba cómo debían rellenar el cuestionario. Se les presentaban las mismas 20 preguntas del examen que acababan de hacer, y tenían que estimar la probabilidad de que ese ítem fuera verdadero y falso. Esto se hacía a través de porcentajes. Por ejemplo, una respuesta de "verdadero=0% y falso=100%" indicaría que se estaba completamente seguro de la falsedad de ese ítem. Por el contrario, una respuesta de V=50% y F=50% significaría que eran igualmente probables las dos alternativas. En la hoja de instrucciones se aclaró, además, que la prueba era voluntaria, que no repercutiría en ningún sentido a la hora de valorar el examen, y que el propósito era el de estudiar la calidad de las pruebas tipo test. Con esto, se pretendía eliminar de algún modo el posible efecto de la «deseabilidad so-

cial» que pudiera hacer que los alumnos se quisieran mostrar más seguros en sus juicios de probabilidad de lo que realmente estaban.

En la mayoría de los trabajos sobre calibración, el método empleado difiere del de este estudio. En general, primero se presentan las preguntas que deben contestar en su totalidad. Se trata, pues, de una tarea de elección forzada. En segundo lugar, se pide que indiquen (también a través de porcentajes) el grado en que se estima que la respuesta dada en ese ítem es correcta. Por tanto, esas respuestas se encuentran entre el 50%-100% porque no es de esperar errar más que el mero azar. En nuestro trabajo, la primera tarea de contestar los ítems es de elección no forzada (pueden dejar ítems sin contestar), y la estimación que se pide es sobre que el ítem sea verdadero (que no correcta su contestación). Se hace hincapié en estas distinciones en el procedimiento por la repercusión que tienen a la hora de interpretar las medidas de precisión en los juicios. Los porcentajes estimados por los sujetos se transformaron en once categorías de probabilidades subjetivas: [0-0,05] [0,05-0,15] [0,15-0,25] [0,25-0,35] [0,35-0,45] [0,45-0,55] (0,55-0,65] (0,65-0,75] (0,75-0,85] (0,85-0,95] (0,95-1].

Para los propósitos de nuestra investigación, se debían formar dos grupos de sujetos según el nivel de riesgo que manifestaban. Como señala Yates (1990), el concepto general de riesgo se puede caracterizar por el enfrentarse a la posibilidad de perder o errar en algún sentido para conseguir un mayor beneficio. Por ello, se pensó que un alumno sería más arriesgado que otro si, ante el mismo conocimiento (la misma nota), contestaba más preguntas. El alumno que actúa así, lo hace supuesta-

mente para intentar acertar más preguntas a riesgo de equivocarse también más. Siguiendo esta idea, se formaron dos grupos de riesgo de la manera siguiente: se ordenaron los alumnos según la nota obtenida en los 20 ítems (aciertos menos errores). Se encontraron 16 notas diferentes. Los alumnos que obtenían la misma nota, se dividieron en dos grupos según el número de contestaciones. Cuando había un número impar de alumnos con la misma nota, se eliminó el central. De esta forma, los alumnos se repartieron uniformemente en dos grupos con idénticas notas (Media = 6,69; DT=3,29), pero distinto número de contestaciones (Media=13,61; DT=1,66 en el grupo de riesgo bajo, y Media=17,17; DT=1,41 en el de riesgo alto). Efectivamente, el número medio de contestaciones fue diferente en los dos grupos de riesgo ($t_{(216)} = -17,04$; $p < .001$).

Para asegurar que los grupos de riesgo estaban bien formados, se dibujó la figura 1 en la que se muestran los porcentajes de contestaciones para cada probabilidad subjetiva. Como se puede apreciar, los alumnos más arriesgados no necesitan estar tan seguros de que el ítem es verdadero o falso para contestarlo. En otras palabras, su «umbral de contestación» es menor.

Resultados

Se calcularon distintas medidas de precisión en los juicios probabilísticos para cada sujeto. Pasemos a comentar cada una de esas medidas, descritas en detalle por Yates (1990).

