

LA EVALUACION DEL DESEMPEÑO: CONCEPTO, CRITERIOS Y METODOS

Mikel Gorriti Bontigui

Responsable de Recursos Humanos de la Dirección de Función Pública del Gobierno Vasco

1. Informe de la Comisión para el Estudio y Preparación del Estatuto Básico del Empleado Público.	2
2. Posibles usos de la Evaluación del Desempeño (ED). Beneficios de no hacerla.	3
3. La ED desde un punto de vista conceptual, los planteamientos taxonómicos de los grandes enfoques.	5
4. El Análisis del Trabajo como fuente de contenidos para la ED.	10
5. Metodologías de Evaluación del Desempeño.	13
Referencias.	16

1. INFORME DE LA COMISION PARA EL ESTUDIO Y PREPARACION DEL ESTATUTO BASICO DEL EMPLEADO PUBLICO

Con fecha 16 de septiembre de 2004 se constituyó la Comisión para el Estudio y Preparación del Estatuto Básico del Empleado Públicoⁱ, encargada de elaborar las orientaciones y criterios básicos que deben guiar la creación de dicho Estatuto, el cual por ser texto básico, incidirá de forma vinculante en nuestras actividades como profesionales de los RRHH en las Administraciones Públicas (AAPP).

Por lo que respecta a la Evaluación de Desempeño (ED) muchas son las

- Evaluación de Desempeño y Puesto de Trabajo:
Para el Informe de la Comisión, el Puesto de Trabajo es la unidad organizativa sobre la que deben realizarse las intervenciones en RRHH que, a su juicio, son las siguientes: a) definir los perfiles de idoneidad de las personas para el desempeño de las funciones; b) la asignación y supervisión de tareas; c) la evaluación de desempeño de los empleados; d) la fijación de retribución, o mejor dicho, de una parte de ella; e) una de las modalidades de carrera profesional, aquella que implica cambio de puesto de trabajo.
- Evaluación de Desempeño y Carrera:
Para el informe es primordial que el futuro Estatuto vincule la carrera con la ED, concebida la primera como: carrera horizontal (en el puesto de trabajo) y la que implica movilidad (cambio de puesto). El informe entiende des-motivador y más formal que eficaz, ligar la carrera a la antigüedad. Mejorar en el puesto de

consideraciones que dicho texto hace, tantas que podríamos decir que, de aprobarse como está planteado, muchos van a ser los cambios que van a ocurrir en las AAPP. Este apartado pretende repasar, de forma sucinta, estas referencias y analizar la trascendencia que desde un punto de vista práctico pueden tener para nosotros.

ⁱ BOE, número 228, de 21 de Septiembre de 2004, Orden APU/3018/2004. Página 31644

trabajo sin moverse de él (ganar más por profundizar en el puesto), así como ir a otro puesto, debe ser consecuencia de una evaluación del desempeño; de estimar en qué medida se han desempeñado bien las tareas y se han satisfecho las responsabilidades enunciadas en las funciones. Por otro lado, el informe distingue el aspecto motivador de la ED al ser esta una condición para la carrera, sea esta una mejora económica o de movilidad.

- Evaluación de Desempeño y Retribución:
El Informe afirma que el futuro Estatuto debería establecer la vinculación expresa entre la ED y los complementos de productividad o cualquier otro elemento retributivo ligado al rendimiento, y, en su caso, con la no percepción del mismo. Es decir, pagar más a quien mejor lo hace, lo cual, no cabe duda, es bastante revolucionario y problemático para lo que estamos acostumbrados en las AAPP.

