

EL SESGO NEGATIVO EN LA INTEGRACION DE INFORMACION PARA CALIFICAR EL DESEMPEÑO: PRIMERO HALLAZGOS¹

Federico R. León

Replicando hallazgos referidos por la literatura en Psicología Social Experimental, el presente estudio de campo encontró que un sesgo negativo prevaleció en un conjunto de conferencias ofrecidas por 8 psicólogos: la calificación para el conjunto fue significativamente menor que la obtenida para las conferencias tomadas individualmente. Esto ocurrió a pesar del hecho que aproximadamente el 50o/o de los sujetos ($N = 50$) calificaron más alto el conjunto que las conferencias y se puede explicar por la existencia de mayores diferencias entre las calificaciones del conjunto y las calificaciones individuales de las conferencias en el caso de sesgo negativo ($\text{conjunto} < \text{conferencias}$) que en el de sesgo positivo ($\text{conjunto} > \text{conferencias}$). Dado que diferencias significativas entre submuestras de sesgo positivo y negativo fueron constatadas en la calificación del conjunto mas no en la de conferencias tomadas individualmente, se concluyó que el sesgo se produce en el proceso de integración de la información para calificar el conjunto.

Replicating findings reported in the experimental social psychological literature, this field study found that a negative bias prevailed in the appraisal of a set of lectures given by eight psychologists: the mean rating for the set as a whole was significantly smaller than the mean rating for the individual lectures. This occurred despite the fact that about 50o/o of the subjects ($N = 50$) produced higher ratings for the set than the lectures and was explained by the existence of greater differences between set and lecture ratings in the cases of negative bias ($\text{set} < \text{lectures}$) than in those of positive bias ($\text{set} > \text{lectures}$). Since significant differences between subsamples of positive and negative bias were observed in the ratings for the set but not in the ratings for the lectures, it was concluded that the bias took place in the process of integrating information for rating the whole.

Introducción

En el mundo contemporáneo —tanto en países de gran avance tecnológico-industrial como en naciones en vías de desarrollo, y en sociedades capitalistas y socialistas por igual— las organizaciones absorben y controlan crecientemente el trabajo de los individuos. Sistemas formalizados son usados por ellas en la evaluación periódica del desempeño laboral de sus miembros,

1 El diseño de este estudio se hizo en el marco de un proyecto de investigación desarrollado para la Escuela de Administración de Negocios para Graduados ESAN entre marzo y octubre de 1980. La recolección y análisis de los datos, así como la redacción del manuscrito, fueron financiados por IYADES con fondos obtenidos en su Primera Conferencia Anual de Psicología del Trabajo y las Organizaciones. Un resumen del trabajo fue presentado al Segundo Congreso Peruano de Psicología (Lima, diciembre de 1981). Colaboró en los análisis Lidia Ortiz. Dirección del autor: IYADES, Apartado 274, Miraflores, Lima 18, Perú.

sean éstos obreros o diplomáticos, vendedores o coroneles, ejecutivos o clérigos, profesores o enfermeros. Tales sistemas exigen por lo general la aplicación del juicio humano, pues suelen consistir en calificaciones (ratings) del rendimiento del subordinado efectuadas por un superior jerárquico en formatos ad hoc. Al igual que en otros contextos, errores sistemáticos de juicio ocurren en el proceso de calificación, y los psicólogos han hecho esfuerzos significativos por identificar tales errores y contribuir a su eliminación.

En la década de 1970, dos tópicos, desarrollo de formatos de calificación y entrenamiento de calificadores, coparon la atención de los investigadores interesados en el mejoramiento de estos métodos de evaluación. El trabajo teórico pertinente al tema fue escaso. Ahora, sin embargo, hay un reconocimiento creciente de la necesidad de desarrollar modelos explicatorios en el área, enfocando la calificación del desempeño como un proceso psicológico y sacando provecho del sustancial cuerpo teórico y empírico que se ha acumulado en campos relacionados, v.g., los de percepción de personas, teorías implícitas de la personalidad, etc. (Dunnette y Borman, 1979; Landy y Farr, 1980). Se supone que una comprensión más cabal de la calificación del desempeño como proceso llevará a una mejoría en la tecnología de calificación, resultando en incrementos de la confiabilidad, validez, y utilidad de las calificaciones. Central en este enfoque es la formulación y verificación empírica de hipótesis concernientes a los mecanismos específicos que subyacen en la calificación. En este artículo se plantea y examina la hipótesis de que un sesgo negativo tiene lugar en el proceso de integración de información por el calificador del desempeño.

