

TRABAJO EFECTUADO POR:

JOSÉ MANUEL DE HARO GARCÍA

Psicólogo Industrial.

Consultor de Recursos Humanos.

ACCÉSIT PREMIO ESTUDIOS FINANCIEROS 1995

Sumario:

Reseña.

- I. La utilidad: introducción.
- II. Conceptos básicos.
- III. Factores determinantes en el cálculo de la utilidad.
 1. La fiabilidad del criterio.
 2. La pertinencia del criterio.
 3. La fiabilidad del predictor.

4. La validez del predictor.
5. La razón de selección.
6. Porcentaje de empleados actuales que tienen éxito.

IV. Desarrollos del concepto de utilidad y situación actual.

1. La función resultado (*payoff*): conceptualizaciones.
 - 1.1. El concepto de validez.
 - 1.2. El ratio de éxito.
 - 1.3. El criterio estandarizado.
 - 1.4. El criterio en unidades monetarias.
2. La medición de la desviación típica del rendimiento en unidades monetarias (*SDy*).
 - 2.1. Las técnicas de medida.
 - 2.2. Los efectos del ambiente.
 - 2.3. Los efectos de las escalas de resultados.
 - 2.4. Los efectos de la población objetivo.

V. Ejemplo de cálculo de la utilidad para medir el impacto económico de un programa de selección.

VI. Conclusiones.

Bibliografía.

RESEÑA

En el presente trabajo se revisa la importancia que tiene el concepto de utilidad como descriptor del valor económico de los programas de selección de personal. Esta revisión se realiza examinando los aspectos básicos del concepto de utilidad, las variables determinantes de su magnitud y los desarrollos de la problemática y situación actual. Tras este análisis teórico, se desarrolla un ejemplo de cálculo de la utilidad en pesetas de un programa real de selección, aplicado recientemente. Los resultados evidencian la importancia, sencillez y magnitud del concepto bajo estudio, al mismo tiempo que nos hace reflexionar sobre el total desconocimiento que todavía se tiene de él en el ámbito empresarial español.

I. LA UTILIDAD: INTRODUCCIÓN

El problema de la utilidad se enmarca dentro de otro problema más amplio que afecta a los especialistas en psicología del trabajo y de las organizaciones. Este problema se refiere a la cuantificación de las contribuciones de tales expertos en dirección de recursos humanos, al éxito u obtención de beneficios de la organización a la que prestan sus servicios. La cuestión es si su contribución es realmente intangible comparada con otras funciones de dirección como finanzas, marketing o ingeniería, o si se puede medir de alguna manera. Dicho en otros términos: ¿cuál es el valor real de los programas de gestión de los recursos humanos en la mejora de la productividad organizacional?, ¿contribuyen sus trabajos a conseguir más beneficios para las empresas?

Los conceptos de utilidad en la psicología del personal pueden ser la única posibilidad de contacto entre los directivos y psicólogos en cuestiones como selección, formación o promoción. Los directivos suelen evaluar las acciones de las compañías en términos económicos, mientras que los psicólogos ilustran la eficacia de sus instrumentos en términos de validez. Si los dos grupos necesitan intercambiar información para tomar decisiones, es necesario un lenguaje común (SCHULER y GULDIN, 1991). El concepto de utilidad puede ser considerado como tal lenguaje común, ya que contiene conceptos psicológicos como la validez o puntuaciones típicas y factores económicos como la desviación típica del rendimiento en pesetas, transformando así, las cantidades psicológicas en medidas económicas.

El análisis de la utilidad ofrece un método para entender mejor el rol de los psicólogos del trabajo y las organizaciones en la mejora de las decisiones de recursos humanos y actuación organizacional. Mediante este análisis se puede describir, predecir y explicar el provecho o deseabilidad de las alternativas de decisión, analizando cómo puede ser utilizada la información en la toma de decisiones (BOUDREAU, 1989). El término utilidad ha sido asociado con un conjunto específico de modelos que reflejan las *consecuencias* (como ventas, beneficios netos o reducción de costos) de ciertos *programas* (selección, reclutamiento, formación, evaluación, establecimiento de metas, compensación o promoción interna) sobre el incremento de la productividad de la fuerza de trabajo.

A continuación vamos a presentar una revisión de los aspectos más relevantes del análisis de la utilidad en selección de personal. Para ello, comenzaremos examinando sus conceptos básicos para establecer un marco de trabajo, examinaremos los factores que influyen en la utilidad y repasaremos la situación actual e implicaciones para la investigación futura.

II. CONCEPTOS BÁSICOS

Los modelos de análisis de la utilidad son *instrumentos de ayuda* en las decisiones (EDWARD, 1977; EINHORN & McCOACH, 1977; EINHORN, KLEINMUNTZ, 1979; FISHER, 1976; HUBER, 1980; KEENEY & RAIFFA, 1976), instrumentos para la descripción, análisis, predicción y explicación de las decisiones. Esta ayuda proporciona un marco de trabajo consistente y estructurado con el que comparar las decisiones óptimas. Los modelos de utilidad en selección ofrecen este marco para considerar las consecuencias de la selección y para comunicar esas consecuencias en términos económicos.

La aplicación de los modelos de utilidad en selección requiere:

1. Un conjunto de *alternativas de decisión* sobre las que elegir (por ejemplo dos o más sistemas de selección).

2. Un conjunto de *atributos* que reflejen las características de las opciones que afectan a los resultados (por ejemplo validez, costos, efectos de tamaño y cantidad de empleados afectados), combinados con una escala de utilidad que refleje los valores de cada nivel del atributo para cada decisión (por ejemplo una escala que transforme los coeficientes de validez en valores de actuación en pesetas); y
3. Una función *resultado*, que permita obtener la utilidad total para cada opción mediante reglas que combinen los atributos entre ellos (por ejemplo una fórmula algebraica que describa la relación entre validez, costo y utilidad).

Considerar los modelos de análisis de la utilidad como instrumentos de decisión es bastante consistente con el desarrollo histórico que se discutirá más adelante.

Otra característica básica en el análisis de la utilidad en selección es que la unidad de análisis la constituyen los programas de mejora de la productividad de los recursos humanos. Tales programas son combinaciones de actividades que afectan al valor organizacional de la fuerza de trabajo.

El análisis de la utilidad es idóneo, sobre todo, para programas que van a ser aplicados a muchos individuos durante algún tiempo (CRONBACH y GLESER, 1965). En selección, la utilidad se centra en programas que usan tests para seleccionar nuevos empleados. Tales programas comprenden reglas que indican cómo debe ser aplicado el test y cómo deben ser usadas las puntuaciones en la elección de los aspirantes. Una característica de estos programas es que afectan a muchas decisiones de contratación sobre individuos. Los programas de selección funcionan proporcionando información que permite aceptar a unos candidatos y rechazar a otros, de modo que se obtenga un grupo más productivo que si no se aplicaran los tests. La decisión sobre cada candidato no es sin embargo el foco de atención del análisis de la utilidad, sino la decisión sobre si se implanta o no un programa que modificará la manera habitual de realizar las contrataciones. El impacto de un programa de decisión sobre la futura fuerza de trabajo será bastante grande, ya que las decisiones que implica el programa afectan a muchos individuos a través de su permanencia en la organización.

Los programas de selección funcionan si incrementan la exactitud de las elecciones en la forma en que son suficientemente importantes para contrarrestar los costos producidos. Sin embargo, cada organización puede definir la importancia de modo diferente, dependiendo de los elementos involucrados, la manera de medir la productividad y el modo en que la selección interactúa con otros programas como compensación y formación. La investigación ha procedido como si tales factores fueran constantes, aunque la investigación futura más integradora debería examinar estas relaciones más explícitamente.

Por último, los modelos de selección pueden ser expresados en términos de tres atributos básicos (BOUDREAU, 1984, 1987a, 1987b; BOUDREAU y BERGER, 1985a). Éstos son:

1. La cantidad, que refleja el número de empleados y períodos de tiempo afectados por las consecuencias del programa.

2. La calidad, que refleja las consecuencias (por persona, por período de tiempo) asociadas con el programa; y
3. El coste, que refleja los recursos requeridos para desarrollar y mantener el programa.

El resultado de un programa de selección puede ser calculado como el producto de la cantidad y la calidad menos el costo. El programa que obtenga las mayores diferencias positivas será el preferido.

Los modelos de utilidad en selección difieren en el modo en que definen cada una de las tres variables, pero todos ellos pueden ser entendidos dentro del marco de trabajo definido por todas las características básicas que acabamos de describir.

III. FACTORES DETERMINANTES EN EL CÁLCULO DE LA UTILIDAD

Todas las decisiones de selección suponen no sólo un coste económico (el precio pagado por reunir y sintetizar la información), sino también un balance de los efectos de las decisiones adoptadas (LEVY-LEVOYER, 1992). Este balance, conocido como utilidad, es crucial para determinar el uso de los diferentes procedimientos de toma de decisiones. Para poder establecer dicho balance, es necesario contar con una serie de datos precisos sobre cada método: validez, número de personas afectadas, coste e importancia que tiene para la organización la calidad de las decisiones (en términos financieros). Es decir, no es suficiente con conocer la validez.

La utilidad de un predictor puede definirse como la mejora conseguida en la calidad de las contrataciones realizadas (calidad en las personas seleccionadas), al emplear un dispositivo determinado de selección en comparación con la que se hubiera obtenido de no usarlo (BLUM y NAYLOR, 1990).

Esa calidad conseguida por el predictor está determinada por los siguientes factores:

1. La fiabilidad del criterio.

Quiere decir que la medida del rendimiento con la que se valida el predictor debe ser estable. Si no fuera estable, no se podría sacar ninguna conclusión exacta respecto al rendimiento predicho para los empleados, porque sería demasiado variable. Para que se considere adecuada, debe ser superior a 0,70 (MUCHINSKY, 1994).

2. La pertinencia del criterio.

Es una característica análoga a la validez del predictor. El criterio seleccionado debe ser una medida apropiada (válida) para poder representar al criterio conceptual de rendimiento laboral. Si un criterio es pertinente, es equivalente o coincidente con el criterio real. El criterio debe ser un indi-

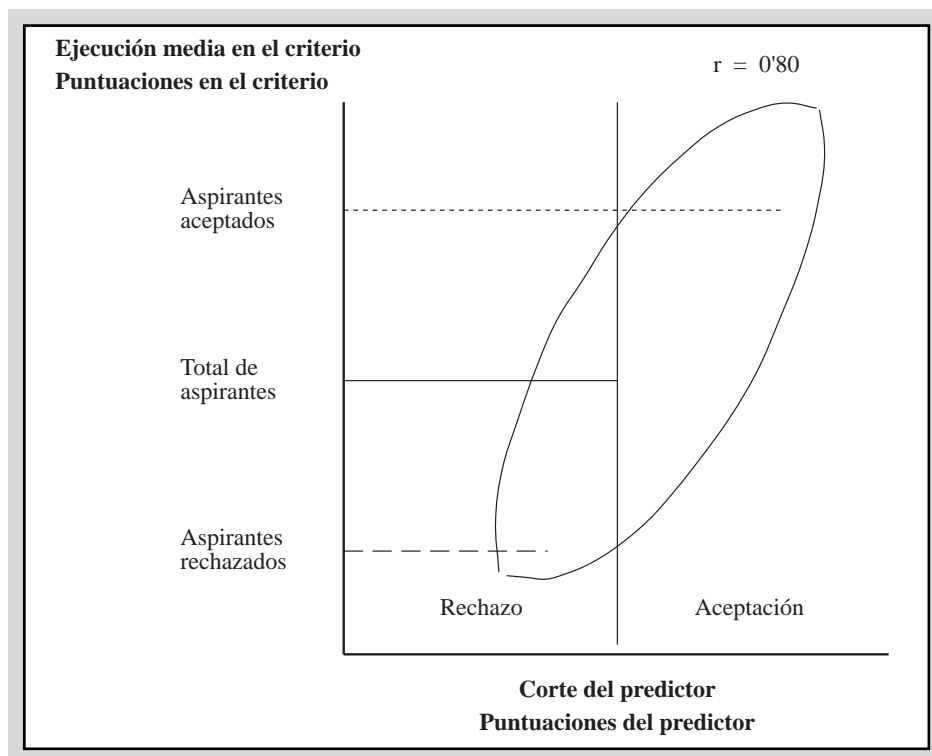
cador de calidad acertado, ya que si careciera de pertinencia, sería muy poco probable que cualquier predictor pudiera proporcionar personas cualificadas. Es difícil cuantificar la pertinencia del criterio, no obstante es necesario contar con el criterio pertinente con el que validar el predictor.

3. La fiabilidad del predictor.

Para que un predictor tenga algún valor debe ser fiable, aunque una fiabilidad alta no garantiza que el predictor sea útil. Esto ocurre cuando tiene una alta fiabilidad, pero una baja validez. En cambio, cuando un test tiene una alta validez, indica que también tiene una alta fiabilidad. El límite superior estimado para la validez de un test es la raíz cuadrada de su fiabilidad; es decir, un test con una fiabilidad de 0'81 tendría una validez superior límite de 0'90. Un test que no tuviera fiabilidad, tampoco tendría ninguna validez. De este modo, la fiabilidad se convierte en una condición necesaria pero no suficiente para la validez del predictor.