- **Evaluación de Desempeño y Formación:** Para el Informe la ED es un elemento esencial con el que detectar las necesidades de formación del puesto, entendidas estas como la distancia que hay entre lo que un trabajador/a sabe y lo que debe de saber para realizar de forma eficaz las tareas de su puesto de trabajo. Podría añadirse que, además, la ED es inevitable si se quiere constatar si la formación ha servido para algo; si además de aprendizaje y retención de los conocimientos y destrezas, ha ocurrido la necesaria transferencia al puesto para que aquella sea rentable.
- **Evaluación de Desempeño y Selección:** El informe reconoce la importancia de la ED para dar naturaleza de predictivas a variables con las que se selecciona a los empleados públicos (investigación), y también porque permiten tener un elemento objetivo a la hora de evidenciar ineficacias que puedan implicar remoción, tanto en el caso de libres designaciones como en el de titulares.
- **Evaluación de Desempeño y Modernización:** Afirman que mediante la ED las AAPP pueden cuestionarse, analizar, legitimar y concluir cuál es su criterio de éxito como organización, lo cual, obviamente es diagnóstico y condición para su modernización.
- **Evaluación de Desempeño y los Profesionales de los RRHH:** En tanto que la ED exige, por su complejidad, el concurso de profesionales cualificados, exige desarrollar en mayor medida los departamentos o servicios de Recursos Humanos. Para el Informe, invertir en ellos, es una de las mejores decisiones que se pueden tomar.
Por último citar un literal del Informe que nos da cuenta de la importancia que la Comisión da a las ED: “Siendo este un factor decisivo de modernización de nuestro sistema administrativo, entendemos que el Estatuto Básico debería establecer la obligatoriedad, para todas las Administraciones, de organizar sistemas de evaluación del desempeño de sus empleados.” (pág. 108).

2. POSIBLES USOS DE LA EVALUACION DEL DESEMPEÑO. BENEFICIOS DE NO HACERLA

Desde un punto de vista general cuatro son los usos para los que puede usarse la ED (Cascio, 1991):

- **Administrativos:** promoción, remoción, remuneración, carrera, formación, etc.
- **Conocimiento del dominio del desempeño:** conocer y delimitar qué significa hacer bien un trabajo.
- **Validación de Inferencias selectivas o formativas:** el desempeño como criterio.
- **Feed-back,** mejora y motivación.

Desde otro punto de vista los propósitos para los que se puede hacer una ED son los siguientes (Viswesvaran, 2001):

- **Evaluación entre personas:** salarios, promociones.
- **Evaluación intra-personas:** identificar potencialidades y flaquezas (formación, motivación).
- **Mantenimiento del sistema:** Investigación (conocimiento del criterio), planificación de RRHH, organización.

- Documental: para justificar decisiones e intervenciones ante recursos legales.

Salvo honrosas excepciones (ya citadas por el Dr. Salgado en la ponencia de la mañana), pocas o muy pocas son las Administraciones que llevan acabo alguna forma de ED. ¿Significa esto que ninguno de los propósitos arriba citados es hoy necesario para las AAPP? No, de hecho la mayoría de ellas tiene necesidad de identificar criterios para salarios, para promociones, para diseñar la formación, etc., el problema es que lo hace desde una inferencia a-sistemática y sin método de lo que es la ED. Esta es una de las tesis de la ponencia: *todos los que nos dedicamos a los RRHH en las Administraciones Públicas hacemos ED de alguna manera, de lo que se trata es de sistematizar el proceso; de darle una visión y método más profesional de manera que ésta se haga de forma más eficaz y justa*. La necesidad de la ED no está en cuestión, de hecho es inevitable, lo que se ve necesario es incorporarla como una práctica cotidiana y profesional de las actividades naturales de los RRHH en las AAPP. Esto no será posible si no se dan dos condiciones que, a mi juicio, tienen que ver con el miedo a implantarla:

- Confiar en que existen métodos y técnicas fiables y válidas para hacerlas.
- Reconocer que la ED tiene que tener trascendencia administrativa, en el sentido arriba explicado (salarios, promociones, remociones, formación, carrera).

En la investigación de la Psicología del Trabajo y de las Organizaciones (PTO), a la ED se le conoce como *el criterio*; como una variable lo suficientemente importante como para ser predicha, mejorada o corregida. Una variable que necesita ser explicada y definida

(la variable dependiente). Desde esta concepción, sólo cuando se sabe qué es hacer bien el trabajo se puede empezar a definir o a inferir sobre todas las demás actividades posibles de RRHH. La cuestión es: en ausencia de de este conocimiento, ¿qué justificación tenemos para inferir perfiles selectivos o contenidos formativos, para definir salarios, para diseñar la carrera, etc., si no tenemos claro qué significa hacer bien un trabajo? ¿Cómo se puede saber qué exigir o medir en un proceso selectivo, si no conozco qué significa hacerlo bien? ¿Cómo puedo proponer mejoras en la formación o reconducir actuaciones de los trabajadores/as, si no sé qué significa hacerlo bien?. Es como poner el carro delante de los bueyes (Guion, 1997).