Precisión total

Puntuación de Probabilidad. La medida de precisión en los juicios probabilísti-

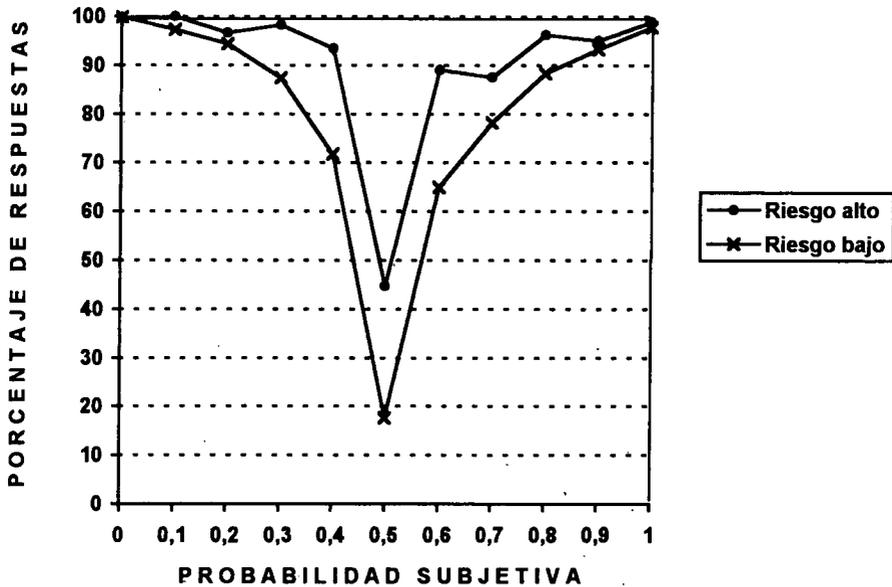


Figura 1. Porcentaje de respuestas según la probabilidad subjetiva estimada por el grupo de riesgo bajo y alto.

cos más ampliamente utilizada se atribuye a Brier (1950), y se conoce como la Puntuación de Brier, Puntuación Cuadrática, o Puntuación de Probabilidad media (\overline{PP}). Ésta se define como:

$$\overline{PP} = \frac{\sum (f_i - d_i)^2}{N} \quad (1)$$

En esta investigación,

f_i es la probabilidad que estima el sujeto de que el ítem i -ésimo sea verdadero
 $d_i=1$ si el ítem i -ésimo es verdadero
 $d_i=0$ si el ítem i -ésimo es falso

Por tanto, \overline{PP} es una función cuadrática que mide el grado de diferencia entre cada uno de los juicios de un sujeto, y lo que ocurre realmente. \overline{PP} se encuadra en el intervalo $[0,1]$. Cuanto más precisos son los juicios de una persona, menor será su \overline{PP} .

En este estudio, los sujetos menos arriesgados mostraron una \overline{PP} media algo menor que los conservadores, aunque la diferencia no resultó estadísticamente significativa (ver tabla 1).

Si un alumno fuera absolutamente incapaz de predecir si el ítem es verdadero o falso, podría estimar en todos los ítems una probabilidad subjetiva igual al 50%. Esa estrategia se denomina *juicio uniforme*. Se puede demostrar, siguiendo a Yates (1990), que a ese sujeto le correspondería una $\overline{PP} = 0,2$. Como indica la tabla 1, el porcentaje de sujetos que superaron el nivel de precisión del juicio uniforme también fue ligeramente superior en el grupo de riesgo bajo.

La precisión en los juicios de probabilidad no es un concepto indiferenciado sino que se puede desglosar en distintos componentes: calibración; discriminación y ruido (Murphy, 1973; Yates, 1982).

Tabla 1. Medidas de precisión de los juicios de probabilidad en cada grupo de riesgo.

Medida de precisión ^a	Riesgo bajo (N=109)		Riesgo alto (N=109)		t (216)	p-bilateral
	Media	(DT)	Media	(DT)		
<i>Total</i>						
$\overline{PP} \downarrow$	0,221	(0,062)	0,234	(0,064)	-1,527	0,128
$\overline{PP} < .25 \uparrow$ ^b	61,9%		54,8%		0,811	0,368
<i>Calibración</i>						
ICC \downarrow	0,084	(0,049)	0,098	(0,047)	-2,041	0,042
Sesgo 0	-0,103	(0,066)	-0,110	(0,068)	0,776	0,439
ICL \downarrow	0,015	(0,015)	0,017	(0,016)	-0,856	0,393
<i>Discriminación</i>						
RM \uparrow	0,112	(0,048)	0,110	(0,048)	0,371	0,711
Pend \uparrow	0,284	(0,139)	0,284	(0,161)	0,049	0,961
<i>Ruido</i>						
Disp \downarrow	0,074	(0,033)	0,085	(0,036)	-2,171	0,031