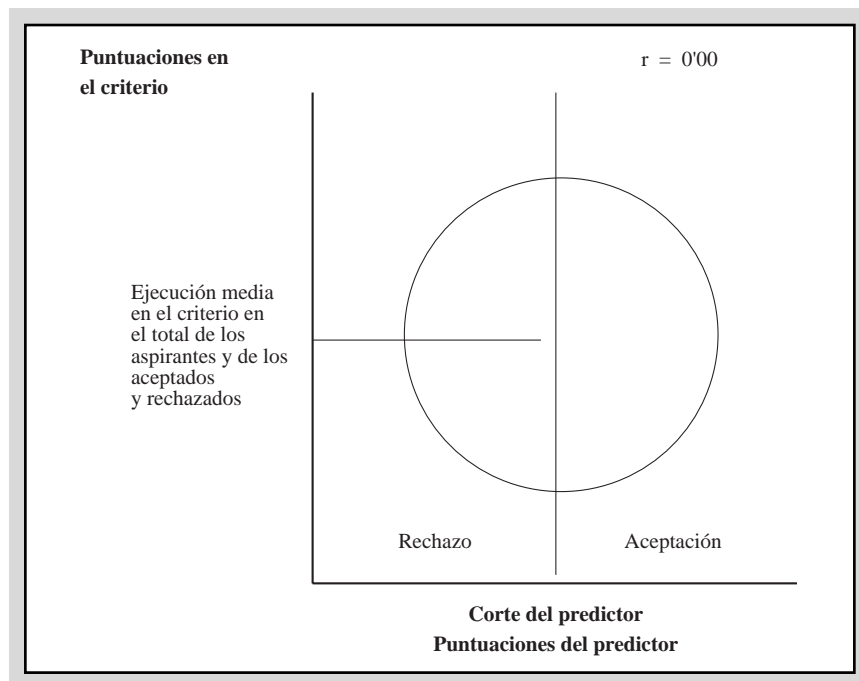
4. La validez del predictor.

La validez del predictor tiene un efecto más directo y visible que los factores anteriores, influyendo en la utilidad más que ningún otro factor. Para ilustrar su importancia, observemos la figura siguiente:



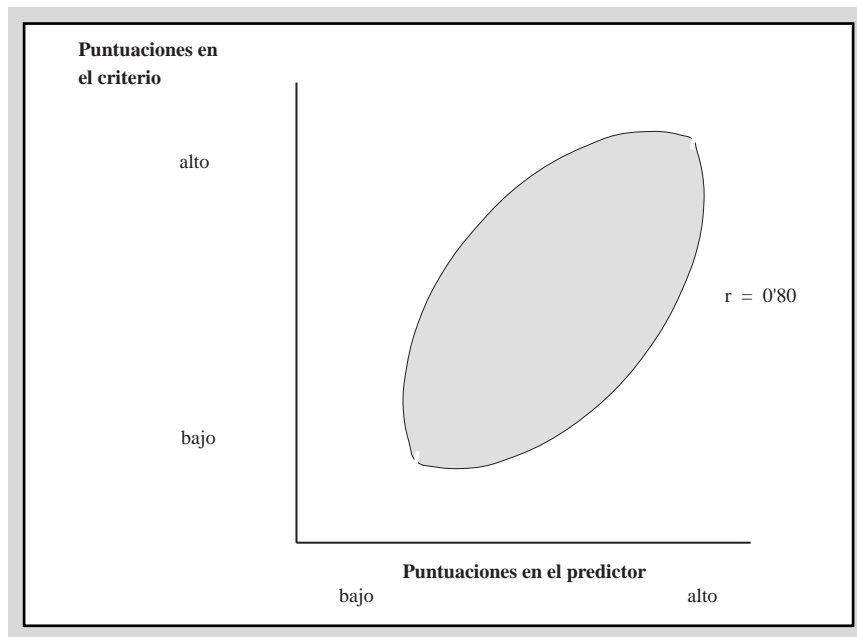
La correlación entre predictor y criterio es de 0'80. Sobre el eje horizontal, que refleja las puntuaciones del predictor, el corte del predictor separa a los aspirantes aceptados de los rechazados. La línea horizontal sólida representa el criterio de ejecución de todo el grupo, y divide por la mitad la distribución global de las puntuaciones. La línea horizontal punteada representa el criterio de ejecución del grupo rechazado, estando por debajo de la ejecución del grupo rechazado. La línea horizontal discontinua representa el criterio de ejecución del grupo aceptado, estando por encima de la ejecución del grupo en su conjunto. Lo que hace que el predictor sea válido, es que identifica a las personas más capaces de todas las que componen el grupo, es decir, las personas que se espera que rindan mejor, las sitúa por encima del punto de corte del predictor y las personas que se espera rindan peor, las sitúa por debajo del punto de corte del predictor.

En el siguiente gráfico se representa una situación en la que un predictor no correlaciona con el criterio.



La línea de corte del predictor separa a los aceptados de los rechazados. En este caso, las tres líneas horizontales están superpuestas, es decir, la ejecución en el criterio del grupo aceptado no es mejor que la del grupo rechazado, coincidiendo en ambos casos con la ejecución total del grupo. Como la utilidad se mide por la diferencia entre la ejecución media del grupo aceptado y del grupo rechazado, y esta diferencia es cero, podemos concluir que cuando la validez es cero, la utilidad también. Por tanto, hay una relación directa entre validez y utilidad. Cuanto mayor es la validez del predictor, mayor es la utilidad, midiéndose ésta como el incremento en la ejecución media del criterio para el grupo aceptado sobre la media del grupo total. De todas formas, hay métodos de validez relativamente baja que podrían tener cierta utilidad, en función de sus valores en los otros factores que nos quedan por ver.

El coeficiente de validez está influido por la fiabilidad del predictor y del criterio, como hemos apuntado más arriba. Esta relación puede expresarse del siguiente modo:



Donde:

r_{pc} = validez obtenida predictor-criterio.

r_{pcr} = validez real predictor-criterio.

r_{pp} = fiabilidad del predictor.

r_{cc} = fiabilidad del criterio.

Sólo cuando r_{pp} y r_{cc} son la unidad, la validez real es igual a la obtenida. Al disminuir la fiabilidad de las dos medidas, disminuye también la validez. Del mismo modo, si la fiabilidad del predictor o criterio fuese cero, la validez sería también cero.

5. La razón de selección.

Se define como el número de ofertas de empleo o vacantes dividido entre el número de solicitantes:

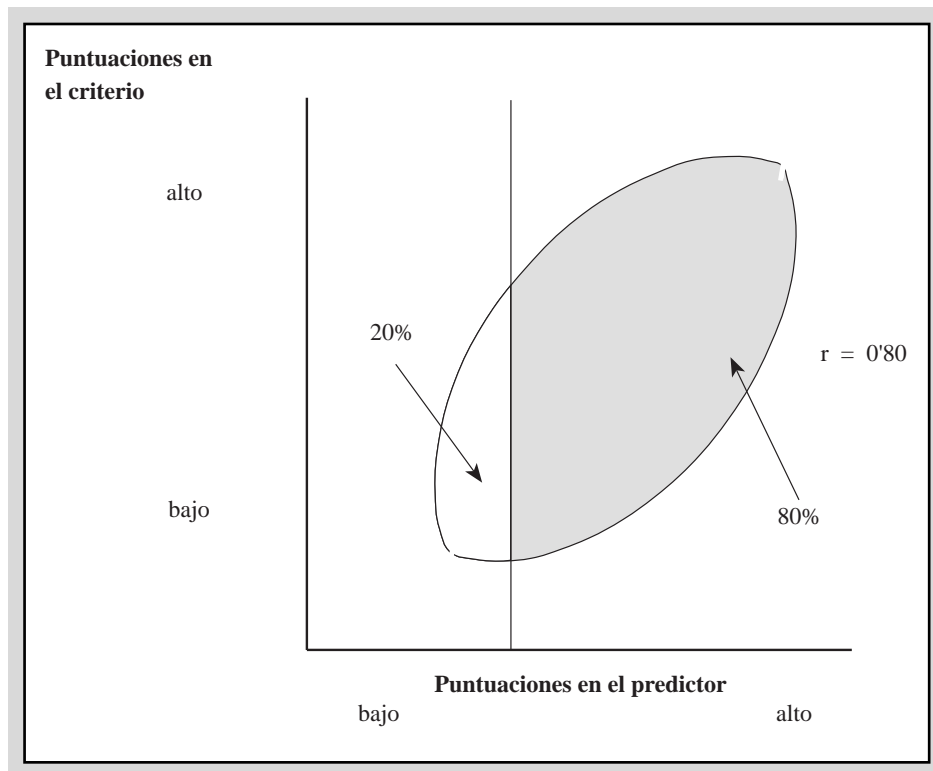
$$RS = \frac{n}{N}$$

Por ejemplo, si se examinan 1.000 solicitudes para seleccionar a tres candidatos para ser responsables de un proyecto, el método de selección va a ser más útil que cuando hay que aceptar al 90 por 100 de los candidatos que se presenten a realizar un curso de ventas. Es decir, cuando la razón de selección es baja (hay un gran número de candidatos por puesto), el procedimiento es más útil que cuando hay una alta razón de selección (si hay pocos candidatos en relación al número de puestos). Cuando la *RS* es igual a 1 (hay tantas ofertas como solicitantes), el uso de cualquier dispositivo de selección tendría poco sentido. Si es mayor que 1, la empresa aceptaría a cualquier candidato, cualquiera que sea su calidad. Cuando *RS* es inferior a 1, se justifica con sentido el uso de técnicas de selección.

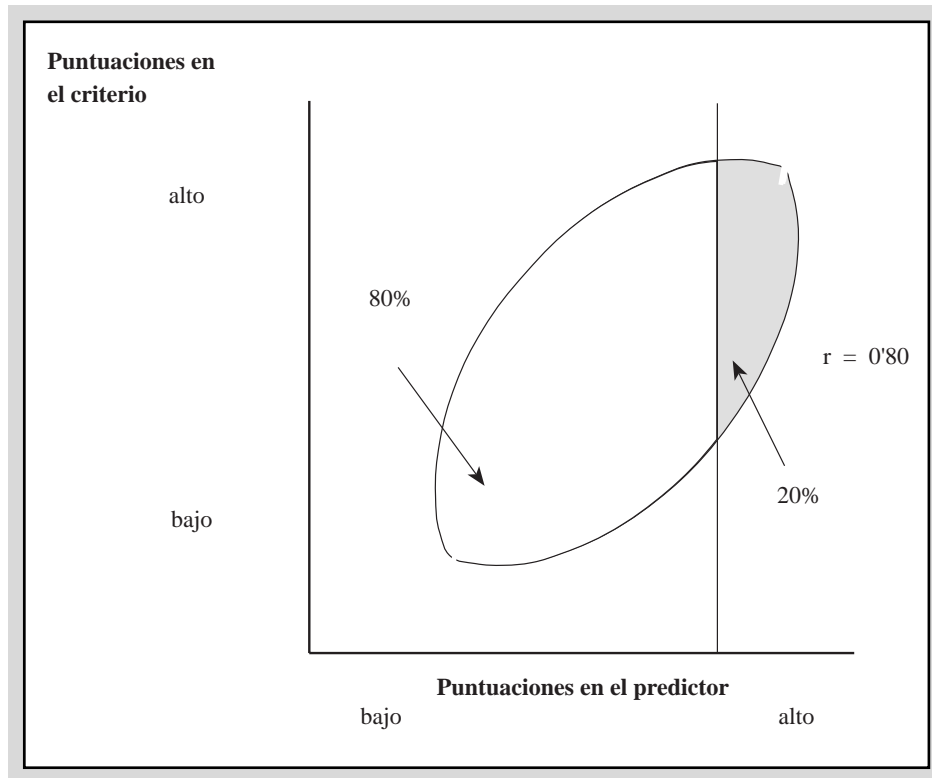
En las siguientes figuras vamos a ver la relación entre la *RS* y la utilidad del predictor.

En esta primera figura vemos las puntuaciones de una muestra grande de personas, con una correlación de 0'80 entre el predictor y el criterio. La superficie rayada representa la proporción de solicitantes a los que se contrata realmente, es decir, la *RS*. En este caso, la *RS* es de 1'00, lo que indica que hay una vacante para cada uno de los aspirantes y por tanto se contrata a todos.

En la siguiente figura, la *RS* se convierte en 0'80.



Sólo hay empleos para el 80 por 100 de los solicitantes, por lo que se selecciona al 80 por 100 de los candidatos que obtengan las mejores puntuaciones en el predictor. Ese 80 por 100 representa la zona subrayada del óvalo, que está a la derecha del punto de corte del predictor. Como las puntuaciones de los rechazados son más bajas, la puntuación promedio en el criterio para los contratados con esta *RS* de 0'80 es más alta que en el caso de la primera gráfica. Este aumento de la calidad del promedio, y por tanto de la utilidad, se aprecia mejor todavía si utilizamos una *RS* de 0'20.