Parece como si en las AAPP hubiese un acuerdo general y tácito de que la ED no es imprescindible porque el hecho de que cada año haya presupuestos exime de la necesidad de demostrar la eficacia de la Administración y por ende, de diferenciar entre buenos y malos trabajadores/as. En realidad todos sabemos quién trabaja bien y quien lo hace peor, y no hay mayor discriminación que no discriminar. Por otro lado, aceptar que todo el mundo realiza el trabajo bien, es tan falso como negar que todo el mundo se diferencia en la forma como hace las cosas. No nos engañemos, a partir del conocimiento no sistemático del desempeño de la gente, nos pasamos el día tomando decisiones sobre las personas que no queremos hacer evidentes porque nos da beneficios:

- Nos evita tener que ser asertivos cuando hay que pasar el mal trago de decirle a alguien que hace mal su trabajo y que algo le va a pasar por ello.
- Nos evita tener que establecer compromisos de cumplimiento que nos estructuren el tiempo.

- Nos evita tener que tomar decisiones retributivas que paguen más a los que mejor hacen y romper el “café para todos”.
- Nos evita evidenciar las ineficacias de los programas de formación que nada o poco tienen que ver con el trabajo para el que se forma, simplemente porque se desconocen las destrezas necesarias para hacer un trabajo tan bien como lo describe su mejor desempeño.
- Nos evita tener que probar que la formación se ha transferido al puesto de trabajo, que éste ha mejorado en eficacia y que la inversión ha sido rentable. Lo cual es mucho trabajo.
- Nos permite raseros generales de igualdad a la hora de la movilidad entre puestos, lo cual tiene más posibilidades de éxito social pero un coste de eficacia.
- Nos evita tener que ejercer el liderazgo sobre compromisos de cumplimiento de plazos, productos o indicadores, que, en definitiva, es trabajo para el que ejerce la jefatura.
- Nos evita cuestionarnos el status quo, al no tener un referente de cuáles son las características de producto o servicio que exigen los ciudadanos.
- Nos impide gestionar los RRHH por razones de oportunidad, de estrategias coyunturales o excesivamente contingentes, que van más allá de las propias exigencias del puesto.
- Etc.

Cualquier Administración que no mide su desempeño, impide el natural cambio gradual y está abocada a cambios bruscos, cualitativos y traumáticos producidos por cambios políticos, tecnológicos, etc. Asimismo, toda Administración que no mide su desempeño tiene más dificultades a la hora de asentar o hacer estables los cambios que se han demostrado eficaces, y así circunscribir el inevitable cambio electoral a su auténtica naturaleza: estrategia política, orientación del servicio público etc.

3. LA EVALUACION DEL DESEMPEÑO DESDE EL PUNTO DE VISTA CONCEPTUAL. LOS PLANTEAMIENTOS TAXONÓMICOS DE LOS TRES GRANDES ENFOQUES

Existen distintas concepciones de lo que es la ED:

- Society for Industrial and Organizational Psychology (SIOP, 1997): Eficacia y valor de una conducta del trabajo y sus efectos.
- SIOP (2003): Medida de una ejecución o conducta laboral relevante.
- Viswesvaran (2001): Un comportamiento organizacional evaluable y que está bajo el control del propio trabajador.

Todas estas definiciones tienen en común la concepción conductual o comportamental de

lo que se evalúa, la relevancia para la organización en la que se hace y que dicha conducta está bajo el control del ejecutor. Según esto, podríamos concluir que Evaluación del Desempeño (ED) es el proceso mediante el que se mide y valora un comportamiento organizacional relevante, bajo el dominio del propio ejecutor o del grupo objeto de evaluación. Esta distinción entre ejecución individual o de grupo depende de los fines de la evaluación. Ambas son posibles.