La integración de información es un proceso ubicuo en la calificación del desempeño en organizaciones, sea que el foco de la calificación esté puesto en un dominio comportamental amplio, v.g., rendimiento general en el puesto de trabajo, o en una categoría estrecha, v.g., calidad del trato con público. En cualquiera de los dos casos, el calificador debe sintetizar en un juicio global para la dimensión calificatoria de interés un sinnúmero de observaciones concretas hechas por él respecto al comportamiento del calificado en situaciones cotidianas de trabajo. Por ejemplo, la calificación del "trato con público" exige integrar varias observaciones hechas sobre el individuo en situaciones de trato con público. Hallazgos de investigación reportados en diversas áreas sugieren que un sesgo hacia la negatividad puede prevalecer en el proceso de integración. Tanto en experimentos de formación de impresiones (Kanouse y Hanson, 1972) como en estudios psicolingüísticos (Wicker, Craghead, Convis, y DeJong, 1979) se encuentra que la calificación que da un sujeto para un todo, v.g., una persona-estímulo, es más negativa de lo que cabría esperar en función del promedio de las calificaciones que el sujeto da para las partes constituyentes de ese todo, v.g., los rasgos usados en la composición de la persona-estímulo. La Tabla 1 proporciona un ejemplo.

De modo similar, en la toma de decisiones con asunción de riesgos, v.g., en apuestas y contextos similares (Kanouse y Hanson, 1972; Lynch, 1978; Shanteau, 1974, 1975), así como en la entrevista de selección de personal (Landy y Trumbo, 1976), se halla que la gente actúa como si diera un peso mayor a los costos anticipados de una decisión que a sus beneficios. La Tabla 2 presenta el mecanismo de pesaje diferencial de la información positiva y negativa que explicaría el fenómeno empírico presentado en la Tabla 1.

TABLA 1

Ejemplo de sesgo negativo: La persona es calificada menos positivamente en conjunto que rasgo por rasgo.

EVALUACION PARCIAL	EVALUACION GLOBAL
Instrucciones Use una escala de -3 a $+3$ puntos para calificar los siguientes rasgos de una mujer: Bella 3 Honestas 2 Femenina 1 Desordenada -1 Conflictiva -2 Torpe -3	Instrucciones Use una escala de -3 a $+3$ puntos para calificar a una mujer que tiene los siguientes rasgos: Bella Honestas Femenina Desordenada Conflictiva Torpe Respuesta: -1

TABLA 2

Interpretación tradicional del sesgo negativo: La información negativa recibe un peso mayor al ser integrada con la positiva.

EVALUACION PARCIAL		INTEGRACION DE LA INFORMACION	
Rasgos	Calificación	Con peso uniforme	Con peso diferencial
Bella	3	1 } 3 2 1	1 } 3 2 1
Honestas	2		
Femenina	1		
Desordenada	-1	1 } -1 -2 -3	2 } -1 -4 -6
Conflictiva	-2		
Torpe	-3		
PROMEDIO		0	-1

Hipótesis 1

Un sesgo negativo, entonces, parece dominar el proceso de integración de la información, trascendiendo áreas comportamentales y estrategias de análisis. Como quiera que no hay razones de peso para creer que la integración de información en la evaluación del desempeño difiere como proceso psicológico de la integración de información en otros contextos, se puede es-