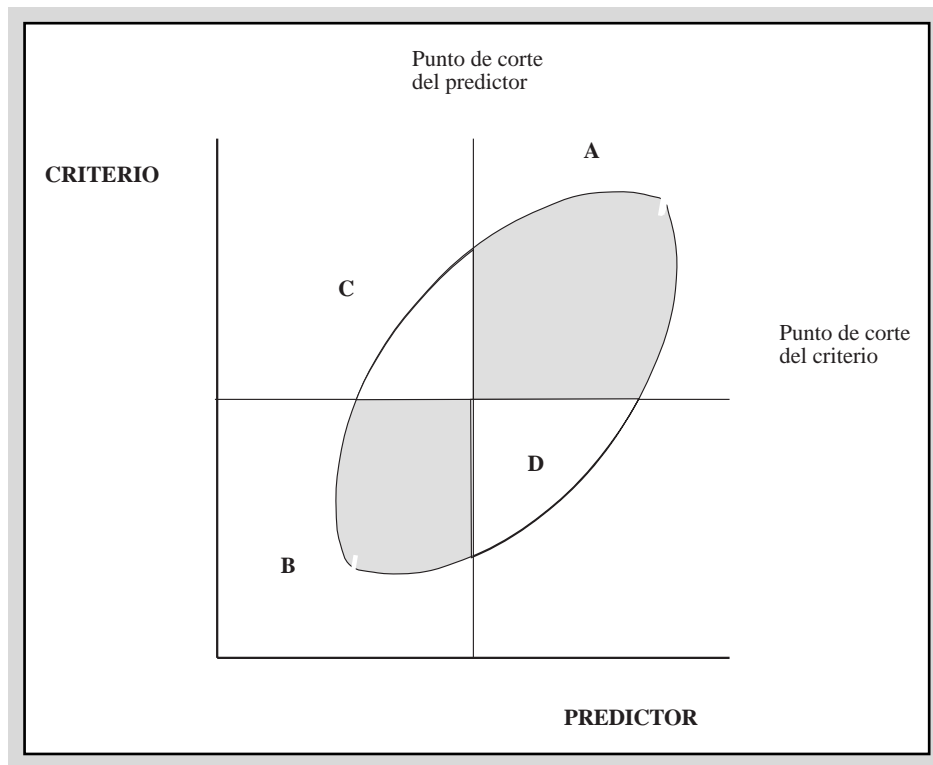
En esta situación, el empleador podrá seleccionar al 20 por 100 de los sujetos que obtengan las mejores puntuaciones en el predictor. Estos individuos se encuentran en la zona subrayada a la derecha del punto de corte. Al haber sólo dos vacantes por cada diez solicitantes, la calidad promedio para este grupo selecto en comparación con el grupo total es bastante grande.

La relación entre la *RS* y la utilidad del predictor es clara: cuanto más pequeña sea la *RS*, mayor será la utilidad del predictor, siempre y cuando la validez sea superior a cero. Cuanto más exigentes seamos a la hora de contratar personas, mayor será la probabilidad de que los admitidos tengan la calidad que deseamos.

6. Porcentaje de empleados actuales que tienen éxito.

Este último factor, también denominado *proporción base*, se refiere a la tasa de éxito o de buen rendimiento de los empleados sin usar el método de selección que se esté estudiando. Dicho de otra manera, sería la proporción de empleados que tienen puntuaciones en el criterio por encima de algún valor considerado como mínimo para que pueda ser considerado eficiente. Esto quiere decir, que si todos los candidatos seleccionados son satisfactorios o si no hay diferencia entre el obrero mejor y peor, el método de evaluación puede ser válido, pero su utilidad será escasa. Al contrario, será muy apreciado tanto si existen fuertes diferencias entre las personas empleadas en la actualidad, como si en ausencia de una selección previa, se produjera una fuerte tasa de fracaso o un bajo rendimiento en los seleccionados.

Consideremos el siguiente gráfico para ilustrar su influencia sobre la utilidad:



Cuando se entrecruzan las dos líneas de corte (predictor y criterio), se forman cuatro porciones de los datos: A, B, C y D. En la parte A, se encuentran los candidatos que están por encima de las líneas de corte del predictor y del criterio (*positivos reales*). Estos candidatos son los que según las pruebas realizadas van a tener éxito y de hecho, tras la medición en el criterio, se comprueba que son eficientes. Son, por tanto, decisiones correctas.

En la parte B, encontramos a los solicitantes (*falsos reales*) que tienen calificaciones por debajo de los puntos de corte del predictor y del criterio. Son aquellos que según las pruebas van a fracasar en el puesto, y al comprobar su rendimiento en el criterio se comprueba que no tienen éxito. Son también decisiones correctas.

La zona C representa a los candidatos que según las puntuaciones del predictor no serían seleccionados, pero que si se comprueba su rendimiento en el criterio, tendrían éxito. A los sujetos de este error (individuos eficaces no seleccionados) se les denomina *falsos negativos*.

En la zona rayada D, vemos a aquellos solicitantes que se contratarían, pero que posteriormente no resultarían satisfactorios en su trabajo. Estas personas representan también equivocaciones, y se les conoce como *falsos positivos*.

Del estudio de las diversas partes de la figura, obtenemos las siguientes proporciones:

1. Razón entre el número de errores en selección y el número de empleados colocado correctamente: $(C + D) / (A + B)$.
2. Porcentaje de los actualmente eficientes: $(A + C) / (A + B + C + D)$.
3. Porcentaje de eficientes usando el predictor: $A / (A + D)$.

La primera de las proporciones depende de: la ubicación del punto de corte del criterio, ubicación del punto de corte del predictor y del coeficiente de validez. Lo que pretende el seleccionador es reducir al máximo los valores C y D.

El segundo porcentaje se refiere al grado de éxito que se obtiene, cualesquiera que sean los métodos usados de selección, antes de la inclusión del predictor cuya utilidad se esté estudiando.

La tercera razón es una expresión de la proporción de solicitantes contratados que tendrán éxito si se utiliza el predictor bajo estudio.

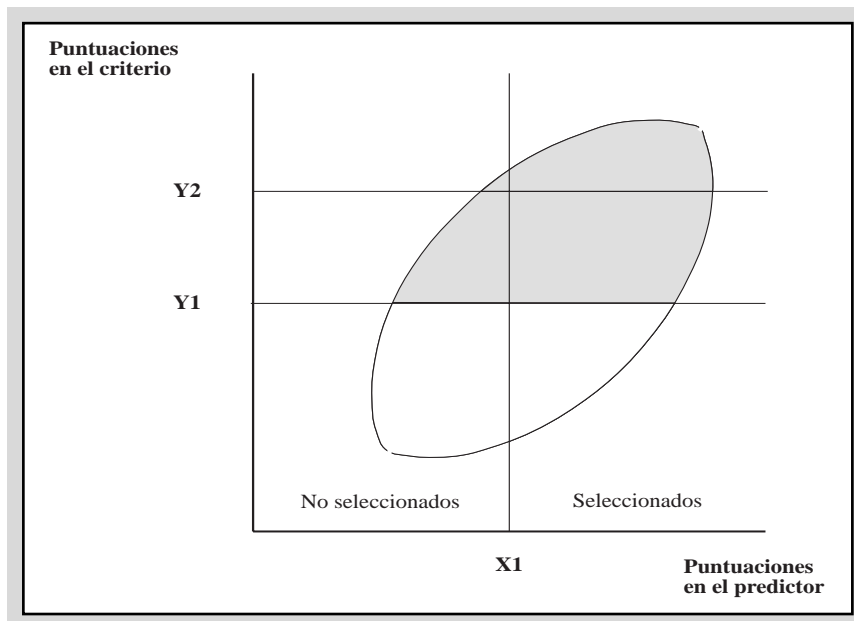
Si el valor de la tercera razón (porcentaje de empleados que tendrán éxito con dicho predictor) es mayor que la segunda (porcentaje de los que son realmente eficientes), el predictor estará aportando algo al proceso de selección. Si dicho valor es inferior, el predictor no será útil.

El método más sencillo para calcular esta utilidad consiste en pasar la prueba, sin servirse de los resultados, hasta obtener datos sobre el rendimiento de los incorporados al trabajo y utilizar un cuadro de doble entrada como el de arriba. Si tras calcular las razones segunda y tercera obtenemos respectivamente los valores de 50 por 100 y 80 por 100, significará que el porcentaje de buenas incorporaciones, ha pasado de un 50 por 100 (sin predictor) a un 80 por 100 (con predictor).

Si se comprueba la magnitud relativa de las razones segunda y tercera, podemos derivar los siguientes principios generales (BLUM y NAYLOR, 1990):

1. Sea cual sea la validez y el punto de corte del criterio, una reducción de la *RS* (razón de selección) dará como resultado un aumento de la validez efectiva. De este modo, puede compensarse la baja validez de una prueba siendo selectivo en las contrataciones.
2. Sea una validez y una *RS* (razón de selección) dada; cuanto menor sea el porcentaje de empleados actuales considerados satisfactorios, mayor será el aumento del porcentaje de los candidatos satisfactorios obtenidos al aplicar el predictor. Es decir, el porcentaje de aumento de la efectividad es igual a la diferencia entre el porcentaje de eficientes usando el predictor y el porcentaje de los actualmente eficientes. Por tanto, el mejor incremento, se dará en la situación en la que el rendimiento sea peor. Así, la posibilidad de obtener el mayor potencial en la ejecución del criterio se da cuando la proporción base es baja (MUCHINSKY, 1994). Por el contrario, si todas las personas en la compañía tienen un rendimiento satisfactorio (proporción base alta), el nuevo predictor no podría mejorar esa situación. En términos relativos, se puede concluir que cuanto más baja sea la proporción base, mayor será el incremento del porcentaje en cuanto a la ejecución satisfactoria de los empleados cuando se utilice un nuevo predictor. Pero en términos absolutos, el mayor aumento en la ejecución media en el criterio tendrá lugar cuando la proporción base sea 0'50. A medida que la proporción base se hace más extrema (alta o baja), decrece la utilidad absoluta del predictor. Con otras palabras, es difícil que un predictor cambie el número real de trabajadores que rinden bien si el grupo actual está rindiendo extremadamente bien o mal.

En resumen, cuanto más cerca esté la proporción base a 0'50, mayor será el aumento en el número real de nuevos empleados que rindan satisfactoriamente.



La ilustración anterior representa la relación entre la tasa de éxito en el puesto y la validez del predictor.

Si la validez del gráfico anterior fuera de 0'50 y la tasa de éxito en el empleo de 0'50 también, si optamos por rechazar a los candidatos cuya puntuación en el predictor sea inferior a X_1 , se disminuye la tasa de decisiones erróneas. Pero, al ser más exigente, y centrar el umbral de rechazo en Y_2 , no se producirá ninguna decisión errónea. Si al mismo tiempo, la tasa de éxito en el empleo sin utilizar ningún predictor es muy baja (Y_2 , que representa sólo al 20%), el número de falsos positivos aumenta y el número de falsos negativos disminuye para un mismo umbral X_1 , y ello sin tener en cuenta la validez del predictor. Hay pues, una relación entre la validez, tasa de éxito y umbral del predictor.

El análisis precedente muestra que el coeficiente de correlación por sí sólo no es indicador suficiente de la utilidad de un método de evaluación. La utilidad está en función del valor de r (coeficiente de correlación), pero corresponde a la mejora de los valores del criterio para el conjunto de la población considerada. La validez, por otro lado, se refiere a la capacidad del predictor para anunciar por adelantado qué candidatos serán buenos y cuáles no. La utilidad sólo refleja la capacidad del método para adoptar un máximo de decisiones correctas.

En cuanto al cálculo, existen desde hace algún tiempo tablas que permiten leer directamente el porcentaje de decisiones acertadas, a partir de la validez del predictor, la razón de selección y la tasa de éxito sin la intervención del predictor. Entre estas tablas, figuran las de TAYLOR y RUSSELL (1939), NAYLOR y SHINE (1965) y GHISELLI y BROWN (1955). Los primeros autores expresaron la relación entre el tamaño del coeficiente de validez, la razón de selección y la tasa de éxito de los empleados actuales. Sus tablas permiten determinar el porcentaje de contrataciones satisfactorias. Sin embargo, las tablas de NAYLOR y SHINE parecen tener ventajas sobre las de TAYLOR y RUSSELL. Así, por ejemplo, NAYLOR y SHINE formulan sus tablas de acuerdo con las diferencias en el puntaje promedio en el criterio de los grupos escogido y original. Este índice de utilidad es más significativo que el de TAYLOR y RUSSELL, quienes utilizan para su tabla, las diferencias en el porcentaje de éxito entre los grupos escogido y original. De igual modo, el uso de las tablas de TAYLOR-RUSSELL implica dividir a los empleados en dos grupos: «eficientes» e «ineficientes», tomando como referencia algún punto arbitrario en la dimensión criterio. El mismo hecho de pretender hacer entre los obreros una dicotomía «buenos-malos», suele ser una sobresimplificación de la realidad. Tanto la productividad como la calidad en el trabajo son variables continuas y es difícil trazar una barrera estricta entre las decisiones correctas e incorrectas. Las tablas de NAYLOR-SHINE no requieren ninguna decisión de este tipo para su uso, resultando más generales en su aplicabilidad. Para estos autores, la varianza del rendimiento es representada por la relación entre la producción del mejor y del peor empleado. GHISELLI y BROWN simplificaron la utilización de las tablas, presentándolas en forma de diagramas donde a partir de la tasa de selección, el valor se proyecta sobre un eje que indica la relación entre el mejor y el peor trabajador.