En cuanto a los criterios para saber si una ED es válida, hay acuerdo general en que este

juicio se debe de hacer en función de los siguientes criterios:

- *Relevancia:* Si lo que se está midiendo, realmente, es un comportamiento importante para la organización en el sentido de que una mala ejecución u omisión, tiene consecuencias graves para ella.
- *Fiabilidad:* si la medida utilizada es consistente, estable o, sistemáticamente, distintos evaluadores evalúan de igual manera al mismo/a trabajador/a en el mismo comportamiento y para el mismo periodo de tiempo.
- *Discriminación:* Si permite que los evaluados se ordenen en función de la medida. La medida genera variabilidad. No valen las medidas que no diferencian a los/as trabajadores/as.
- *Practicidad:* Definida en función de sus posibilidades de medida: claridad en la definición de lo que se mide, costo, tiempo, etc.

En la última década y fruto de distintas investigaciones que tienen que ver con la relevancia de la ED, se ha concluido que existen tres ámbitos o tipos de desempeño para cualquier organización y para cualquier trabajo: el desempeño referido a tarea, el desempeño contextual y las conductas contra-productivas. Las primeras se refieren a comportamientos directamente exigidos para la ejecución de las tareas, las segundas se refieren a comportamientos de naturaleza social o psicológica, que actúan como catalizadores del desempeño de tarea facilitándolo o dificultándolo. Por último, las conductas contra-productivas son aquellas que son

contrarias a los legítimos intereses de la organización (robos, abusos de alcohol y drogas, uso de recursos laborales para fines personales, etc.).

Antes de entrar en las descripciones de estas tipologías, conviene señalar un estudio en el que se concretaron las dimensiones relevantes para la ED independientemente de las tipologías citadas. Aunque existen otras taxonomías, la más general es la de Viswesvaran et al. (1996) que mediante estudios factoriales y partiendo de una hipótesis léxica de 486 términos (todos los términos que en inglés se referían a desempeño laboral) identificó las siguientes dimensiones relevantes para evaluar cualquier desempeño:

- Evaluación global del trabajo (suma de todas las que vienen a continuación).
- Productividad (cantidad de trabajo producido).
- Esfuerzo, cantidad para un buen producto.
- Conocimiento.
- Competencia interpersonal (capacidad para empezar y mantener relaciones interpersonales).
- Competencia administrativa (cómo se coordina con otros para la ejecución de su trabajo).
- Calidad de sus productos (grado en el que el producto o servicio satisface la necesidad del cliente o un estándar de la organización).
- Comunicación (competencia en la transmisión de contenidos).
- Liderazgo (conseguir que un grupo trabaje).
- Aceptación de reglas (comportamientos que manifiestan el compromiso con la organización y la transmisión de sus valores).

Debe tenerse en cuenta que el valor principal de esta taxonomía es nomológico, es decir, identifica una red de dimensiones y sus relaciones, cuyos contenidos son comportamientos propios de cada trabajo. Viveswaran (1996) concluyó que el desempeño es una estructura multidimensional formada por un factor general de desempeño en el ápice y por otros sub-factores menores, jerárquicamente relacionados.

Desempeño de Tarea

El desempeño de tarea es la clásica concepción de ejecución; la que se refiere a los comportamientos que generan el producto o servicio que satisface las responsabilidades que asume la organización. El principal predictor de la ejecución de tarea es la aptitud

cognitiva formada por los conocimientos del puesto, las destrezas, entendidas como formas estandarizadas de resolver problemas, y los hábitos, entendidos como respuestas características (Borman, et al., 2001).

En la siguiente tabla se exponen resumidas las dimensiones que fruto de estudios factoriales y otras metodologías formales, se han concluido relevantes para evaluar el desempeño de tarea.ⁱⁱ

ⁱⁱ Es razonable observar que de las dimensiones expuestas en la tabla, no todas se refieren a desempeño de tarea propiamente dicho y que algunas de ellas pueden considerarse precursoras o condiciones del mismo (destrezas de supervisión y liderazgo, comportamientos interpersonales, etc). Incluso muchas de ellas podrían ser de lo que llamamos desempeño contextual. En sentido estricto las referidas a la ejecución de tarea pura serían las marcadas en negrilla. La diferencia entre variable independiente y dependiente no es diáfana en el estudio del desempeño.