perar que un sesgo negativo prevalezca también en la calificación del desempeño en organizaciones. El primer objetivo del estudio reportado aquí fue el de determinar empíricamente si los calificadores del desempeño en efecto evalúan al todo menos positivamente que a las partes. La literatura internacional no da indicios de que el sesgo negativo así definido haya sido verificado en situaciones de evaluación del desempeño, a pesar de la importancia práctica que tendría una confirmación de la hipótesis. Esta implicación práctica tal vez no sea inmediatamente obvia y deba ser explicitada. Si bien en las organizaciones es deseable contar con evaluaciones múltiples, que reflejen los distintos aspectos del trabajo de un individuo, v.g., cantidad de trabajo, cuidado de máquinas, etc., a la hora de tomar decisiones de personal se suele necesitar un valor único que exprese el aporte total del individuo a la organización (Schmidt y Kaplan, 1971). Este valor único suele obtenerse promediando las calificaciones múltiples ofrecidas por el supervisor para los distintos aspectos del desempeño o usando un juicio global con el cual el supervisor califica el aporte total del individuo. Si el todo tendiese a ser calificado menos positivamente que las partes, las dos formas no serían comparables: los trabajadores se beneficiarían con el promedio de evaluaciones parciales, perjudicándose con el juicio global.

Hipótesis 2

Kanouse y Hanson (1972) sugirieron que el sesgo hacia la negatividad podría no ser universal, estando más bien moderado por diferencias individuales; es decir, algunos calificadores podrían ser menos dados que otros a privilegiar lo negativo en sus calificaciones globales. Un segundo objetivo de este estudio fue el de estimar el grado de las posibles diferencias individuales existentes en sesgo negativo. Distintos métodos tendrían que usarse para cancelar los efectos organizacionales indeseables de este sesgo si hubiese una variabilidad inter-individual importante versus si no la hubiese. Por ejemplo, se podría añadir una constante a las evaluaciones globales si el sesgo fuese uniforme pero no si hubiese diferencias individuales significativas.

Hipótesis 3

Finalmente, el estudio que se reporta aquí permitió examinar otra cuestión, de índole más teórica. Tanto las investigaciones sobre el tema en otras áreas como los objetivos de investigación enunciados atrás (Hipótesis 1 y 2) se refieren a relaciones empíricas. Estas relaciones empíricas han sido interpretadas como evidencia de que un sesgo negativo domina el proceso de integración de información (véase Tabla 2). Los fenómenos observados, sin embargo, admiten una interpretación alternativa: que la distorsión esté en la calificación de componentes. Recuérdese la situación de la Tabla 1, donde el todo es calificado -1 mientras que el promedio para las partes es 0. Si se tiene en cuenta ciertos hallazgos de investigaciones recientes, que revelan cambios sistemáticos y claras distorsiones en la calificación de componentes (León, 1980, 1981), habrá que considerar verosímil que un sesgo positivo en la calificación de las partes y no uno negativo en la integración de ellas en un todo mayor, cause el fenómeno observado. La Tabla 3 explicita esta hipótesis. El tercer objetivo del estudio fue el de evaluar la validez de esta interpretación alternativa vis-a-vis la tradicional.

TABLA 3

Interpretación alternativa del sesgo negativo: La evaluación global expresaría los valores reales, mientras que las evaluaciones parciales expresadas estarían distorsionadas por un sesgo positivo (+ 1).

RASGOS	VALORES REALES (No expresados)	EVALUACIONES PARCIALES (Valores reales + 1)
Bella	2	3
Honesta	1	2
Femenina	0	1
Desordenada	- 2	-1
Conflictiva	- 3	-2
Torpe	- 4	- 3
PROMEDIO	- 1	0

Método

Contexto

Las calificaciones del desempeño por analizar fueron obtenidas en un ámbito académico: la Primera Conferencia Anual de Psicología del Trabajo y las Organizaciones, que consistió en 11 exposiciones y se desarrolló en seis noches (IYADES, 1981). Hubo ocho expositores; cinco de ellos aportaron una exposición cada uno y los tres restantes dos, en distintas fechas. Con excepción de la primera noche, hubo dos exposiciones por fecha. El público consistió en 121 personas, en su mayor parte psicólogos y funcionarios de personal, aunque también asistieron estudiantes de psicología y otros.

Formato de Evaluación

En la última fecha de la Conferencia se entregó a cada asistente una hoja de evaluación con 12 líneas por llenar, referidas a las 11 exposiciones en el orden del programa (EXPOS) y a la Conferencia Anual en su conjunto (CAC). Cruzadas con estas líneas había tres columnas referidas a dimensiones de evaluación: Calidad de la Presentación, Riqueza de la Información y Utilidad Práctica de la Información. La tarea consistió en usar una escala evaluativa de siete puntos que iba desde -3 (Pésimo) hasta + 3 (Excelente), pasando por 0 (Regular), para calificar cada exposición, así como CAC, respecto a cada dimensión.