Veamos a continuación un fragmento de la tabla de NAYLOR y SHINE, para unos valores determinados, a modo de ejemplo.

Pi	Zxi	Li	Li/Pi
0'9987	-3'00	0'0044	0'00
0'9505	-1'65	0'1023	0'11
0'7517	-0'68	0'3166	0'42
0'5080	-0'02	0'3989	0'79
0'4090	0'23	0'3885	0'95
0'2514	0'67	0'3187	1'27
0'1056	1'25	0'1826	1'73
0'0559	1'59	0'1127	2'02

Para poder entender el uso de la tabla, es necesario conocer lo que significan las expresiones que vienen en la parte superior y otras que son relevantes.

Así:

Pi = la proporción de selección.

Zxi = el punto de corte del predictor en puntuaciones típicas.

Li = ordenada de la distribución normal en Zxi.

rx_y = coeficiente de validez, y

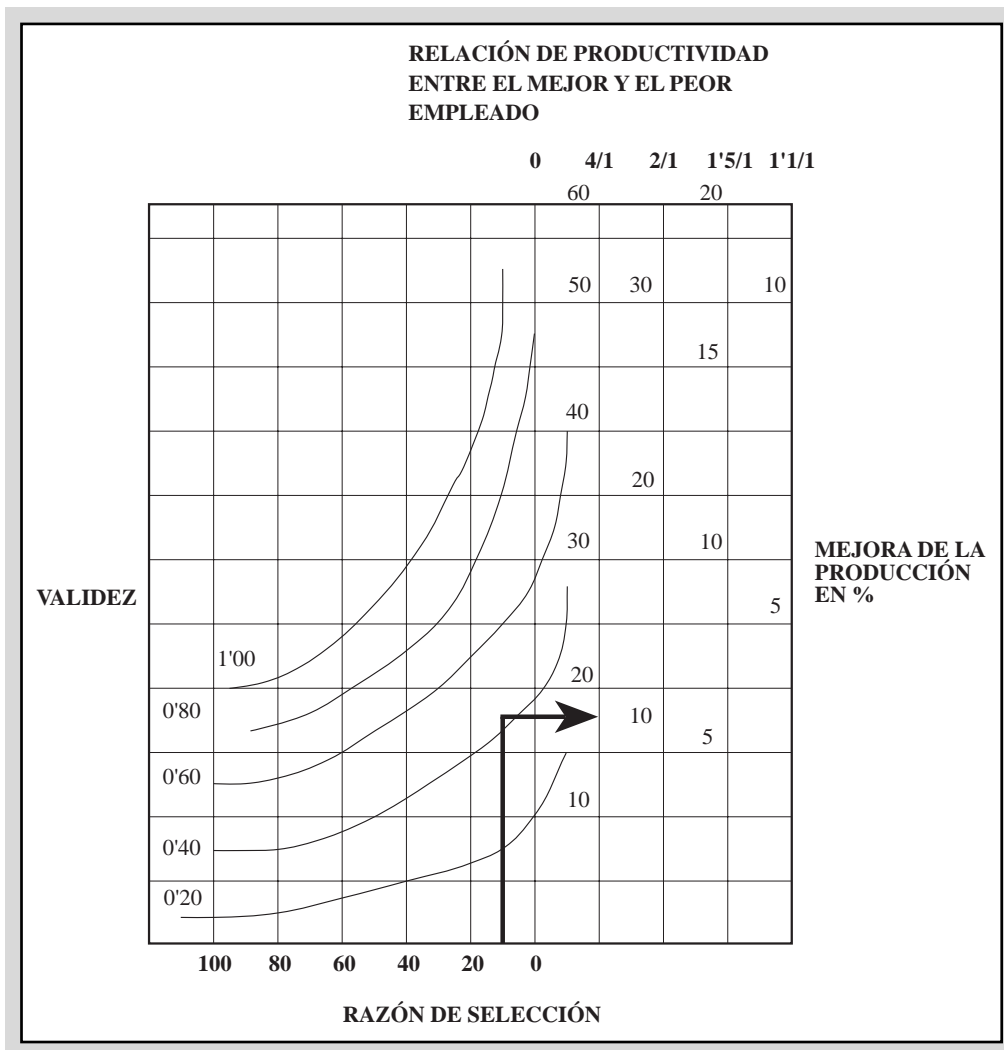
Z_{yi} = la media del criterio en puntuaciones típicas de todos los casos por encima del punto de corte.

La tabla se basa en la siguiente ecuación: $Z_{yi} = r_{xy} (L_i/P_i)$, y presupone la distribución normal de las variables.

La tabla puede utilizarse para calcular el incremento de la puntuación media en el criterio para un predictor o predictores con una relación determinada con el criterio, o para calcular la razón de selección necesaria para lograr una puntuación media determinada en el criterio. Así, por ejemplo, para una proporción de selección de 0'25 (se selecciona sólo al 25% de los candidatos) y una validez del predictor de 0'35, ¿cuál sería el nivel promedio de desempeño de los seleccionados? Utilizando la tabla, para el valor de $P_i = 0'25$, el cociente L_i/P_i sería igual a 1'27. Si aplicamos la fórmula $Z_{yi} = r_{xy} (L_i/P_i)$, tendremos que $Z_{yi} = (0'35) (1'27) = 0'44$. Esto significa, que la puntuación media en el criterio de los seleccionados, utilizando una razón de selección de 0'20, es de 0'44 unidades Z mejor que la media en el criterio de los no seleccionados.

En el caso de la regresión múltiple, para poder utilizar las tablas es necesario transformar el valor del predictor múltiple ($Z' y_i$) en el valor del punto de corte (Zx_i) mediante la fórmula $Zx_i = Z' y_i / R_{xy}$, donde R_{xy} es el coeficiente de correlación múltiple entre el predictor y criterio. Aunque hay tablas más detalladas, que tienen en cuenta los efectos de los tamaños y las varianzas desiguales de grupos (véase RORER, HOFFMAN, LA FORGE y HSIEH, 1966 y DARTINGTON y STAUFFER, 1966); debido a su facilidad de manejo y poco grado de elaboración la de NAYLOR y SHINE es bastante aconsejable. Sin embargo, no está exenta de críticas. Así, se basa en la suposición de que la relación entre el predictor y el criterio es lineal, además de obtener el coeficiente de validez por procedimientos de validez concurrente. Cuando la relación no es lineal la tabla es inapropiada para determinar la utilidad de una prueba.

Veamos también un ejemplo de la tabla-diagrama de GHISELLI y BROWN (1955):



La línea de trazo grueso indica la mejora en la productividad producida por una razón de selección del 10 por 100, una validez de 0'40, y una relación de productividad entre el mejor y el peor empleado de 4 a 1. El resultado sería de aproximadamente el 18 por 100 de mejora en la productividad al emplear ese predictor con respecto a no emplearlo.

La relación entre la validez y el ratio de selección, también puede verse en el siguiente cuadro (SMITH, 1986), que representa el ahorro en libras esterlinas (utilidad) para diferentes coeficientes de validez y ratios. En esta tabla se aprecia, que conforme aumenta el coeficiente de validez y disminuye el ratio de selección, aumenta la cantidad de dinero ahorrado, es decir, su utilidad.

COEFICIENTE DE VALIDEZ	RATIOS		
	1/3	1/6	1/10
0'1	39	45	61
0'2	77	91	122
0'3	116	137	183
0'4	154	182	245
0'5	192	228	306
0'6	231	273	368
0'7	270	368	429

Para concluir esta sección, resumiremos en un cuadro los efectos de las siete variables descritas sobre la utilidad de un predictor. De ellos, los que más peso tienen en el cálculo de la utilidad son los tres últimos (validez de predictor, razón de selección, tasa de éxito). También, son los que más fácilmente pueden ser manipulados por la organización.

Teniendo en cuenta por tanto, los siete factores, ¿cuál es la combinación que proporciona la mejor utilidad de un predictor? La tabla que sigue muestra los valores de estas variables para una utilidad alta (MUCHINSKY, 1994).

VARIABLE	VALOR PARA UNA UTILIDAD ALTA
Fiabilidad del criterio	Alta
Pertinencia del criterio	Alta
Fiabilidad del predictor	Alta
Validez del predictor	Alta
Razón de selección	Baja
Proporción de éxito	0'50
Coste	Bajo

IV. DESARROLLOS DEL CONCEPTO DE UTILIDAD Y SITUACIÓN ACTUAL

La historia del análisis de la utilidad en selección puede ser considerada bajo una perspectiva de toma de decisiones, centrándose en la definición de bondad de cada modelo y usando los conceptos de cantidad, calidad y costo para evaluar los valores de los modelos de decisión en cada desarrollo histórico.

La función resultado, expresión de la utilidad (*payoff*), ha sido definida históricamente de cuatro modos:

1. Basada en el coeficiente de validez.
2. Basada en el ratio de éxito.
3. Basada en el nivel de criterio estandarizado, y
4. En términos de niveles del criterio en dinero.

A continuación repasaremos estas cuatro perspectivas.

1. La función resultado (*payoff*): conceptualizaciones.

1.1. El concepto de validez.

El coeficiente de validez (la correlación entre una medida del predictor y una medida del criterio de conducta) es el atributo de selección que ha tenido una historia más larga. Su valor, y el de otros índices derivados, permiten concluir que sólo grandes diferencias en el coeficiente de validez producen importantes diferencias en el valor de un test. Como indicadores de la utilidad de un test para tomar decisiones son bastante deficientes. Así, implica una visión poco realista de la función resultado. El coeficiente de correlación mide las desviaciones al cuadrado desde una función lineal de predicción, tratando cualquier desviación del valor predicho por la función lineal como indeseable. La sobrepredicción es tratada como equivalente a la infrapredicción.

En términos de los tres atributos básicos del programa (cantidad, calidad y costo), los modelos basados en la correlación no reflejan ni la cantidad de períodos de tiempo afectados por las decisiones de selección ni la cantidad de empleados afectados en cada período de tiempo. Por tanto, proporciona sólo evidencia indirecta del efecto del predictor sobre la calidad de la fuerza de trabajo. Por último, fallan al reconocer que el desarrollo y aplicación de programas de selección lleva consigo costos.

1.2. El ratio de éxito.

El ratio de éxito representa el porcentaje de individuos seleccionados que tienen éxito en el puesto. De acuerdo con las tablas de TAYLOR-RUSSELL (1939), cuando otros parámetros permanecen constantes:

1. Altos coeficientes de validez producen ratios de éxito mejores que bajos coeficientes.
2. Bajos ratios de selección proporcionan altos ratios de éxito.
3. Tasas cercanas al 0'50 proporcionan ratios de éxito mejores que tasas cercanas al 0 (ningún candidato tiene éxito) y que las tasas cercanas al 1 (todos tienen éxito, aunque no se realice selección).

Con respecto a la función resultado, con un criterio dicotómico (satisfactorios o insatisfactorios), se perderá información, porque el valor del rendimiento no es igual para todos los que se encuentran en el grupo de los satisfactorios, ni igual para todos los que se encuentran en el grupo de los insatisfactorios (CRONBACH y GLESER, 1965; CASCIO, 1982; y HUNTER y SCHMIDT 1982). La situación más típica es que existan diferencias dentro de cada uno de los grupos. En tales situaciones sería más apropiado un criterio continuo. Todavía aquí, el modelo de TAYLOR-RUSSELL puede proporcionar apoyo para tales situaciones. CASCIO (1982) sugiere que puede ser más apropiado para criterios verdaderamente dicotómicos (rotación), o donde las diferencias en resultados por encima de un nivel aceptable no produce cambios significativos en los beneficios (tareas administrativas o técnicas), o cuando tales diferencias no se pueden medir (enfermería o enseñanza). En términos de los tres atributos básicos del programa (cantidad, calidad y costo), el modelo de TAYLOR-RUSSELL (como el anterior) ignora la cantidad de empleados afectados y el número de períodos de tiempo que durará el proceso. El modelo hace una buena contribución, al describir el cambio en calidad producido por el programa, pero esta medida de la calidad debe ser interpretada de modo diferente de una situación a otra. También, el modelo de TAYLOR-RUSSELL ignora los costos.