Autor/es	Naturaleza de la medida	Dimensiones
Toops (1944)	Distinguiendo entre adecuación: calidad y ausencia de errores, y cantidad: volumen de producto	<ul style="list-style-type: none"> • Unidades de producción • Calidad del trabajo. • Tiempo de permanencia • Destrezas de supervisión y liderazgo.
Wherry (1957)		<ul style="list-style-type: none"> • Resultados • Calidad • Tiempo perdido en la ejecución • Cambio de trabajos • Tiempo de formación o promoción • Satisfacción
Bernardin & Beatty (1984)	Definen la ejecución como los registros de resultados producidos en un puesto durante un periodo de tiempo. Aunque la ejecución es la combinación de varios predictores, la ejecución sólo puede medirse en términos de resultados. De las 6 características que propusieron, afirmaron que el entendimiento de dichas dimensiones significaba la relación entre ellas.	<ul style="list-style-type: none"> • Calidad • Cantidad • Tiempo • Costo-eficacia • Necesidad de supervisión • Impacto interpersonal
Campbell (1990)	Describe 8 dimensiones pero avisa que la correlación entre las 8 es baja por lo que todo intento de ahondar en ellas sería ineficaz para descubrir esa estructura latente.	<ul style="list-style-type: none"> • Eficacia en tareas específicas. • Eficacia en tareas no específicas (para todos) • Comunicación oral y escrita. • Esfuerzo demostrado (en acabar una tarea) • Disciplina personal sostenida • Facilitación de ejecución pares y equipos • Supervisión • Gestión o administración.
Murphy (1989)	Entiende la ejecución como el compendio de cuatro dimensiones	<ul style="list-style-type: none"> • Conductas que hacen perder el tiempo • Ejecución de las tareas o roles definidos para el puesto. • Interpersonales: de ayuda y de equipo. • Conductas destructivas: violencia, robos, etc.

A diferencia de las otras tipologías de desempeño, las referidas a tareas varían con el puesto, sin embargo las referidas al desempeño contextual y las conductas contra-productivas son generales (Borman et alt., 2001).

Desempeño Contextual

Como se ha dicho más arriba, el desempeño contextual trata de variables moderadoras referidas al contexto organizacional, social y psicológico de la ejecución, que actúan como catalizadoras facilitándola, dificultándola, incluso inhibiéndola. Variables que pueden

actuar tanto a nivel individual como grupal. Para Borman y Motowidlo (1993) las razones por las que existen este tipo de variables son: a) porque definen el ambiente de trabajo, b) porque son comunes a la mayoría de los trabajos de una organización, c) porque son más motivadoras que cognitivas, c) porque aunque deseables, son más implícitas que explícitas.

La principal taxonomía de desempeño contextual es de Borman et alt. (2001) y consta de tres dimensiones con sus sub-dimensiones:

Dimensión	Definición	Sub-dimensiones
<i>Ayuda a otros</i>	Aportar ayuda a compañeros de trabajo ofreciéndoles sugerencias, dándoles conocimientos o destrezas necesarias, haciendo directamente lo que tienen que hacer o dándoles apoyo emocional para solucionar sus problemas personales. Informándoles de cosas que pueden ser de su interés, o que deberían saber y poniendo los intereses del grupo por delante de los intereses personales. Mostrar consideración, cortesía y tacto en las relaciones así como demostrar motivación y confianza.	<ul style="list-style-type: none"> • Cooperación • Cortesía • Motivación: (aplaudir los éxitos ajenos, bromear en tiempos de adversidad, mostrar confianza en sus capacidades para triunfar, ayudarles para superar las dificultades).
<i>Vinculación organizacional</i>	Representar a la organización eficazmente defendiéndola y expresando satisfacción y lealtad más allá de momentos o circunstancias. Compromiso con la misión de la organización y sus objetivos asumiendo los roles y procedimientos de la misma, y proponiendo sugerencias.	<ul style="list-style-type: none"> • Representación • Lealtad • Compromiso
<i>Desempeño más allá del puesto</i>	Trabajar más de lo que se le pide, no importando las dificultades. Tomar la iniciativa para hacer todo lo que haya que hacer con el fin de conseguir los fines de la organización, y no sólo lo que le corresponde. Desarrollar el propio conocimiento y destreza aprovechando las oportunidades que da la organización o asumiendo este esfuerzo a costa de su propio tiempo y recursos.	<ul style="list-style-type: none"> • Persistencia. • Iniciativa. • Auto-desarrollo.