Sujetos

De las hojas devueltas, 50 contenían calificaciones para el todo (CAC) además de calificaciones para las partes (EXPOS). De estas 50, 35 contenían información completa (tres calificaciones para cada exposición) y formaron un Grupo Completo.

Las 15 que contenían menos de 33 calificaciones para EXPOS formaron un Grupo Incompleto. En orden a examinar posibles diferencias entre los dos grupos, se calculó el promedio de calificaciones CAC a través de las dimensiones de evaluación en cada caso al interior de cada grupo. Es decir, se obtuvo un promedio CAC por sujeto. La diferencia entre las medias de estos valores en el Grupo Completo ($MCAC = 1.24$) y el Grupo Incompleto ($MCAC = 1.29$) no fue estadísticamente significativa, $F(1,49) < 1$. Las calificaciones EXPOS fueron promediadas primero verticalmente, resultando en una MEXPOS por columna para cada sujeto. El promedio de las tres MEXPOS constituyó una $MMEXPOS$ por sujeto. La diferencia entre las medias de los últimos valores en los grupos Completo ($MMEXPOS = 1.35$) e Incompleto ($MMEXPOS = 1.38$) tampoco fue estadísticamente significativa, $F(1,49) < 1$. En consecuencia, se decidió reunir ambos grupos en una sola muestra para el resto de los análisis ($N = 50$).

Mediciones de Sesgo

Se efectuó una medición de sesgo para cada sujeto en cada columna de la hoja de evaluación mediante la fórmula $SESGO = CAC - MEXPOS$. Es decir, se computó el promedio de calificaciones dadas por el sujeto a las exposiciones en una dimensión evaluativa (por ej., Calidad) y se sustrajo este promedio de la calificación dada por el sujeto a la Conferencia Anual en conjunto en la misma dimensión evaluativa (v.g., Calidad).

De esta manera, una cifra resultante con signo negativo indicaría Sesgo Negativo (Todo, CAC, menor que promedio de partes, MEXPOS), mientras que una con signo positivo indicaría Sesgo Positivo (Todo, CAC, mayor que promedio de partes, MEXPOS). Luego, se obtuvo para cada sujeto un promedio de los tres sesgos —aquellos medidos para Calidad, Riqueza, y Utilidad. La consistencia interna de los promedios de sesgo se estimó mediante el coeficiente Alfa (Cronbach, 1951), tratándose a los tres componentes del promedio— los sesgos calculados para las dimensiones de Calidad, Riqueza y Utilidad — como ítems de un test o escala. Alfa fue igual a .80, indicando una consistencia interna satisfactoria para fines de investigación.

Resultados

Diferencias Individuales en Sesgo

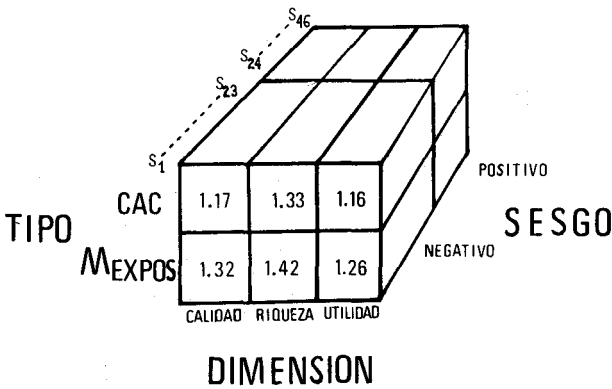
Las medidas de sesgo sirvieron para examinar la Hipótesis 2 del estudio, relativa a las diferencias individuales que podrían existir en la tendencia a cometer sesgo negativo. Grandes diferencias individuales fueron reveladas por el análisis de los sesgos promedio. De los 50 casos, 25 mostraron promedios negativos; 23 mostraron promedios positivos; y dos mostraron prome-

dio 0. La masa principal de la distribución de los promedios de sesgo se situó alrededor del punto 0, aunque una cola de ella se alargó hacia el polo negativo de la escala, causando un rango que iba desde -3.36 hasta + .63.