^a \downarrow : mejor a medida que aumenta; \uparrow : mejor a medida que disminuye; 0: cero es el mejor valor

^b Comparación de porcentajes a través del test Chi-cuadrado con la corrección de continuidad

Calibración

Se pueden distinguir tres índices de calibración. Analicemos cada uno de ellos:

Índice de Calibración por Categorías de juicios (ICC). Ya vimos cómo en la Puntuación de Probabilidad se comparaba cada juicio de probabilidad con lo que ocurría realmente en cada uno de los ítems (0: falso; 1: verdadero). En el ICC, en cambio, se toman todos los ítems donde se emitió el mismo juicio de probabilidad, y se compara ese juicio con la proporción de esos

ítems que realmente eran verdaderos. Por ejemplo, supongamos que un alumno ha estimado una probabilidad de 0,7 en 10 ítems. Con una calibración perfecta, 7 de esos 10 ítems serían realmente verdaderos, y 3 falsos. El ICC se calcula de la siguiente manera:

$$ICC = \frac{\sum N_j (f_j - \bar{d}_j)^2}{N} \quad (2)$$

donde,

j indica las diferentes categorías de probabilidades subjetivas (0; 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9 y 1);

N_j es el número de juicios registrados en la categoría j de probabilidad subjetiva; \bar{d}_j es la proporción real de ítems verdaderos, tomando aquellos en los que se estimó la categoría j de probabilidad subjetiva.

Tal como se desprende de la fórmula, cuanto menor sea la puntuación en el ICC, mejor será la calibración. Los datos revelan que el ICC es peor (mayor) en el grupo de mayor riesgo (tabla 1). Parece que las probabilidades subjetivas de los arriesgados se alejan más de la realidad. Por tanto, los que más se arriesgan menos saben cuánto se arriesgan.

La mayor parte de los trabajos de calibración aportan gráficos de calibración. Para dibujarlos, se deben reunir los juicios de todos los sujetos en cada grupo, es decir,

como si los alumnos se hubieran ido turnando para enjuiciar. Cada grupo (o «macro sujeto») obtuvo, entonces, una única puntuación en cada medida de precisión de los juicios. Esas puntuaciones son las que aparecen dentro de los gráficos.

En los gráficos de calibración, el ICC se representa como las desviaciones de la función respecto a la diagonal 1:1. Si un sujeto calibra perfectamente en todas las categorías de sus juicios, su curva de calibración coincidirá con la diagonal 1:1. En la figura 2, se muestran las curvas de calibración de cada grupo de riesgo. En ellas, es difícil apreciar las diferencias encontradas en el análisis de los datos individuales. Al juntar todos los juicios en cada grupo, las diferencias en el ICC parecen disolverse.

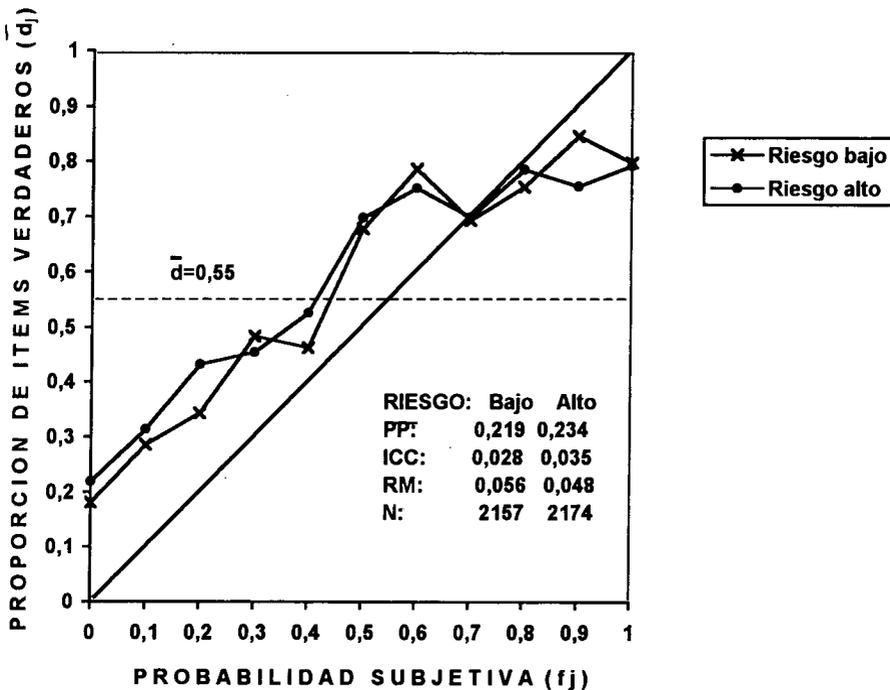


Figura 2. Curvas de calibración en el grupo de riesgo bajo y alto.