Así como el principal antecedente del desempeño de tarea es la aptitud cognitiva (conocimiento, destreza y hábito), el principal antecedente del desempeño contextual es la personalidad. Así, un estudio de Miller, Griffin y Hart (1999) realizado con empleados

públicos encontró una correlación de .42 (N=104) entre el factor Concienciaⁱⁱⁱ de personalidad (disciplina, sostenimiento de esfuerzo, responsabilidad, etc) y el desempeño contextual.

Conductas Contra-Productivas

Una conducta contra-productiva (CCP) es todo comportamiento intencional de un miembro de la organización, contrario a los legítimos intereses de ésta (Gruys y Sackett, 2003). A diferencia de los otros dos modelos, la característica de este tipo de desempeño es que no es deseable, pero por desgracia ocurre, y es pertinente evaluarlo si se concibe éste como todo comportamiento organizacional relevante, evaluable y bajo el dominio del/la trabajador/a.

Existe un estudio reciente sobre este tipo de comportamientos que ha pretendido identificar la estructura de estas conductas, así como cuáles de ellas son las más frecuentes o fáciles de incurrir (Gruys y Sackett, 2003). Para identificar la estructura se identificaron de la literatura 256 CCP, y una vez eliminados los enunciados redundantes quedaron 66. Mediante análisis factoriales y pruebas de consistencia interna de las categorías, se llegó a las siguientes conclusiones:

ⁱⁱⁱ El factor Conciencia es uno de los "Cinco Grandes" factores de la personalidad (Big Five: BF). Modelo que, hoy por hoy, se entiende como modelo dominante de la estructura de la personalidad. Los otros cuatro factores son: Extroversión, Estabilidad Emocional, Agradabilidad y Apertura a la Experiencia.

Categorías de CCP	N	α
Robos y conductas relacionadas	10	.77
Destrucción de propiedades de la organización	4	.66
Mal uso de la información (falsificar información, mentir, no guardar secretos)	5	.71
Mal uso del tiempo y de los recursos	13	.90
Comportamiento Temerario (no seguir directivas de seguridad)	4	.71
Falta de Atención	5	.77
Baja calidad del trabajo	3	.86
Uso y abuso de alcohol	3	.59
Uso y abuso de drogas	4	.71
Abusos verbales (palabrotas, insultos)	8	.82
Abusos físicos o sexuales	7	.82

Gruys, M. L. y Sackett, P.R. (2003)

Las 10 CCP consideradas como más susceptibles de incurrir son las siguientes:

Conductas Contra-Productivas	Media (1-7)	Desviación
Uso de correo electrónico para fines personales	4.66	2.17
Hacer fotocopias de material propio en el trabajo	4.28	1.95
Usar Internet para fines no relacionados con el trabajo	3.47	1.99
Hacer negocios personales en tiempo de trabajo	3.41	1.80
No leer/conocer las directivas de seguridad	3.28	1.66
Tiempo excesivo y no aprobado para comidas o café	3.12	1.98
Discutir con un compañero de trabajo	2.89	1.41
Perder el tiempo durante el trabajo	2.89	1.49
Hacer llamadas personales a larga distancia	2.86	1.85
Fingir una enfermedad para no ir a trabajar	2.70	1.80

Gruys, M. L. y Sackett, P.R. (2003)

El estudio, además, concluyó lo siguiente:

- La mayoría de las CCP son comunes a la mayoría de las organizaciones.
- Cuando una persona se ve envuelta en un CCP es probable que se vea envuelta en otros.
- Existe soporte empírico para el uso de las 11 categorías de CCP.
- A más edad y experiencia menor probabilidad de verse envuelto en CCP.
- Existe evidencia de que hay correlación entre las 11 categorías, lo que indica que existe una estructura jerárquica de éstas, con un único factor en la cúspide como sustento de las CCP.