Prevalencia del Sesgo Negativo

Los dos casos son promedio 0 de sesgo fueron excluidos de la muestra para los restantes análisis. En orden a subdividir el total de los otros 48 casos en sub-muestras de Sesgo Negativo (SN) y Sesgo Positivo (SP) de igual tamaño, se excluyó también dos casos de SN, escogidos estrictamente al azar. En la muestra final, entonces, hubo 46 casos, agrupados en sub-muestras de 23 sujetos que cometieron sesgo negativo (SN) y 23 sujetos que cometieron sesgo positivo (SP). La Figura 1 presenta los promedios CAC y MEXPOS obtenidos en la muestra final para cada dimensión evaluativa. Puede notarse allí que los resultados son consistentes con la Hipótesis 1 del estudio, pues revelan relaciones empíricas iguales a las de la literatura revisada: calificaciones del todo que son menores que el promedio de calificaciones para las partes.

FIGURA 1
Promedios CAC y MEXPOS obtenidos en cada dimensión evaluativa y esquema del modelo analítico usado.



A fin de determinar la significación estadística de los resultados, las calificaciones del desempeño ofrecidas por los 46 sujetos fueron sometidas a un análisis mixto de la varianza con mediciones repetidas según el diseño SPF-2.32 de Kirk (1968), que también se representa en la Figura 1. La variable intersujetos fue Sesgo (SN, n = 23; SP, n = 23) y las variables intrasujetos fueron dimensiones de evaluación (Calidad, Riqueza, Utilidad) y Tipo de Calificación (CAC, MEXPOS). La variable dependiente fue siempre la escala de siete puntos usada por los sujetos en la evaluación del desempeño; en el caso de CAC ella es una calificación simple y en el caso de MEXPOS es un promedio de calificaciones. Adviértase que varios supuestos del diseño SPF son violados en este análisis: los de asignación al azar de casos a celdas SN y SP, asignación aleatoria de secuencias de calificación a los sujetos, y, probablemente, igualdad y simetría de las matrices de varianzas y covarianzas. El enfoque analítico escogido, sin embargo, es preferible a alternati-

vas tales como computar tests t redundantemente para los múltiples subconjuntos de datos generados por el diseño. Mientras que la redundancia de los tests t sería inmanejable en un contexto como el presente a no ser que se decidiera sacrificar poder en gran medida, las violaciones del modelo SPF pueden contrarrestarse eficientemente usando grados de libertad conservadores en el cálculo de probabilidades para las ratios F .

De las nueve F s resultantes del análisis mixto, tres fueron significativas. En lo que sigue, los grados de libertad indicados para ellas ya están corregidos por la fórmula conservadora de Geisser-Greenhouse (Kirk, 1968, p. 303). En primer lugar, hubo un efecto principal significativo de Sesgo, $F(1,44) = 5.40$, $p < .05$. Los sujetos con promedios negativos de sesgo ofrecieron calificaciones ($M = 1.01$) medio punto más bajas que las de quienes cometieron sesgo positivo ($M = 1.54$). De aquí se puede concluir que los primeros son más severos y/o los últimos son más benevolentes al calificar. En segundo lugar, hubo un efecto principal significativo de Tipo de Calificación, $F(1,44) = 6.55$, $p < .05$. Las calificaciones del todo ($MCAC = 1.22$) fueron significativamente menores que los promedios de calificación de partes ($MMEXPOS = 1.33$). Esto, confirmando definitivamente la Hipótesis 1 del estudio, no contradice a la Hipótesis 2. Nótese que el número de personas que cometen sesgo negativo puede no ser mayor que el de las personas que cometen sesgo positivo, pero si las primeras comenten sesgos "más grandes", habrá una prevalencia cuantitativa del sesgo negativo en el agregado de casos.