1.3. El criterio estandarizado.

La mayor crítica realizada al ratio de éxito era que el criterio dicotómico fallaba al reflejar la variación real en la actuación. Esta visión de la utilidad, intentó solucionar esto, definiendo un criterio continuo como función resultado. BROGDEN (1946a,b) usó los principios de la regresión lineal para demostrar la relación entre el coeficiente de correlación e incrementos en el criterio medido en una escala continua. Asumiendo una relación lineal entre las puntuaciones del criterio y del predictor, se derivó la mejor estimación lineal y no sesgada del cambio en puntuaciones típicas del criterio (Z_y) correspondientes a cambios en las puntuaciones típicas del predictor (Z_x) en la población de solicitantes. La ecuación de predicción lineal sería: $Z_y = (r_{xy}) (Z_x)$. Así, si conocemos la puntua-

ción media estandarizada del predictor de un grupo selecto de solicitantes, nuestra mejor predicción de la media estandarizada del criterio sería el producto del coeficiente de validez y la puntuación estandarizada del predictor.

Este modelo de utilidad refleja un criterio continuo (expresado en puntuaciones típicas) como función resultado, e incluye como atributos la validez y la media estandarizada del predictor. NAYLOR y SHINE (1965) mostraron en sus tablas la media en puntuaciones típicas del predictor para cada ratio de selección. En este modelo, al estar expresado el criterio en unidades típicas, es difícil interpretarlo en unidades más naturales al proceso de decisión (como dinero, unidades producidas o costos reducidos). También esta función resultado refleja sólo la diferencia entre la media del criterio en puntuaciones típicas de los que han sido seleccionados usando el predictor y la media en puntuaciones típicas en el criterio de los que se seleccionarían sin el predictor. La utilidad total de el programa no se evalúa, sólo el incremento producido de usar el predictor a no usarlo.

Con respecto a los tres conceptos básicos del modelo de utilidad (cantidad, calidad y costo), la cantidad de empleados y tiempo no se refleja. La calidad del criterio es más bien una unidad estadística más que una unidad tangible y el costo del programa de selección se omite todavía.

1.4. El criterio en unidades monetarias.

El mayor inconveniente de utilizar las puntuaciones típicas del criterio como función resultado es que resulta bastante difícil de interpretar en unidades reales. Los servicios de selección, a menudo, son considerados como costes y medidos en unidades monetarias. Con un criterio estandarizado, uno se puede hacer preguntas como: ¿es útil gastar 6.000.000 de pesetas en seleccionar 50 personas al año, con el fin de obtener 0'5 puntos más en el criterio, en unidades típicas de desviación, que si no utilizáramos el predictor? Obviamente muchos directores de personal no están familiarizados con el concepto de desviación típica, y encontrarían difícil asignar un valor económico a 0'5 unidades típicas de desviación de incremento en el criterio.

BROGDEN (1946a, b, 1949) y CRONBACH y GLESER (1965) calcularon fórmulas de utilidad en términos de la función resultado más que en puntuaciones típicas del criterio, incluyendo el concepto de costo. Ellos fueron los que dieron origen a la noción de utilidad en una escala de valores económicos. Para conseguir esto, introdujeron un factor escalar que transformaba las unidades típicas del criterio en unidades monetarias, y además añadieron un término para los costos del programa de selección. El factor escalar era el valor en unidades monetarias de una desviación típica de diferencia en el criterio (SDy). El factor de costo es el producido por administrar el predictor a un solicitante individual (C). Finalmente, la utilidad, simbolizada como U , representa la diferencia en unidades monetarias entre realizar la selección utilizando el predictor o sin él (habitualmente denominada la utilidad incremental del predictor). El resultado se puede expresar con la siguiente ecuación: $U = (SDy) (rxy) (Zx) - C/SR$. El costo por solicitante (C) se divide por el ratio de selección (SR), para calcular el costo total para obtener cada seleccionado. La ecuación anterior representa

el incremento en unidades monetarias del criterio de las personas que han sido seleccionadas con un predictor x , en una población de solicitantes donde el coeficiente de validez es rx ; una desviación típica en unidades de desviación de diferencia en el criterio es SDy ; la media en puntuaciones típicas en el predictor de los seleccionados en Zx y el coste por seleccionado de usar el predictor es C/SR . Para expresar la ganancia total de usar un predictor para seleccionar a N candidatos, habría que multiplicar los beneficios por el número de seleccionados y multiplicar el costo por solicitante por el número de aspirantes.

CRONBACH y GLESER (1965) también recomendaron calcular la diferencia en utilidad entre dos tests, sustituyendo el coeficiente de correlación por la diferencia entre coeficientes y el costo por la diferencia entre costos. Finalmente, para incorporar la duración de los efectos del grupo de empleados seleccionados, SCHMIDT, HUNTER, McKENZIE y MULDROW (1979) multiplicaron el componente de beneficio de este modelo por la permanencia estimada en el puesto de los individuos contratados.

Situar los niveles del criterio en dólares por persona y período de tiempo parece estar más en conformidad con los objetivos organizacionales de incrementar los beneficios. Este modelo de utilidad incorpora un factor escalar (SDy) que permite transformar el criterio en unidades monetarias. La medida de SDy ha generado grandes controversias como se verá más adelante.

Este modelo incorpora los tres conceptos básicos del análisis de la utilidad en selección (cantidad, calidad y costo). La cantidad se incorpora en el número de seleccionados y en el tiempo de permanencia media. La calidad se incorpora en el producto de rx , Zx y SDy (produciendo el valor incremental en unidades monetarias del criterio por persona y por período de tiempo, es decir, la mejora en pesetas por persona y año al emplear ese predictor). Los costos para desarrollar y llevar a cabo el programa de selección están contenidos en C .

Este modelo permaneció largo tiempo desconocido por los psicólogos del trabajo y de las organizaciones, aunque representó una importante alternativa a la teoría de la medida tradicional como marco de trabajo. Las razones de esta ausencia de atención son poco claras. Es probable que la complejidad algebraica del modelo provocara desaliento en los directivos, encontrando los investigadores dificultades para comunicar los propósitos e importancia del modelo. Más aún, los investigadores han venido asumiendo incorrectamente que todos los parámetros deben ser medidos con bastante exactitud para poder ser aplicados, puesto que su propósito era producir una estimación puntual de la utilidad. Esta concepción errónea existe todavía, como se verá más adelante.

HUNTER y SCHMIDT (1982) y SCHMIDT *et al.* (1979) observaron la limitada aplicación de este modelo y propusieron tres concepciones erróneas para explicarlo:

1. La creencia de que las ecuaciones de utilidad carecen de valor, a menos que los datos se ajusten exactamente al modelo lineal homocedástico y todas las distribuciones marginales sean normales.

2. La creencia de que la validez de los tests es específica de la situación, pudiendo hacer aplicaciones del análisis de la utilidad sólo cuando se haya realizado el estudio de validez en la situación particular (recordemos que la investigación sobre la generalización de la validez sugiere que la mayoría de la variabilidad en los coeficientes de validez observados se debe a artefactos de los estudios más que a diferencias reales), y
3. La creencia de que el parámetro que hace referencia a los valores estandarizados en unidades monetarias (SDy) es difícil o imposible de medir. En otro apartado discutiremos los problemas de su medida.

2. La medición de la desviación típica del rendimiento en unidades monetarias (SDy).

La desviación típica del valor en dólares del rendimiento en la población de candidatos (SDy) fue caracterizada como el «talón de Aquiles» del análisis de la utilidad por CRONBACH y GLESER (1965), debido a la creencia de que sólo se podía medir con procedimientos complejos y costosos (ROCHE, 1965). Esto dificultó la aplicación amplia de los conceptos de utilidad a la psicología del personal. No obstante, el interés en los cálculos de la utilidad de las prácticas de personal se ha ido renovando, debido, principalmente, a los nuevos procedimientos de estimación de SDy , que han evolucionado desde los datos basados en costes contables a los datos basados en criterios conductuales. La gran cantidad de investigación en este campo (de 5 estudios entre 1953-1979, se pasó a 21 entre 1979-1986) considera la exactitud de la medida de SDy como un requerimiento fundamental para la investigación del análisis de la utilidad en selección (BURKE y FREDERICK, 1984, 1985; WEEKLEY, O'CONNOR, FRANK y PETERS, 1985).

A continuación examinaremos cuatro aspectos relacionados con la medición de SDy : las técnicas de medida, los efectos del ambiente, los efectos de las escalas de resultados y la población objetivo.

2.1. Las técnicas de medida.

Debido a que las diferencias en SDy pueden producir grandes diferencias en la utilidad total, los procedimientos de cálculo de SDy ocupan un lugar relevante. En la actualidad, estos procedimientos se pueden clasificar en cuatro grandes grupos (BOUDREAU, 1989):

1. De contabilidad de costes (costes contables).
2. De estimación global.
3. De estimación individualizada, y
4. Reglas proporcionales.

Examinémoslos con más detenimiento.

2.1.1. Procedimientos basados en costes contables.

Se caracterizan por usar las técnicas contables para vincular un valor a unidades de rendimiento o producción para cada individuo. La desviación típica de estos valores de rendimiento individual representa SD_y . La dificultad y arbitrariedad de este método han sido citadas frecuentemente como argumentos en favor de métodos más simples (CASCIO, 1982; CASCIO y RAMOS, 1986; HUNTER y SCHMIDT, 1982 y SCHMIDT *et al.*, 1979).

2.1.2. Procedimientos de estimación global.

Se basan en las estimaciones de expertos del valor del rendimiento en dólares de 2, 3 ó 4 percentiles de una distribución hipotética de rendimiento, representando SD_y , las diferencias medias entre los percentiles estimados (BOBKO *et al.*, 1983; BURKE y FREDERICK, 1984, 1985; HUNTER y SCHMIDT, 1982; SCHMIDT *et al.*, 1979; WEEKLEY *et al.*, 1985 y WROTEN, 1984). Este modelo general de estimación, propuesto por SCHMIDT *et al.* (1979), utiliza las estimaciones de los supervisores del valor de los productos y servicios producidos. SCHMIDT y HUNTER (1982) partieron de la idea de que los mandos intermedios eran capaces de evaluar la productividad del personal bajo su responsabilidad, así como las diferencias entre los mejores y los peores. Así, pedían a los mandos que hicieran una evaluación financiera del valor de un empleado medio, de un empleado excelente y de un empleado mediocre. Para facilitar esta evaluación, utilizaban un cuestionario que les sugería que pensarán en el coste que supondría para la organización tener que adquirir en el exterior los bienes o servicios prestados por el personal evaluado. Otra ayuda para la estimación, era definir a los empleados excelentes como pertenecientes al 15 por 100 de los mejores y los mediocres como pertenecientes al 15 por 100 de los peores, en una clasificación jerárquica de todos los empleados bajo sus órdenes. Al corresponder estos valores a la media más o menos una desviación típica dentro de una distribución normal, era posible calcular la desviación típica que caracteriza la dispersión de los «valores financieros» de una categoría de empleados para la organización. Las estimaciones de los ejecutivos son objeto de medias que permiten calcular el valor financiero asociado a una mejor selección. SCHMIDT *et al.* (1980), demostraron, que la diferencia de valor financiero entre un programador excelente y un programador medio, así como entre un programador medio y uno mediocre era de 10.000 dólares anuales.

Existe evidencia limitada sobre la exactitud de la estimación global de SD_y , estando basadas las pruebas en medidas objetivas del rendimiento deficientes y discutibles. BOBKO *et al.* (1983) encontraron que la distribución actual de ingresos de venta (núm. de pólizas vendidas, valor medio de la póliza, etc.) para agentes de venta, se distribuía normalmente, y que la SD_y estimada basada en las diferentes medias entre los percentiles 85 y 50, y entre los percentiles 50 y 15, no eran significativamente diferentes de la distribución de ventas actuales. Sin embargo, cuando a los encuestados se les preguntaba por el valor total de los productos o servicios, y por lo que pagarían en una organización externa por realizarlos, los valores se anclaban en niveles salariales más que en niveles de ventas. BURKE y FREDERICK (1984) también encontraron estimaciones de SD_y del valor total más bajas (1% de la desviación típica de las ventas) ancladas en varias actividades incluyendo las ventas.

REILLY y SMITHER (1986) encontraron que estudiantes graduados que participaron en una simulación de negocios (a los que se les había proporcionado datos para estimar las desviaciones típicas actuales), produjeron estimaciones globales de SDy un poco más altas que la información de la simulación (ventas netas y nuevas ventas) y mucho más altas que la simulación para ingresos netos. La SDy estimada del valor total fue del 49 por 100 de las ventas actuales repetidas, 3'45 veces de las nuevas ventas actuales y 1'92 veces del ingreso neto. Por tanto, la investigación es dispersa y los resultados confusos, proporcionando poca evidencia de que las estimaciones globales de SDy reflejan las ventas actuales o información sobre la productividad. Estas estimaciones dejan, por tanto, alguna duda sobre la presunción de que los estimadores usan una variable distribuida normalmente como una base cognitiva para sus procesos de estimación, incluso aunque el objetivo indicador del rendimiento en el puesto se distribuya normalmente (como en el estudio de BOBKO y cols. citado anteriormente). Desde un punto de vista lógico, la equivalencia en dólares de la diferencia entre los percentiles 85 y 50, y entre los percentiles 50 y 15, no es generalmente un indicador inequívoco de una distribución normal como base cognitiva del proceso de estimación de los estimadores porque cualquier distribución asimétrica (p.e. rectangular) podría también proporcionar esta equivalencia.