4. EL ANALISIS DEL TRABAJO COMO FUENTE DE CONTENIDOS PARA LA EVALUACION DEL DESEMPEÑO

De todo lo escrito hasta aquí puede deducirse que una ED no está completa si no se realiza midiendo los tres tipos de desempeño, y que de ellos, los referidos a tarea deben ser concretos del puesto de trabajo objeto de evaluación. En realidad esto no es así, los contenidos y la exhaustividad de los mismos depende de los objetivos de la evaluación. Así, si lo que se pretende es validar predictores selectivos o formativos, no es necesario realizar evaluaciones de desempeño sobre dimensiones exhaustivas o específicas de cada puesto. Valdría con una evaluación global y sobre algunas de las dimensiones generales citadas, siempre que éstas fueran fiables y discriminaran lo suficiente entre buenos y malos. A esto han contribuido las conclusiones que en cuanto a los constructos

de desempeño han permitido los estudios meta-analíticos y la generalización de la validez (Guion, 1997). Ahora bien, si lo que se pretende es conocer el dominio de la ejecución; conocer cuál y qué es el éxito en un puesto de trabajo concreto, sí es necesario un análisis más exhaustivo que permita determinar criterios más detallados (no necesariamente más específicos del puesto o situación). Lo mismo ocurre si lo que se pretende es identificar necesidades formativas mediante la ED, diseñar formación, analizar la polivalencia entre puestos, o identificar las destrezas críticas entre puestos para determinar peajes formativos de carrera, etc. A modo de resumen y más con intención de reflexión que de afirmación, se adjunta la siguiente tabla:

Acción	ED Exhaustivo	ED General
Dominio del Puesto	x	
Selección (Diseño)		x
Selección (Validación)		x
Formación (Diseño)	x	
Formación (Validación)		x
Carrera (Diseño)	x	
Carrera (Validación)		x
Promoción		x
Remoción	x	
Salario (subir)		x
Salario (bajar)	x	
Feed-Back: Motivación		x
Feed-Back: Mejora	x	

En la mayoría de las organizaciones, las prácticas referidas a RRHH no se hacen de forma aislada, las necesidades que éstas satisfacen están siempre presentes. Como se ha dicho más arriba, incluso en las AAPP se hace evaluación de desempeño aunque esta es de un modo no sistemático, formal, profesional o evidente. Conocer un puesto de forma exhaustiva es necesario para la mayoría de las prácticas de RRHH que exijan diseño o justificación. Por ello el análisis de trabajo es algo inevitable si se quiere cumplir con las

propuestas de prácticas en RRHH que propone el Informe de la Comisión para la elaboración del Estatuto Básico del Empleado Público.

La Dirección de Función Pública del Gobierno Vasco ha realizado y finalizado un análisis de todos los puestos (AP) de funcionarios de su Administración General. La metodología empleada parte de una descripción general del puesto y de un inventario de tareas que cada trabajador confecciona mediante un

cuestionario al que se accede, previa identificación y verificación, a través de su Intranet. Cada tarea de dicho inventario es valorada por el/la propio/a trabajador/a según cuatro criterios de relevancia o criticidad (frecuencia, duración, consecuencias de ejecución inadecuada y dificultad de aprendizaje). Como resultado de dichas valoraciones se obtienen las tareas críticas o más representativas del puesto, que posteriormente se someten a un Análisis de Tarea (AT), que consiste en una entrevista de un/a analista con el experto en la misma, en la que se identifican: el diagrama de flujo (los pasos de la ejecución), las destrezas (qué es necesario saber hacer para realizarla bien), y los estándares de ejecución (qué significa hacerla bien). Éste último apartado es el que aporta los contenidos para el conocimiento del dominio de la ejecución, el que identifica qué significa hacer bien o mal el trabajo y por tanto, los contenidos relevantes para una posible evaluación del desempeño.

Cuando en el AT se identifican los estándares de ejecución, se hace con referencia a tres aspectos de cada tarea crítica: la ley, las condiciones y los indicadores.

- *La ley.* Identificar la ley de una tarea consiste plasmar por escrito, el valor añadido de la misma; lo que satisface cuando se hace bien y qué deja de conseguirse si se hace mal. En definitiva es un criterio de bondad, mediante el que se puede juzgar una ejecución como buena, regular o mala. Las leyes suelen ser jerárquicas, es decir, que hay algunas cuya trasgresión, por sí sola, significa que la tarea esta mal hecha, y otras cuya trasgresión, no significa que esté mal, pero tampoco bien (regular).
- *Las condiciones:* se entiende por condición de ejecución, los requisitos

que debe cumplir un servicio si el fruto de la tarea es tal, o una ejemplificación si el fruto de la tarea es un producto (el mejor ejemplo de informe). Mediante las condiciones se identifican aspectos esenciales de la bondad, la forma como la ley se encarna en el producto o cumple en el servicio.