Locus del Sesgo

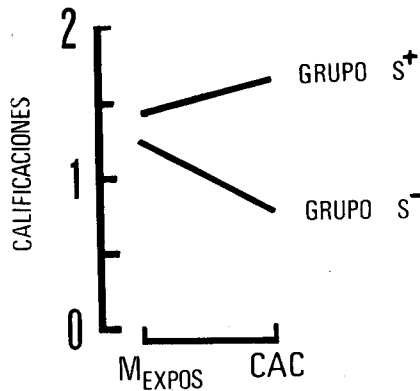
El tercer resultado del ANOVA fue muy interesante porque sirvió para evaluar la Hipótesis 3 del estudio. Se trata de una interacción significativa entre Sesgo y Tipo de Calificación, $F(1,44) = 46.61$, $p < .01$, que se grafica en la Figura 2. Esta interacción tiene dos aspectos. Uno concierne obviamente a los efectos opuestos de Tipo de Calificación a cada nivel de Sesgo. Los análisis simples indican efectos significativos tanto en el Grupo SN, $F(1,44) = 30.92$, $p < .01$, como en el Grupo SP, $F(1,44) = 8.74$, $p < .01$; pero mientras que ellos ocurren en una dirección en el primer grupo ($MCAC = .78$; $MMEXPOS = 1.25$), vienen en dirección contraria en el segundo ($MCAC = 1.66$; $MMEXPOS = 1.41$). Este resultado, que en la Figura 2 se expresa en la inclinación de la curva SN hacia la derecha y de la curva SP hacia la izquierda, estaba predeterminado, pues el criterio de formación de los grupos SN y SP fue, respectivamente, que el todo fuese calificado menos positivamente que el promedio de partes y viceversa. Lo que no estaba predeterminado es el segundo aspecto de la interacción: que la diferencia en MEXPOS entre los grupos SN ($M = 1.25$) y SP ($M = 1.41$), $F(1,44) < 1$, no fuera estadísticamente significativa, en contraste con la diferencia en CAC ($M_{SN} = .78$, $M_{SP} = 1.66$; $F(1,44) = 14.26$, $p < .01$), que sí lo fue. En la Figura 2 ello se refleja en la distancia de los promedios SN y SP entre sí, mayor cuando se refieren a CAC que cuando se refieren a MEXPOS. Esto quiere decir que las diferencias al calificar el desempeño que existen entre quienes cometen sesgo negativo y quienes cometen sesgo positivo ocurren principalmente en la calificación del todo, no en la calificación de las partes.

TABLA 4

Número de sujetos en cada grupo que ofrecieron por lo menos una calificación positiva/ neutra / negativa (n) y proporción promedio de cada tipo de calificación en el total de respuestas del sujeto (M).

GRUPOS	CALIFICACIONES					
	Positivas		Neutras		Negativas	
	n	M	n	M	n	M
SESGO POSITIVO	23	.85	17	.11	12	.12
SESGO NEGATIVO	23	.73	17	.25	9	.18

FIGURA 2
Promedios M_{EXPOS} y CAC obtenidos por los grupos Sesgo Positivo y Sesgo Negativo.



A fin de identificar posibles diferencias en la calificación de partes que pudieran existir sin haber sido detectadas en los análisis previos, se analizó la frecuencia de calificaciones EXPOS positivas, neutras, y negativas en los Grupos SN y SP. La Tabla 4 presenta los resultados. En ambos grupos todos los sujetos ofrecieron calificaciones positivas pero la proporción de esta clase de respuestas por sujeto fue significativamente mayor en el grupo

SP que en el grupo SN, $t(44) = 2.39$, $p < .05$. La relación se invierte en el caso de las respuestas neutras produciendo una diferencia también significativa, $t(32) = -6.09$, $p < .01$. En el caso de las respuestas negativas hubo mayor proporción por sujeto en el grupo SN, pero la diferencia no fue estadísticamente significativa. Finalmente, se evaluó la calidad de las calificaciones EXPOS mediante mediciones de halo, restricción de rango, y acuerdo interjueces. El halo fue medido en cada sujeto por la desviación estándar de las calificaciones de Calidad, Riqueza, y Utilidad ofrecidas; a mayor D.E., menor halo. Para medir la restricción de rango se computó en cada sujeto la desviación estándar de las calificaciones EXPOS asignadas en una dimensión, y luego el promedio de las D.E.s a través de dimensiones; a mayor D.E. promedio, menor restricción de rango. La medición del acuerdo interjueces exigió computar a través de sujetos, primero la desviación estándar de las calificaciones asignadas a una EXPO en una dimensión, y luego el promedio de las D.E.s a través de EXPOS y dimensiones; a mayor D.E. promedio, menor acuerdo. Los grupos SN y SP no mostraron diferencias significativas en halo, restricción de rango, o acuerdo interjueces.