Un segundo tipo de calibración es la calibración a la larga. Si la calibración fuera perfecta, entonces, la media de todos sus juicios sobre un suceso (\bar{f}) debería coincidir con la proporción de veces que ocurre realmente (\bar{d}). La calibración a la larga se operativiza mediante el *Sesgo*:

$$\text{Sesgo} = \bar{f} - \bar{d} \quad (3)$$

o con su cuadrado, el *Índice de Calibración a la Larga* (ICL):

$$\text{ICL} = \text{Sesgo}^2 \quad (4)$$

Cuanto mayor es el valor absoluto del Sesgo, peor es el ICL. Con un único sujeto, el Sesgo y el ICL proporcionan una información redundante. Sin embargo, con un grupo de sujetos, pueden revelar diferentes cuestiones. Por ejemplo, si la mitad de los sujetos de un grupo obtuvieran un Sesgo de +0,15 y la otra mitad de -0,15, el Sesgo medio resultaría nulo. En cambio, los juicios de los sujetos individualmente aportarían un pobre ICL.

En los trabajos en los que se pide que los sujetos estimen la probabilidad (0,5-1) de que la respuesta dada sea correcta, el Sesgo es un buen indicador de la sobre/infraconfianza (Lichtenstein y Fischhoff, 1977). Un Sesgo positivo se da cuando se sobreconfía, es decir, cuando se estima haber acertado más de lo que realmente se acierta. Por el contrario, la infraconfianza se traduce en un Sesgo negativo. Sin embargo, en esta investigación, la tarea consiste en estimar la probabilidad de que el ítem, contestado o no, sea verdadero (no que la respuesta dada sea correcta). Así pues, un Sesgo negativo como el que aparece en los dos grupos de riesgo (véase la tabla 1) no indica infraconfianza, sino una

mayor seguridad en la falsedad que en la veracidad de los veinte ítems. Ese Sesgo negativo podría deberse a que hay más ítems verdaderos que falsos, con lo cual la tasa base (\bar{d}) es mayor de 0,5. También podría deberse a un sesgo de respuesta por el que se tuviera mayor confianza en las preguntas falsas que en las verdaderas. En cualquier caso, lo que sí parece claro es que el riesgo de los sujetos no se relaciona significativamente con el Sesgo, ni con ICL.

El Sesgo se representa en los gráficos de covariación (figuras 3a y 3b) mediante la intersección entre la línea vertical de la tasa base (\bar{d}) y la horizontal de la probabilidad subjetiva media (\bar{f}). Si esa intersección se sitúa en la misma diagonal 1:1, entonces el Sesgo será nulo, ya que la probabilidad subjetiva media coincidirá con la tasa base. Si la intersección queda por encima de la diagonal, el Sesgo será positivo. Si queda por debajo, como en los dos gráficos presentados en la figura 3 (a y b), el Sesgo será negativo. Esto confirma lo que ya se había obtenido numéricamente.

Discriminación

La calibración compete a la habilidad para indicar apropiadamente las distintas probabilidades de que se dé un suceso. A diferencia de esto, la discriminación se refiere a la tendencia a decir algo diferente de alguna forma en las ocasiones en que ocurre un suceso que en las que no se da. La medida en la que una colección de juicios alcanza el ideal de discriminación se refleja en el estadístico de *Resolución de Murphy* (RM):