El principal problema del procedimiento de estimación global lo constituye la variabilidad entre los estimadores, unido a la dificultad del proceso de estimación (algunos estimadores incluso rehúsan la tarea por considerarla difícil). La variabilidad interestimadores podría estar producida por el hecho de que los estimadores usan escalas diferentes al hacer sus estimaciones. REILLY y SMITHER (1985) atribuyen parte de la varianza entre estimadores a diferentes interpretaciones del valor total: el valor total de los productos y servicios por un lado, y el costo de contratar a una firma externa para realizar estos servicios. BOUDREAU (1983a) ha señalado correctamente que el valor de los productos y servicios no necesita ser igual al coste de obtenerlos en una firma externa.

Una sustancial reducción de la variabilidad entre estimadores fue conseguida por BURKE y FREDERICK (1984, 1986) y EDWARDS, FREDERICK y BURKE (1988), modificando el procedimiento original en el modo seguido para establecer un punto de referencia sobre la escala de actuación en dólares, equivalente para todos los estimadores. Para ello, los estimadores estimaban el percentil 50. Después, se les informaba individualmente del percentil medio de todos los estimadores y se realizaba la estimación de los percentiles restantes por cada estimador.

Como conclusión, ofrecemos las orientaciones de SCHULER y GULDIN (1991) para mejorar la aplicación del modelo general de estimación (GEM). En primer lugar, entrenar a los estimadores sobre las asunciones, cálculos e interpretaciones de los percentiles para eliminar la variabilidad entre estimadores. En segundo lugar, proporcionar un punto ancla para la escala por un procedimiento de *feedback* del percentil medio 50, para reducir la variabilidad de los estimadores. En tercer lugar, eliminar cualquier ambigüedad en las interpretaciones del cuestionario, y por último y en cuarto lugar, tomar más de los tres percentiles usuales (15, 50 y 85) en cuenta a la hora de chequear si la asunción general de una distribución normal del rendimiento en el trabajo (siendo real en la mente de cualquier estimador) no es violada.

2.1.3. Procedimientos de estimación individualizada.

Estos métodos transforman algunas características medibles de cada individuo de la muestra (p.e. sueldo, ventas...) en dólares, usando un factor escalar como la media del salario o la media de las ventas, representando la desviación típica de estos valores SDy (ARNOLD *et al.* 1982; BOBKO *et al.* 1983; BURKE y FREDERICK, 1984; CASCIO y RAMOS, 1986; DUNNETTE *et al.* 1982; JANZ y DUNNETTE, 1974 y REILLY y SMITHER, 1985). Hay tres técnicas dentro de estos procedimientos:

1. El CREPID de CASCIO y RAMOS (1986).
2. El de JANZ y DUNNETTE (1977), y
3. El de BOBKO *et al.* (1983).

A. CREPID (*Cascio Ramos Estimate of Performance In Dollars*).

Se debe a CASCIO (1982) y CASCIO y RAMOS (1986). Comparado con el anterior, este procedimiento estima SDy analíticamente. Esto incluye un análisis del puesto, una estimación separada y ponderada de las principales actividades del puesto y finalmente la distribución del rendimiento global de los trabajadores en el puesto, desde donde se deduce SDy . El trabajo se descompone en «actividades principales». Cada actividad es medida sobre cuatro dimensiones: tiempo/frecuencia, nivel de dificultad, importancia y consecuencia del error. Las estimaciones resultantes de cada actividad en tales dimensiones se multiplican, dando lugar a un peso total de la actividad. Para asignar un valor en dólares a cada actividad, el salario medio para el puesto es dividido entre las actividades de acuerdo con la importancia proporcional de cada una de ellas. Después de esta fase de análisis del puesto, los superiores son preguntados sobre el rendimiento de los empleados en cada actividad, usando una escala de 0 a 2. Para transformar estas estimaciones en dólares, se multiplica por el valor en dólares asignado a esa actividad. Una vez que a cada empleado se le ha asignado un valor en dólares para cada actividad, se suman todos estos valores para proporcionar el valor en dólares total del rendimiento anual del empleado.

EDWARDS, FREDERICK y BURKE (1988) señalaron que este procedimiento consumía bastante tiempo. Así, su utilidad podría ser incrementada sustancialmente usando datos de archivo del rendimiento de los trabajadores. Estos autores demostraron empíricamente la adecuación de esta serie de datos para el procedimiento CREPID.

En cuanto a la exactitud de las estimaciones proporcionadas por el método CREPID, REILLY y SMITHER (1985) informaron de una fiabilidad entre estimadores de 0'87, mientras GREER y CASCIO (1987) obtuvieron 0'70. Por añadidura a este objetivo de medida de la exactitud, la exactitud subjetivamente percibida es también alta porque los directivos evaluaron sus propias estimaciones en el modelo CREPID como más exactas que las estimaciones del modelo GEM (EDWARDS *et al.*, 1988).

De la comparación de los procedimientos se reportan las siguientes resultados (SCHULER y GULDIN, 1991):

1. El procedimiento CREPID produce estimaciones de SDy más pequeñas comparadas con el GEM (EDWARDS *et al.*, 1988; GREER y CASCIO, 1987; REILLY y SMITHER, 1985; WEEKLEY *et al.*, 1985). GREER y CASCIO (1987) demostraron que el método CREPID proporcionaba un rango mucho más estrecho de valores comparado con el GEM, un mayor grado de validez aparente, y producía un valor medio del mérito que era el 98 por 100 (sólo 56% en el GEM) del que producía la aproximación basada en costes contables.
2. No hay una evaluación concluyente sobre la convergencia de CREPID y/o estimaciones de SDy GEM con el procedimiento de reglas proporcionales 40-70 por 100.
3. Desde un punto de vista teórico, los tres métodos difieren en su contexto de evaluación (EDWARDS *et al.*, 1988; RAJU *et al.*, 1990; STEFFY y MAURER, 1988). CREPID y el procedimiento de reglas proporcionales 40-70 por 100 tienen la misma evaluación base-salario. El salario, un coste de servicio, es lo que los economistas y contables califican como una evaluación *ex-post* (p.e. un activo del valor actual a la organización o al propietario), mientras que el GEM se basa en una evaluación *ex-ante* (p.e. un activo del valor presente neto resultante de su uso por la organización). Como STEFFY y MAURER señalaron, nosotros no podemos esperar necesariamente resultados convergentes de los procedimientos basados en diferentes contextos de evaluación.

B. La aproximación de JANZ y DUNNETTE (1977).

Este segundo procedimiento de estimación individualizado comprende también la identificación de las actividades críticas del puesto. Sin embargo, más que distribuir el salario en cada actividad según su tiempo, frecuencia o importancia, este procedimiento requiere expertos en el trabajo para estimar el costo en dólares asociado con los diferentes niveles de eficacia en cada dimensión de rendimiento del puesto. Se investigan así, las consecuencias de varios niveles de eficacia en la determinación de su impacto en las actividades en las que el costo y/o el valor puede ser unido.

C. La aproximación de BOBKO *et al.* (1983).

Por último, este tercer procedimiento de estimación individualizado implica tener expertos que asignen dólares a los empleados individuales directamente. BOBKO usó este método para derivar una SDy estimada basada en niveles de ventas, representando el nivel individual estimado las ventas de cada persona. BURKE y FREDERICK (1984) también usaron niveles individuales de ventas. WROTEN (1984) adoptó una posición similar, pero no tenía datos de ventas disponibles. Simplemente preguntó a sus supervisores por una estimación anual en dólares del rendimiento de cada empleado. LEDVINKA *et al.* (1983) usó la nómina total más los beneficios dividido entre el número de reclamaciones de seguros como el valor por demanda y entonces multiplicó este valor por la desviación típica actual de los seguros procesados.

La estimación individualizada tuvo la ventaja de asignar un valor específico a cada empleado que puede ser analizado y examinado explícitamente para ver su conveniencia. Un análisis así sería útil para determinar qué atributos individuales contribuyen a las diferencias en los juicios. Los métodos que incluyen el análisis conductual del trabajo (CREPID, JANZ y DUNNETTE), pueden ser más entendibles y creíbles a aquellos que están familiarizados con el trabajo, aunque no existe evidencia absoluta sobre este problema. Incluso, cada método hace ciertas asunciones básicas sobre la naturaleza de la función resultado (*payoff*). CREPID se basa en la asunción de que el salario medio está a la altura de la productividad media, una posición no apoyada por la teoría económica y claramente violada por organizaciones con sistemas de pago basados en el tiempo de permanencia. Las medidas basadas en las ventas se basan en la asunción de que las ventas reflejan suficientemente las diferencias en rendimiento. Esta asunción puede ser útil, pero olvida tareas relevantes como el entrenamiento, que reduce las ventas individuales pero incrementa las del grupo. El procedimiento de JANZ y DUNNETTE se basa en la asunción de que los efectos de la conducta en el puesto sobre los costos e ingresos pueden ser exactamente investigados por los directivos.

Los métodos individualizados son a menudo más complejos, costosos en tiempo y dinero que otros métodos, no existiendo evidencia clara y convincente de que mejoran las decisiones con respecto a ellos.

2.1.4. Reglas proporcionales.

Esta aproximación multiplica alguna variable disponible relacionada con la productividad (p.e. el salario medio, ventas medias...) por una proporción para estimar SD_y (HUNTER y SCHMIDT, 1982; SCHMIDT y HUNTER, 1983; WEEKLEY *et al.*, 1985 y CASCIO y RAMOS, 1986). Las reglas proporcionales emergieron en parte de las observaciones sobre la relación entre las estimaciones de SD_y y los niveles medios de los salarios, y en parte, del deseo de proporcionar una medida de SD_y sencilla que pudiera ser usada cuando una estimación global no fuese posible. Este método consiste en multiplicar el salario medio de un puesto por alguna proporción (p.e. entre 40 y 70%) para derivar la SD_y estimada para el grupo de empleados.

HUNTER y SCHMIDT (1982) revisaron los estudios empíricos y compararon la SD_y estimada, descubriendo que esta SD_y resulta por término medio del 16 por 100 del salario medio. Estos autores también revisaron dos de sus propios estudios (usando un procedimiento de estimación global), y observaron que el SD_y era el 60 por 100 del salario anual (en un estudio de analistas de presupuestos) y del 55 por 100 del salario anual (en otro estudio de analista de programadores). Ellos estimaron que *la media real para SD_y cae en algún lugar entre el 40 y el 70 por 100 del salario bruto anual*. También revisaron datos empíricos sobre niveles de productividad medidos en unidades de producción, concluyendo que la estimación de la desviación típica del rendimiento del empleado sin miedo a exagerar, puede ser fijada en el 20 por 100 de la media de la producción y que se apoya con los datos que no es erróneo utilizar como SD_y el 40 por 100 del salario medio.

Las reglas proporcionales propuestas por SCHMIDT y HUNTER son intrigantes porque sugieren que la estimación simple de SD_y puede ser bastante factible en casi todas las situaciones. Sin embargo, esta sobresimplificación se obtiene asumiendo que el salario medio es en efecto igual a

alrededor de la mitad del valor medio de los productos vendidos, lo cual no se cumple en sistemas de pago basados en la permanencia o sistemas de pago negociados, así como por condiciones del mercado tales como el desempleo o mercados de trabajo internos. Uno debe asumir también, que SDy es igual a casi el 20 por 100 de la media de los productos vendidos, lo cual no se da en varios estudios de los revisados por SCHMIDT y HUNTER (1983).

La revisión de BOUDREAU (1989) de los estudios de utilidad hasta 1986, descubrió 17 estimaciones de SDy por debajo del 40 por 100 del salario, 18 dentro del rango del 40-70 por 100 y 29 por encima del 70 por 100. Ninguna cayó por debajo del 20 por 100, 9 cayeron entre el 23 y el 34 por 100 y algunas fueron del 100 por 100 o más. Mientras estos resultados apoyan el conservadurismo de esta regla de decisión, es posible que usar la regla proporcional de HUNTER y SCHMIDT, al producir estimaciones tan conservadoras y subestimadas, pueda llevar a rechazar la utilidad potencial de los programas de recursos humanos.

La investigación existente sugiere que las diferencias estimadas en SDy usando diferentes métodos, son a menudo menores al 50 por 100. Sin embargo es tentador considerar que el hecho de que estas diferencias puedan ser multiplicadas por factores de cientos o miles de unidades monetarias a la hora de obtener el valor final de utilidad. Incluso, una pequeña diferencia multiplicada por valores tan grandes puede implicar inmensas diferencias en la utilidad total. Por tanto, es interesante realizar más investigación sobre SDy para ir rebajando poco a poco tales diferencias y proporcionar mejores estimaciones de la utilidad total.