Discusión

La Hipótesis 1 del presente estudio, sugerida por los hallazgos de la literatura experimental sobre formación de impresiones y toma de decisiones bajo incertidumbre, afirmaba que en la evaluación del desempeño el todo resultaría calificado por debajo del promedio de partes. Esta hipótesis fue confirmada por los resultados del estudio reportado aquí: la Primera Conferencia Anual de Psicología del Trabajo y las Organizaciones en conjunto mereció calificaciones significativamente más bajas que las del promedio de las exposiciones en las que consistió. La similitud de nuestros resultados con los de la literatura revisada, sin embargo, termina allí pues al procederse a evaluar la Hipótesis 2, relativa a diferencias individuales en sesgo, se encontró que cerca del 50o/o de los sujetos cometía sesgo positivo, algo que probablemente no ocurrió en los experimentos. La prevalencia del sesgo negativo encontrada en el presente estudio, entonces, sólo concierne a la intensidad del sesgo, no a su extensión; es decir, se refiere a la magnitud absoluta del sesgo, mayor en el negativo que en el positivo, pero no a la frecuencia de ocurrencia de uno y otro.

La tercera pregunta formulada se refirió a los procesos psicológicos subyacentes en los fenómenos observados: ¿se explican éstos por un sesgo negativo en la integración de información o por uno positivo en la calificación de partes? Los resultados del estudio parecen contradecir la segunda explicación (nuestra Hipótesis 3). Si el todo fuese evaluado menos positivamente que las partes debido a que la evaluación de partes está inflada mientras que la evaluaban del todo refleja los valores reales, se tendría que esperar calificaciones parciales mayores entre sujetos que cometen sesgo negativo que entre quienes cometen sesgo positivo. Este no fue el caso del presente estudio, donde la gente que cometió sesgo negativo produjo una menor proporción de calificaciones positivas que los sujetos del grupo Sesgo Positivo, y donde las diferencias en el promedio de calificaciones EXPOS entre los dos

grupos no fueron significativas. En cambio, las calificaciones globales (CAC) fueron significativamente más altas en el grupo Sesgo Positivo que en el grupo SN, sugiriendo que es en el proceso de integración de la información para calificar el todo que se produce el sesgo. El diseño del estudio, sin embargo, no permite afirmar concluyentemente que sea un proceso de pesaje diferencial de la información positiva versus la negativa, como postula la interpretación tradicional, lo que causa que el todo sea calificado menos positivamente que las partes. Podría ser que la calificación global nazca por selección de puntajes representativos o alguna otra forma no aditiva de procesamiento de la información, en lugar de originarse en una real síntesis de todos los elementos que supuestamente intervienen en su formación.

El sesgo negativo empírico, ocurriendo sistemáticamente en los experimentos citados, demostró robustez al recurrir en un contexto natural de evaluación del desempeño. No obstante ello, resulta difícil de precisar la generalizabilidad de los resultados del presente estudio a las situaciones organizacionales más comunes. En primer lugar, el contexto usual de evaluación del desempeño en organizaciones impone una relación todo-partes distinta de la que operó aquí. Mientras que aquí las partes fueron exposiciones concretas de personas y el todo fue una entidad abstracta, impersonal, en la situación tradicional de evaluación del desempeño en organizaciones las partes son aspectos del desempeño individual y el todo es el rendimiento global de una persona. En segundo lugar, el formato de evaluación empleado aquí fue una escala gráfica bastante simplificada; en las organizaciones de trabajo están siendo usados formatos de distribución forzada, escalas comportamentalmente ancladas, evaluación por objetivos, y otros métodos evaluativos complejos. Claro que los resultados reportados aumentan la probabilidad subjetiva de encontrar en el contexto convencional la prevalencia de un sesgo negativo. Las diferencias de contexto, sin embargo, son suficientes para exigir la verificación de la Hipótesis 1 en ámbitos organizacionales más comunes antes de dar por sentada su validez universal.