$$\text{RM} = \frac{\sum N_j (\bar{d}_j - \bar{d})^2}{N} \quad (5)$$

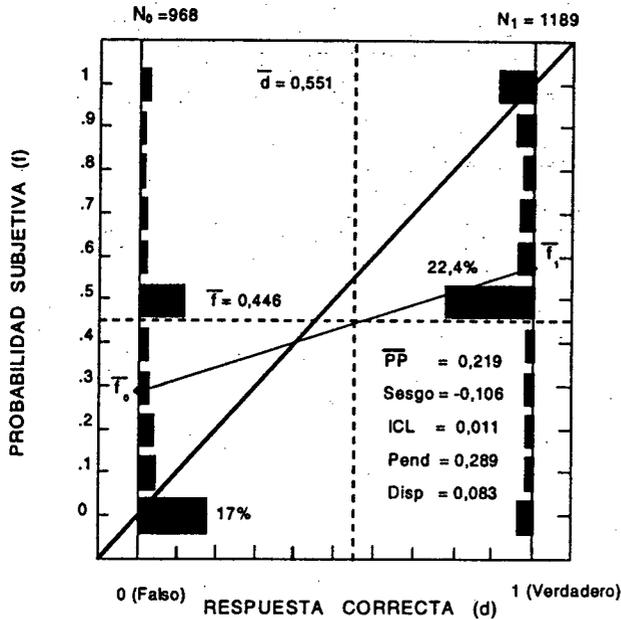


Figura 3a. Gráfico de covariación para el grupo de riesgo bajo.

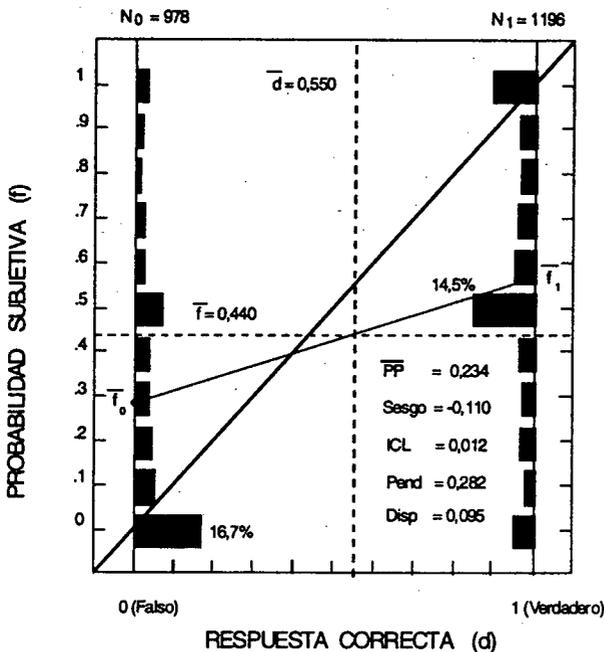


Figura 3b. Gráfico de covariación para el grupo de riesgo alto.

El estadístico RM resultó muy parecido en los dos grupos de riesgo (tabla 1). Esto es lógico si se piensa que los grupos se formaron igualando la nota (aciertos menos errores), que está directamente relacionada con la capacidad para discriminar entre los ítems verdaderos y falsos.

El estadístico RM también queda plasmado en los gráficos de calibración (figura 2). En ellos, la tasa base (\bar{d}) se dibuja en una línea horizontal. RM será pésimo cuando se solape esa línea horizontal con la curva de calibración. Ese caso se produciría cuando el sujeto emitiera las diferentes probabilidades subjetivas aleatoriamente, sin poder discriminar cuándo el ítem es verdadero o falso. Entonces, sería de esperar que a cualquier probabilidad subjetiva le correspondiera a la larga una proporción real de ítems verdaderos igual a la tasa base. En cambio, si alguien fuera capaz de discriminar perfectamente los ítems, estimaría probabilidades mayores de 0,5 cuando el ítem es verdadero, y menores de 0,5 cuando es falso. Por tanto, la curva de calibración para todas las probabilidades subjetivas mayores de 0,5 tendría una altura de 1 (todos esos ítems serían verdaderos), y una altura de 0 para las menores de 0,5 (todos esos ítems serían falsos). Por ello, RM se representa como las distancias entre cada punto (altura) de la curva de calibración, y la altura de la tasa base. En la figura 2, se advierte que las curvas de calibración de los dos grupos de riesgo se acercan por igual a la tasa base, esto es, presentan una discriminación parecida.