2.2. Los efectos del ambiente.

Este segundo aspecto relacionado con la medición de SDy indica que existen diferentes trabajos con las mismas funciones y puestos con la misma denominación pero diferentes competencias. Así, hay trabajos donde los trabajadores emplean más discreción con respecto a la calidad o cantidad de trabajo y donde la variación en la producción tiene grandes implicaciones para las metas organizacionales. En estos trabajos los valores de SDy deberían ser más altos, teniendo en cuenta que la variabilidad en las habilidades y la motivación de la fuerza de trabajo sea la misma. Por tanto, el ambiente concreto del puesto es un factor relevante que se ha de tener en cuenta a la hora de calcular SDy , ya que al ser diferente, influye en la obtención de SDy .

2.3. Los efectos de las escalas de resultados.

La función resultado (*payoff*), ya comentada más arriba, se refiere a todas las consecuencias de una decisión dada que afectan a la persona que toma la decisión o a la institución que representa (CRONBACH y GLESER, 1965). Las medidas de resultados deben reflejar diferentes consecuencias (p.e. incremento de la productividad, reducciones del coste de trabajo, imagen mejorada de la organización, altos niveles de ganancias financieras...) en diferentes situaciones de decisión, consistentes con los objetivos de los que toman las decisiones. Las escalas de las funciones de resulta-

dos (*payoff*) usadas en la investigación del análisis de la utilidad, se centran en las consecuencias del incremento de la calidad en la fuerza de trabajo. Así, el resultado de la mejora en la selección depende de cómo se mida el incremento en la calidad producida por tales programas.

Las formas generales de medir la mejora de la calidad en la fuerza de trabajo son:

1. Por el incremento de la cantidad y la calidad.
2. Por el incremento de la calidad de la producción, y
3. Por la reducción de los costes de producción.

Una escala de resultados definida en términos de beneficio puede reflejar alguno o todos estos usos. Las escalas de resultados que reflejan el aumento de ingresos y reducciones de costos dominan la literatura de la utilidad en selección, aunque están aumentando las escalas basadas en beneficios. La mayoría de las primeras aplicaciones sobre análisis de la utilidad en selección, se centraron en la reducción del costo como el producto más saliente de la mejora en la calidad. SCHMIDT *et al.* (1979) midieron el *SDy* en términos del valor anual de los productos y servicios, así como del coste de emplear una firma externa que proporcionara estos productos. HUNTER y SCHMIDT (1982) interpretaron la función resultado como «el valor de la producción como ventas» o «lo que el empleado carga al cliente». No obstante, evaluar las inversiones en selección como contribución en ingresos puede inflar artificialmente los valores de la utilidad de la selección.

BRODGEN y TAYLOR (1950) propusieron el criterio en dólares para reflejar los ingresos por ventas generados cuando un producto es vendido, menos los costes de producción. CRONBACH y GLESER (1965) proporcionaron un concepto muy general de *payoff*, incluyendo todas las consecuencias importantes para los que toman las decisiones, consistente con una definición del beneficio. Sólo un estudio reciente (REILLY y SMITHER, 1985) adoptó una función de resultados reflejando la contribución de beneficios, con resultados que sugieren que los estudiantes graduados, en sus simulaciones, diferían la mayoría de ellos en sus estimaciones de *SDy*, cuando eran cuestionados sobre ingresos netos, más que sobre «nuevas ventas» o «valor total». BOBKO *et al.* (1983) encontraron que los supervisores de ventas exhibían más variabilidad en las estimaciones de *SDy* cuando intentaban estimar el valor anual de la compañía que cuando estimaban las ventas anuales de la compañía en dólares. La investigación futura debería centrarse en cómo las mejoras en productividad deben ser usadas para que la función resultado refleje mejor la decisión actual.

2.4. Los efectos de la población objetivo.

Todas las estimaciones de *SDy* están basadas en la población actual (BOBKO *et al.* 1983; BURKE y FREDERICK, 1985; JANZ y DUNNETTE, 1977; SCHMIDT *et al.*, 1979; SCHMIDT y HUNTER, 1983; WROTEN, 1984), porque es más familiar a los supervisores. Sin embargo, técni-

camente esta población no es la interesante. Para el análisis de la utilidad, la población de interés es la población de solicitantes a los que se aplica las pruebas, que pueden diferir en muchos aspectos de la otra.

Así, la población actual es una muestra restrictiva de solicitantes del trabajo, estando *SDy* sesgado a la baja. Los cambios en la población de aspirantes en el tiempo debido a diferentes procedimientos de reclutamiento o influencias del mercado laboral, también pueden operar incrementando o disminuyendo la variabilidad en el rendimiento, provocando que los niveles de *SDy* de los solicitantes sean mayores o menores que los de los empleados actuales. Por otro lado, estimar *SDy* en los empleados actuales motiva a los estimadores a considerar a todos los empleados con experiencia que ellos han tenido, incluyendo empleados con diferentes niveles de permanencia. Si el rendimiento varía con la permanencia, la *SDy* estimada reflejará la variabilidad de la permanencia. Una *SDy* estimada de los empleados, sobreestimaré la *SDy* de los aspirantes. Por último, puesto que todos los grupos de empleados con denominaciones similares del puesto forman la población de interés, al existir diferencias reales en las asignaciones de las tareas o ambientes de trabajo dentro de tales grupos, se van a producir diferencias en la variabilidad del rendimiento.

Las estimaciones de *SDy* basadas en la población actual, pueden ser un reflejo inexacto del *SDy* actual del sistema de selección. La mayoría de los autores que discuten el problema adoptan el argumento de que las estimaciones de *SDy* basadas en los empleados actuales son conservadoras, debido a la restricción del rango. Sin embargo, ningún estudio ha comparado las estimaciones de *SDy* basadas en la población de candidatos con las estimaciones de *SDy* basadas en la población actual de trabajadores.

V. EJEMPLO DE CÁLCULO DE LA UTILIDAD PARA MEDIR EL IMPACTO ECONÓMICO DE UN PROGRAMA DE SELECCIÓN

El presente ejemplo pertenece a una muestra de 150 sujetos preseleccionados de una población de 3.381 personas. La población procedía de las solicitudes de empleo recibidas por la empresa GENERAL ELECTRIC PLASTICS DE ESPAÑA, S.A., ya bien de forma espontánea o como consecuencia de los anuncios que se publicaron en la prensa regional durante los años 1992 y 1993 para cubrir plazas de alumnos en cursos de pre-empleo para operadores de planta química.

El proceso consistía en preseleccionar a grupos de 30 personas, que recibirían un curso previo selectivo de 102 horas, de forma que al final del mismo y a través de una evaluación final, se determinara quiénes pasarían a formar parte de una bolsa de trabajo desde la que se irían contratando a los operarios que fueran necesarios para desarrollar los trabajos en la Planta de *Compounding* en Cartagena.

Desde la preselección inicial hasta la decisión última de contratación, el aspirante sería evaluado en una serie de variables que se describen en el apartado siguiente. Al final del curso habría que adoptar un criterio que permitiese combinar todos los datos disponibles del mejor modo posible y así obtener una decisión correcta.

Las características más representativas de la muestra de 150 sujetos preseleccionados para realizar los cursos de pre-empleo eran las siguientes: edad media de 27 años, varones en el 91 por 100 de los casos y mujeres el 9 por 100 restante, experiencia laboral media de 57 meses, nivel de estudios predominante FP 2 (61%) y situación laboral mayoritaria de desempleo (60%).

VARIABLES E INSTRUMENTOS DE MEDIDA

Las variables medidas fueron las siguientes:

1. *Capacidad intelectual*, evaluada mediante el subtest *ACI* (Apreciación de la capacidad intelectual) de la *batería de operarios*. Es una prueba de tipo *omnibus*, en la que se han hecho intervenir los principales aspectos o factores primarios de la inteligencia. En su composición intervienen elementos de comprensión y razonamiento verbal, aptitudes numérica, espacial, mecánica y comprensión espacial. Proporciona una medida rápida y sencilla de la dotación intelectual (15 minutos de tiempo máximo). Se utilizó la versión de TEA Ediciones, S.A. de 1986.

2. *Comprensión mecánica y aptitud espacial*, evaluada mediante la prueba *Palancas*, también integrada en la *batería de operarios*. Esta prueba aprecia la aptitud espacial y comprensión mecánica en poblaciones de operarios y profesionales de oficio. Está constituida por 24 elementos formados por dibujos que representan varias palancas con diversos puntos de giro, dos o más de los cuales son fijos. La tarea consiste en averiguar en qué dirección se moverían las palancas si se aplicase una fuerza en el extremo de una de ellas. Es aplicable y parece discriminar bien en poblaciones de operarios industriales, y en general en todas aquellas tareas en las que es fundamental la comprensión mecánica y el manejo de máquinas o sus representaciones gráficas. Su tiempo máximo de ejecución es de nueve minutos, y se utilizó la versión de TEA Ediciones, S.A. de 1986.

3. *Atención y aptitud perceptiva*, medida por la prueba *formas idénticas*. Esta prueba parece ligada a cierto tipo de aptitudes perceptivo-espaciales no medidas por tests clásicos. Se han encontrado correlaciones que oscilan entre 0'30 y 0'50 con tests que definen los siguientes factores de Thurstone: visualización estática y orientación (S1 y S2 del autor), velocidad y rapidez perceptiva y algo de los factores inductivo y de razonamiento. La tarea consiste en identificar las figuras que son idénticas a las que aparecen como modelo o patrón, en un tiempo máximo de cuatro minutos. Se utilizó la versión de TEA Ediciones, S.A. de 1979.

4. *Comprensión de instrucciones de cierta complejidad*, medida por la prueba *Instrucciones Complejas (IC)*. Esta prueba, proporciona una estimación rápida y aproximada de la capacidad de comprensión y ejecución de órdenes escritas. Preparada inicialmente para apreciar un aspecto de las aptitudes llamadas administrativas, la nueva versión, se dirige también al exámen del personal ocupado en tareas técnico-mecánicas. Está formada por instrucciones escritas que deben ser leídas e interpretadas por el sujeto para la realización de la prueba. La tarea que se propone al sujeto, implica más bien la puesta en funcionamiento de un conjunto de habilidades que deben ser poseídas con-

junta y armónicamente. Las correlaciones con otros tests indican que el IC exige para su perfecta ejecución un buen nivel de razonamiento lógico (factor R), y en menor grado de inteligencia general y cálculo numérico (factores G y N). Cabe añadir además, que requiere desde el punto de vista de la personalidad un buen control emocional. Su tiempo máximo de ejecución es de 75 minutos. Se utilizó la forma B de la versión de TEA Ediciones, S.A. de 1972, cuyo autor es MARIANO YELA.

5. *Factores de personalidad*, medidos por el *Cuestionario factorial de personalidad 16 PF de R.B. CATTELL*. Este cuestionario se basa en estudios sistemáticos desarrollados a lo largo de más de 30 años de estudios factoriales que han evidenciado la existencia de 16 factores estables y relativamente independientes. Desde su primera versión en 1939 han aparecido numerosas revisiones, la última de las cuales es de 1994. Consta de 187 cuestiones con tres alternativas de respuesta (en las formas A y B), y 105 cuestiones (en las formas C y D). Los factores evaluados son de tipo bipolar, y vienen definidos tanto por términos técnicos como por sus adjetivaciones más corrientes. Los polos pueden ser denominados del siguiente modo:

- A. Sizotimia-Afectotimia (Reservado-abierto).
- B. Menos-Más inteligencia (Pensamiento concreto-abstracto).
- C. Menos-Más fuerza del ego (Emocionable-estable).
- E. Sumisión-Dominancia (Conformista-agresivo).
- F. Desurgencia-Surgencia (Moderado-impulsivo).
- G. Menos-Más fuerza del superego (Evasión reglas-perseverante).
- H. Trectia-Pármia (Tímido-atrevido).
- I. Harria-Prensia (Duro-sensible).
- L. Alaxia-Protensión (Confiado-suspicaz).
- M. Praxernia-Autía (Práctico-imaginativo).
- N. Sencillez-Astucia (Espontáneo-calculador).
- O. Imperturbabilidad-Culpabilidad (Tranquilo-aprensivo).
- Q1. Conservadurismo-Radicalismo (Tolerante-crítico).
- Q2. Adhesión social-Autosuficiencia (Dependiente-autosuficiente).
- Q3. Poca integración-Autocontrol (Indisciplinado-controlado).
- Q4. Poca-Mucha tensión energética (Relajado-tenso).

Existen otros cuatro factores de segundo orden: ansiedad, extraversión, socialización controlada e independencia. La amplia extensión de 16 PF en muchos países y su abundante bibliografía permiten considerar este instrumento como uno de los mejores cuestionarios factoriales existentes para evaluar la personalidad en distintas situaciones. En este estudio se aplicó la forma C de la adaptación española efectuada por TEA Ediciones, S.A. en 1975.

6. *Valores interpersonales*, medidos por el *Cuestionario de valores interpersonales (SIV)* de L.V. GORDON. El autor se ha fijado en un tipo específico de valores: los que inciden de manera preponderante en las relaciones de un individuo con sus semejantes y que tienen relevancia en su ajuste personal, social, familiar y profesional. Los incluidos en esta prueba son los siguientes:

- 1.º Estímulo (S). Deseo de comprensión por parte de los demás.
- 2.º Conformidad (C). Aceptación de las normas socialmente correctas.
- 3.º Reconocimiento (R). Consideración de los propios méritos por parte de los demás.
- 4.º Independencia (I). Capacidad de decidir y actuar según propios criterios.
- 5.º Benevolencia (B). Generosidad y ayuda a otras personas, y
- 6.º Liderazgo (L). Autoridad y mando sobre otros.