¿Cuáles son entonces las implicaciones prácticas de los resultados del presente estudio y a qué realidades se aplican? Los resultados parecen generalizables a situaciones de evaluación de programas docentes y también a casos de evaluación del desempeño de subsistemas organizacionales tales como secciones o departamentos en conjunto, en particular si se usan escalas gráficas simples. En estos casos saldrá perjudicado el objeto de evaluación al usarse una calificación global y se beneficiará con una evaluación basada en el promedio de sus componentes, es decir, los desempeños individuales.

Trascendiendo estas preocupaciones prácticas inmediatas no hay duda que valdría la pena explorar sistemáticamente los correlatos del sesgo negativo con vistas a esclarecer sus causas. En el presente estudio se halló que quienes cometían sesgo negativo se caracterizaban por ofrecer menor frecuencia de calificaciones positivas para las exposiciones que quienes cometían sesgo positivo. Cabe preguntarse si no declinará la frecuencia de sesgos negativos al incrementarse la proporción de elementos positivos en la situación evaluativa, y vice versa, si no aumentará la frecuencia de sesgos positivos al reducirse la proporción de elementos negativos y neutros. En esta perspectiva, es interesante notar que el estudio descrito aquí involucró pocos elementos negativos en comparación con los experimentos reportados en la literatura. Nótese las distribuciones de la Tabla 4. Añádase el

que los promedios de calificaciones EXPOS fueron positivos en todos los casos y sólo tres exposiciones bordearon el promedio 0, es decir, elicitaban muchas evaluaciones negativas. En cambio, en los experimentos revisados siempre hubo una intención de presentar a los sujetos distribuciones simétricas de información positiva y negativa. Esto sugiere como probable que en los experimentos haya habido una mayor frecuencia de sesgos negativos que positivos y conduce a la hipótesis de una prevalencia de sesgos *positivos* en situaciones de evaluación donde los buenos rendimientos dominan ampliamente el cuadro. Esta hipótesis, así como aquellas que tomen en cuenta la posibilidad de que el sesgo responda a disposiciones personales más o menos permanentes, deben ser motivo de estudios específicos.

REFERENCIAS

CRONBACH, L.J. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, 16, 297 - 334.

DUNNETTE, M.D., y BORMAN, W.C. Personnel selection and classification systems. *Annual Review of Psychology*, 1979, 30, 477 - 525.

IYADES (Instituto de Psicología Aplicada al Desarrollo Económico y Social). *Folleto de la Primera Conferencia Anual de Psicología del Trabajo y las Organizaciones*. Lima: Autor, 1981.

KANOUSE, D.E., y HANSON, L.R., Jr. Negativity in evaluations. En E.E. Jones (Ed.), *Attribution: Perceiving the causes of behavior*, Morristown, N.J.: General Learning Press, 1972.

KIRK, R.E. *Experimental design: Procedures for the behavioral sciences*. Belmont, Calif.: Brooks/Cole, 1968.

LANDY, F.J., y FARR, J.L. Performance rating. *Psychological Bulletin*, 1980, 87, 72 - 107.

----- y TRUMBO, D.A. *Psychology of work behavior*. Homewood, Ill.: Dorsey Press, 1976.

LEON, F.R. Dynamics of work decision-making: Interactions among outcome valences. *Interamerican Journal of Psychology*, 1980, 14, 31-42.

----- The role of positive and negative outcomes in the causation of motivational forces. *Journal of Applied Psychology*, 1981, 66, 45-53.

LYNCH, J.G., Jr. Why additive utility models fail as descriptions of choice behavior. *Journal of Experimental Social Psychology*, 1979, 15, 397 - 417.

- SCHMIDT, F.L., y KAPLAN, L. B. Composite vs. multiple criteria: A review and resolution of the controversy. *Personnel Psychology*, 1971, 24, 419 - 434.
- SHANTEAU, J. Component processes in risky decision making. *Journal of Experimental Psychology*, 1974, 103, 680 - 691.
- An information-integration analysis of risky decision making. En M. F. Kaplan y S. Schwartz (Eds.), *Human judgment and decision processes*. New York: Academic Press, 1975.
- WICKER, F.W. CARGHEAD, J.C., CONVIS, C.L., y DEJONG, R.H. Negative dominance in hedonic integration on of noun-noun pairs. *Motivation and Emotion*, 1979, 3, 129 - 141.