La *Pendiente* (Pend) es otra medida que pone en evidencia la capacidad para discriminar las veces en que se da un suceso de las que no ocurre. Si los juicios tienen una buena precisión en este sentido, enton-

ces la media de las probabilidades subjetivas que se den cuando el ítem sea verdadero (\bar{f}_1) tenderán a ser mayores que cuando sea falso (\bar{f}_0). Por esto, la *Pendiente* se calcula como:

$$Pend = \bar{f}_1 - \bar{f}_0 \quad (6)$$

Naturalmente, esta medida de discriminación entre ítems verdaderos y falsos también depende del conocimiento de la asignatura. Por esta razón, la *Pendiente* resultó prácticamente idéntica en los dos grupos de riesgo (ver tabla 1), ya que se igualaron en la nota.

En los gráficos de covariación, queda plasmada la *Pendiente* en la inclinación de la recta que une \bar{f}_0 con \bar{f}_1 . La discriminación máxima se corresponderá con la diagonal 1:1 (*Pendiente* = 1-0 = 1). Como se puede observar en los gráficos de covariación (figuras 3a y 3b), las dos *pendientes* tienen casi la misma inclinación.

Ruido

El último aspecto de la precisión de los juicios que se analizó fue el ruido que implica una forma especial de variación. Se calcula por las varianzas de los juicios condicionales a que el ítem sea verdadero ($Var(f_1)$) y falso ($Var(f_0)$). La medida de *Dispersión* (Disp) proporciona una información conjunta de dicha variabilidad, que no es más que una media ponderada de dichas varianzas condicionales:

$$Disp = \frac{N_1 Var(f_1) + N_0 Var(f_0)}{N_1 + N_0} \quad (7)$$

donde,

N_1 es el número de ítems verdaderos, y
 N_0 es el número de ítems falsos.

Los alumnos arriesgados obtuvieron un valor en la Dispersión significativamente mayor que los conservadores (ver tabla 1). La Dispersión se muestra en los gráficos de covariación. Estos gráficos no son más que dos diagramas de barras enfrentados: el de la izquierda para las probabilidades subjetivas cuando el ítem es falso; y el de la derecha para las probabilidades subjetivas cuando el ítem es verdadero. Existirá más Dispersión cuanto mayores sean las varianzas tanto en el gráfico de barras de la izquierda como en el de la derecha. En los gráficos de covariación, se puede apreciar que en el grupo de mayor riesgo (figura 3a) la variación de los juicios es algo mayor que en el de menor riesgo (figura 3b).

Discusión

Los alumnos se dividieron en dos grupos con idéntica nota pero con diferente nivel de riesgo. Los más arriesgados contestaban más preguntas ante la misma confianza subjetiva, exponiéndose a fallar más para intentar acertar más.

Tanto la nota como las medidas de discriminación entre ítems verdaderos y falsos reflejan de alguna manera el conocimiento que tiene el alumno de la asignatura. Por ello, era de esperar, como así resultó, que las medidas de discriminación fueran similares, ya que las notas eran iguales. En cualquier caso, lo que nos interesaba especialmente era contestar la pregunta que encabezaba el título de este artículo: ¿los que más se arriesgan, saben más cuánto se arriesgan?. De acuerdo con nuestros resultados, ocurre justo lo contrario. Los más arriesgados calibraron peor, es decir, las probabilidades subjetivas que estimaban se alejaban más de la proporción real

de ítems verdaderos. Por tanto, parece que los que más se arriesgan saben menos cuánto se arriesgan.

La diferencia en calibración entre los dos grupos de riesgo no se apreció claramente en los gráficos de calibración. En ellos, se deben reunir los juicios de probabilidad de todos los sujetos de cada grupo, por lo que cada grupo obtiene una única puntuación en cada medida de precisión. Como señala Björkman (1992) el análisis individual pone de manifiesto que las puntuaciones grupales no cuentan la historia completa. La calibración y resolución son atributos de los individuos con una considerable variación de sujeto a sujeto. Por esta razón, es más adecuado el análisis individual de los datos, aunque sea a través de las medias y desviaciones típicas de cada grupo.

Nuestros resultados conllevan implicaciones, no sólo para el estudio teórico de los juicios, sino también para entender la actitud de búsqueda de riesgo que caracteriza a los jugadores de azar. Así, la conducta claramente arriesgada de los jugadores que se exponen sistemáticamente a perder su dinero, puede que esté relacionada con una inadecuada calibración de sus juicios probabilísticos. Si esto se confirmara, sería interesante incorporar los métodos para corregir el Sesgo (Fischhoff, 1982) a las terapias cognitivas para el tratamiento del juego patológico. De hecho, algunos de estos métodos ya se han aplicado con relativo éxito como parte de una terapia más amplia de reestructuración cognitiva (Ladouceur, Sylvain, Letarte, Giroux y Jacques, 1998; Sylvain, Ladouceur y Boisvert, 1997).