El cuestionario consta de 90 elementos, todos ellos formulados según una dirección positiva y puntuando en alguna de las seis escalas de valores. Se presentan agrupados en tríadas y el sujeto debe elegir entre las tres proposiciones de cada grupo la que le parece más y menos importante. Su brevedad (15 minutos), su relativa independencia de la inteligencia y otras aptitudes cognoscitivas, su aceptabilidad por parte de los candidatos y la importancia de sus escalas como indicadores del éxito en muchas tareas aconsejan su inclusión en ciertas baterías psicológicas. Se utilizó la versión de TEA Ediciones, S.A. de 1977.

EL PROCEDIMIENTO DE CÁLCULO DE LA UTILIDAD

Para las pruebas anteriores, se obtuvo un coeficiente de validez (R) media de 0'30 con respecto al criterio: «valoración obtenida de los alumnos durante el curso». La *SD* (la desviación típica de la distribución del criterio profesional) se estimó por el procedimiento de reglas proporcionales de HUNTER y SCHMIDT (1982). Así, teniendo en cuenta un salario medio de 2.500.000 pesetas brutas anuales, se obtuvo una cantidad de 750.000 pesetas (30% del salario bruto). La nota media en puntuaciones típicas obtenida por los seleccionados en los predictores (Z), se estimó a partir de la proporción de selección (procedimiento de NAYLOR y SHINE), de modo que para una proporción de 0'10 (10% de seleccionados con relación al número de examinados) le correspondió una Z

de l'25. Por último, el coste de selección por candidato fue de unas 35.000 pesetas. *Aplicando la ecuación de BROGDEN y CRONBACH y GLESER para el cálculo del beneficio esperado del programa de selección, se obtuvieron las siguientes cantidades:*

1. POR EMPLEADO Y AÑO.

$$(R \times SDy \times Z) - (C/P) = (0'30 \times 750.000 \times 1'5) - (35.000 \times 0'10) = 277.750 \text{ ptas.}$$

2. POR 150 EMPLEADOS Y 1 AÑO.

$$(R \times SDy \times Z) - (C/P) \times N = 277.750 \times 150 = 41.662.500 \text{ ptas.}$$

3. POR 150 EMPLEADOS Y 5 AÑOS.

$$(R \times SDy \times Z) - (C/P) \times N \times A = 277.750 \times 150 \times 5 = 208.312.500 \text{ ptas.}$$

4. POR 150 EMPLEADOS Y 10 AÑOS.

$$(R \times SDy \times Z) - (C/P) \times N \times A = 277.750 \times 150 \times 10 = 416.625.000 \text{ ptas.}$$

VI. CONCLUSIONES

1. En selección de personal se ha olvidado el aspecto económico referente al beneficio de los programas implementados, habiéndose centrado principalmente en aspectos técnicos como la validez o fiabilidad de las pruebas.

2. El concepto de utilidad es el lenguaje común entre directivos y psicólogos a la hora de tomar decisiones sobre la adopción de programas de selección, al reunir aspectos psicológicos como la validez y aspectos económicos como la desviación típica del rendimiento en pesetas.

3. Los modelos de utilidad son instrumentos de ayuda en las decisiones, auxiliando en la descripción, análisis, predicción y explicación de las mismas.

4. La utilidad de un predictor puede definirse como la mejora conseguida en la calidad de las contrataciones realizadas al emplear un dispositivo de selección en comparación con la que se obtendría al no usarlo.

5. En el cálculo de la utilidad influyen variables como: la fiabilidad del criterio y predictor, la pertinencia del criterio, la validez del predictor, la razón de selección y el porcentaje de empleados actuales que tienen éxito.

6. Aspectos relevantes en la investigación en los modelos de utilidad son: la conceptualización de la función resultado y la determinación de la desviación típica del rendimiento en unidades monetarias (SD_y).

7. Mediante un procedimiento relativamente sencillo es posible calcular el ahorro en pesetas producido al aplicar un conjunto de pruebas de selección.

8. El ahorro obtenido (la utilidad) en pesetas puede alcanzar cifras millonarias (en nuestro ejemplo, que se obtuvo con datos reales, pudo llegar a varios cientos de millones, si se calcula para 10 años y 150 personas).

BIBLIOGRAFÍA

- ARNOLD, J.D., RAUSCHENBERGER, J.M., SOUBEL, W. y GUION, R.M. (1983). «Validation and utility of a strength test for selecting steel workers». *Journal of Applied Psychology*, 67, 588-604.
- BOBKO, P., KARREN, R. y PARKINGTON, J.J. (1983). «Estimation of standard desviations in utility analysis: an empirical test». *Journal of Applied Psychology*, 68, 170-176.
- BOUDREAU, J.W. (1984). «Decision Theory contributions to HRM research and practice». *Industrial relations*, 23, 198-217.
- BOUDREAU, J.W. (1987). *Utility analysis applied to internal and external employee movement: an integrated theoretical perspective*. Ithaca, New York: Author.
- BOUDREAU, J.W. (1989). «Selection utility analysis. A review and agenda for future research». En M. Smith and I.T. Robertson (eds), *Advances in selection and assessment*. Chichester: Wiley & Sons.
- BOUDREAU, J.W. & BERGER, C.J. (1985a). «Decision-Theoretic utility analysis applied to external employee movement». *Journal of Applied Psychology*, 70, 581-612.
- BLUM, M.L. & NAYLOR, J.C. (1990). *Psicología industrial*. Trillas. México.
- BROGDEN, H.E. (1946a). «On the interpretation of the correlation coefficient as a measure of predictive efficiency». *Journal of Educational Psychology*, 37, 65-76.
- BROGDEN H.E. (1946b). «An approach to the problem of differential prediction». *Psychométrica*, 14, 169-182.
- BROGDEN, H.E. (1949). «When testing pays off». *Personnel Psychology*, 2, 171-183.
- BROGDEN, H.E. y TAYLOR, E.K. (1950). «The dollar criterion-applying the cost accounting concept to criterion construction». *Personnel Psychology*, 3, 133-154.
- BURKE, M.J. y FREDERICK, J.T. (1984). «Two modified procedures for estimating standard desviations in utility analysis». *Journal of Applied Psychology*, 69, 482-489.

- BURKE, M.J. y FREDERICK, J.T. (1985). *A comparison of economic utility estimates for alternative rational SDy estimation procedures*. Paper presented at the 45th annual meeting of the Academy of management. San Diego, CA.
- CASCIO, W.F. (1982). *Costing Human Resources: The Financial impact of behavior in Organizations*. Boston, MA: Kent.
- CASCIO, W.F. y RAMOS, R. (1986). «Development and applications of a new method for assessing job performance in behavioral/economic terms». *Journal of Applied Psychology*, 1, 20-28.
- CRONBACH, L.J. & GLEESER, G.C. (eds) (1965). *Psychological tests and personnel decisions*. 2nd edn. Urbana: University of Illinois Press.
- CRONSHAW, S.F. y ALEXANDER, R.A. (1985). «One answer to the demand for accountability: selection utility as an investment decision». *Organizational Behavior and Human Decisions Processes*, 35, 102-118.
- DARLINGTON, R.B. & STAUFFER, G.F. (1966). «Use and evaluation of discrete tests information in decision making». *Journal of Applied Psychology*, 50, 125-129.
- DUNNETTE, M.D., ROSSE, R.L., HOUSTON, J.S., HOUGH, L.M., TOQUAM, J., LAMMLEIN, S., KING, K.W., BOSSHARDT, M.J. y KEYES, M. (1982). *Development and validation of an industry-wide electric powerplant operator selection system*. Edison Electric Institute.
- EDWARDS, W. (1977). «Use of multiatributte utility measurement for social decision making». En D.E. BELL, R.L. KEENEY y H. RAIFFA (eds), *Conflicting objectives in decisions*. New York: Wiley.
- EDWARDS, J.E., FREDERICK, J.T. y BURKE, M.J. (1988). «Eficacy of modified CREPID SDy on the basis of archival organizational data». *Journal of Appied Psychology*. 73, 529-535.
- EINHORN, H.J. & McCOACH, W. P. (1977) . «A simple multiatributive utility procedure for evaluation». *Behavioral science*, 22, 270-282.
- EINHORN, H.J., KLEINMUNTZ, D.N. & KLEINMUNTZ, B. (1979). «Linear regresion and processtracing models of judgement». *Psychological review*, 86, 465-485.
- FISCHER, G.W. (1976). «Multidimensional utility models for risky and riskless choice». *Organizational Behavior and Human Performance*, 17, 127-146.
- GHISELLI, E.E. (1955). *The measurement of occupational aptitude*. University of California Press, Berkeley.
- GHISELLI, E.E. y BROWN, C.W. (1955). *Personnel and Industrial Psychology*. MacGraw Hill.
- GREER, O.L. y CASCIO, W.F. (1987). «Is cost accounting the answer? Comparison of two behaviorally based methods for estimating the standard desviations of job performance in dollars with a cost accounting based approach». *Journal of Applied Psychology*, 72, 588-595.
- HUBER, G.P. (1980). *Managerial decisions making*. Glenview, III: Scott, Foresman & Co.
- HUNTER, J.E. y SCHMIDT, F.L. (1982). «The economic benefits of personnel selection using psychological ability tests». *Industrial Relations*, 21, 293-308.
- KEENEY, R.L. & RAIFFA, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*. New York: Wiley.
- JANZ, J.T. y DUNNETTE, M.D. (1977). «An approach to selection decisions: Dollars and sense». En J.R. HACKMAN *et al.* (eds): *Perspectives on Performance in Organizations*. New York: McGraw Hill, 119-126.

- LEVINDKA, J., SIMONET, J.K., NEINER, A.G. y KRUSE, B. (1983). *The dollar value of JEPS at life of Georgia*. Unpublished technical report.
- LEVY-LEBOYER, C. (1992). *Evaluación del personal*. Madrid. Díaz de Santos.
- MUCHINSKY, P.M. (1994). *Psicología aplicada al trabajo: Una introducción a la Psicología Industrial y organizacional*. Bilbao. Descleé de Brouwer, S.A.
- NAYLOR, J.C. & SHINE, L.C. (1965). «A table for determining the increase in mean criterion score obtained by using a selection device». *Journal of Industrial Psychology*, 3, 33-42.
- RAJU, N.S., BURKE, M.J., y NORMAND, J. (1990). «A new approach to utility analysis». *Journal of Applied Psychology*, 75, 3-12.
- REILLY, R.R. y SMITHER, J.W. (1985). «An examination of two alternative techniques to estimate the standard desviations of job performance in dollars». *Journal of Applied sychology*, 70, 651-661.
- ROCHE, W.J. (1965). «A dollar in fixed-treatment employee selection». En L.J. CRONBACH y G.C. GLESER (eds): *Psychological testing and personnel decisions*. Urbana: University of Illinois Press.
- RORER, L.G., HOFFMAN, P.G., LA FORGE, G.E. & HSIEH, K. (1966). «Optimal cutting scores to discriminate groups of unequal size and variance». *Journal of Applied Psychology*, 50, 153-164.
- SCHULER, H. y GULDIN, A. (1991). *Methodological issues in personnel selection research*. En C.L COOPER y I.T. ROBERTSON: *International Review of Industrial and Organizational Psychology*, 6, 213-264.
- SCHMIDT, F.L., HUNTER, J.E., MCKENZIE, R.C. y MULDROW, T.W. (1979). «Impact of valid selection procedures on work-force productivity». *Journal of Applied Psychology*, 64, 609-626.
- SCHMIDT, F.L., GAST-ROSEMBERG, I., HUNTER, J.E. (1980). «Validity generalization results for computers programmers», *Journal of Applied Psychology*, 65, 643-661.
- SCHMIDT, F.L. y HUNTER, J.E. (1983). «Individual differences in productivity: an empirical test of estimates derived from studies of selection procedure utility». *Journal of Applied Psychology*, 68, 407-414.
- SMITH, M. (1986). «Selection, Where are the best prophets?» *Personnel Management*, December, pág. 63.
- STEFFY, B.D. y MAURER, S.D. (1988). «The dollar productive impact on the human resource function: Conceptualization and measurement». *Academy of Management Review*, 13, 271-286.
- TAYLOR, H.C. & RUSSELL, J.T. (1939). «The relationship of validity coefficients to the practical effectiveness of tests in selection: Discusion and tables». *Journal of Applied Psychology*, 23, 565-578.
- WEEKLEY, J.A., O'CONNOR, E.J., FRANK, B. y PETERS, L.W. (1985). «A comparison of three methods of estimating the standard desviation of performance in dollars». *Journal of Applied Psychology*, 79, 122-126.
- WROTEN, S.P. (1984). *Overcoming the futilities of utility applications: measures, models and management*. Presented at the American Psychological Association's 92nd Meeting, Toronto, august.