Por último, queremos señalar un aspecto que merece especial atención en el estudio del riesgo: la naturalidad de la si-

tuación en la que se aplican las pruebas. En nuestra investigación, la primera tarea de contestar un examen real permitió manifestar el riesgo en los sujetos. Y no sólo porque se trataba de una tarea de elección no forzada sino, lo que es más importante, porque las posibles consecuencias de contestar o no a las preguntas, aprobar o suspender, resultaban relevantes para el alumno. Muchas de las investigaciones sobre toma de decisiones bajo riesgo emplean situaciones imaginarias en las que el sujeto debe decidir suponiendo que se enfrenta ante dos o más alternativas. Estas se describen a través del conjunto de las posibles consecuencias y de sus probabilidades. El riesgo de los sujetos ante una misma tarea puede variar respecto a la situación natural o de laboratorio, por lo que resulta crucial cuidar la validez ecológica en este tipo de estudios.

Referencias

- Bersabé, R. (1996). *Sesgos cognitivos en los juegos de azar: La ilusión de control*. Tesis doctoral publicada. Universidad Complutense de Madrid.
- Bersabé, R., y Martínez Arias, R. (1999). La superstición en los juegos de azar. *Clínica y Salud*, 10, 151-167.
- Bersabé, R., y Martínez Arias, R. (2000). Superstition in gambling. *Psychology in Spain*, 4, 28-34.
- Björkman, M. (1992). Knowledge, Calibration, and Resolution: A linear Model. *Organizational Behavior and Human Decision Processes*, 51, 1-21.
- Fischhoff, B. (1982). Debiasing. En D. Kahneman, P. Slovic, y A. Tversky (Eds.), *Judgment under uncertainty: Heuristic and biases* (pp. 422-444). Nueva York: Cambridge University Press.
- Gaboury, A., y Ladouceur, R. (1989). Erroneous perceptions and gambling. *Journal of Social Behaviour and Personality*, 4, 411-420.
- Griffiths, M. D. (1994). The role of cognitive bias and skill in fruit machine gambling. *British Journal of Psychology*, 85, 351-369.
- Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica*, 77, 217-273.
- Ladouceur, R., y Gaboury, A. (1988). Effects of limited and unlimited stakes on gambling behavior. *Journal of Gambling Behavior*, 4, 119-126.
- Ladouceur, R., Sylvain, C., Letarte, H., Giroux, I, y Jacques, C. (1998). Cognitive treatment of pathological gamblers. *Behaviour Research and Therapy*, 36, 1111-1119.
- Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, 32, 311- 328.
- Langer, E. J. (1975). *The psychology of control*. Beverly Hills, CA: Sage.
- Lichtenstein, S., Fischhoff, B., y Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. En D. Kahneman, P. Slovic, y A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306-334). Nueva York: Cambridge University Press.
- Martínez Arias, R. (1991). El proceso de toma de decisiones. En R. Martínez Arias y M. Yela (Eds.), *Pensamiento e Inteligencia* (pp. 411-494). Madrid: Alhambra Universidad.
- Murphy, A. H. (1983). A new vector partition of the probability score. *Journal of Applied Meteorology*, 12, 595-600.
- O'Connor, M. J. (1989). Models of human behavior and confidence in judgment:

- a review. *International Journal of Forecasting*, 5, 159-169.
- Sylvain, C., Ladouceur, R., y Boisvert, J. M. (1997). Cognitive and behavioral treatment of pathological gambling: A controlled study. *Journal of Consulting and Clinical Psychology*, 65, 727-732.
- Wagenaar, W. A. (1988). *Paradoxes of Gambling Behaviour*. Hove: Lawrence Erlbaum.
- Walker, M. B. (1992). *The Psychology of Gambling*. Oxford: Pergamon.
- Yates, J. F. (1982). External correspondence: Decompositions of the mean probability score. *Organizational Behavior and Human Performance*, 30, 132-156.
- Yates, J. F. (1990). *Judgment and Decision Making*. Englewood Cliffs, Nueva Jersey: Prentice Hall.