- *Los indicadores:* los indicadores identifican los efectos o señales tangibles y cuantificables en las que se manifiesta la ley. Normalmente suelen ser aspectos cuantificables como errores, cantidad crítica a producir, quejas, número de veces que algo debe repetirse. La principal característica del indicador es que no está bajo el dominio directo del trabajador/a sino que es consecuencia de sus acciones, efectos de su proceder.

Ejemplo: Departamento de Sanidad. Dirección de Salud Pública (Antón, 2005)

Tarea 176. Realizar inspecciones sanitarias en establecimientos alimentarios

Ley:

- *La correcta ejecución de la tarea debe reflejar con objetividad y precisión la situación sanitaria de los establecimientos inspeccionados conforme a normas.*

Condiciones:

- *Cumplir el plazo de 15 días.*
- *La **Hoja de datos administrativos del establecimiento** debe incluir las posibles variaciones en esos datos desde la última inspección, incluida la actividad que desarrolla.*
- *El **protocolo** ha debido ser cumplimentado de forma completa y*

- siguiendo las instrucciones técnicas para hacerlo.
- El **formulario de clasificación** del establecimientos según el riesgo debe estar completo y según las instrucciones técnicas (criterios...) para hacerlo.
 - El **acta** debe ser:
 - Según el modelo oficial.
 - Por duplicado: una copia para el establecimiento y la otra para la administración.
 - Con la siguiente estructura:
 - Encabezado: lugar y fecha, inspector/a, nombre del establecimiento y la actividad, dirección, teléfono, representante del establecimiento.
 - Cuerpo del acta.
 - Debe ser objetiva describiendo lo que ha constatado en la inspección.
 - Debe describir una por una todas las deficiencias (mayores, serias o críticas) reflejadas en el protocolo, excepto De acuerdo con la experta el enunciado de la tarea se ha modificado ligeramente: Realizar inspecciones sanitarias en establecimientos alimentarios por quien ejecuta la inspección.
 - Debe ser leída a la persona del establecimiento, quien deberá firmarla y quedarse con una copia. Si no quiere firmarla, puede hacerlo otra persona (otro inspector/a, testigo...).
 - El o la representante del establecimiento debe poder añadir observaciones.
 - Los **registros en la aplicación** deben estar completos y siguiendo todos los pasos.
 - La **carta** de requerimiento debe:
 - En caso de deficiencias serias o mayores que lo requieran debe transmitir literalmente lo que pone en el acta (nº de acta, fecha de la inspección, hechos...), indicar el plazo el plazo que se da al establecimiento para corregir las deficiencias y estar firmada por el Director Territorial.
 - En caso de deficiencias leves debe señalarlas (no aparecen en acta) y solicitar que se corrijan antes de la siguiente inspección.
 - Deben enviarse con acuse de recibo.
- Indicadores:**
- <2 días de incumplimiento del plazo propuesto por el coordinador para la realización de la inspección.
 - <1% de alegaciones de establecimientos que prosperan.
 - 98% de inspecciones que tienen el VºBº del personal técnico de Bilbao y/o de Lakua.
 - 10-15% de inspecciones con deficiencias leves (p.e. fallos con la herramienta informática o al rellenar el protocolo porque los criterios están poco claros) detectadas en las auditorías externas realizadas por la Subdirección de otra provincia.
 - 0% de inspecciones con deficiencias graves (p.e. autorizar un establecimiento con deficiencias serias o críticas) detectadas en las auditorías externas realizadas por la Subdirección de otra provincia.
- Conocer las tareas más importantes de un puesto a este nivel permite elaborar

instrumentos de ED que pongan contenidos a las dimensiones identificadas para el desempeño de tarea, o identifiquen otras de manera que un/a trabajador/a pueda ser evaluado/a respecto de contenidos relevantes

(ya se ha comentado que el desempeño contextual y las conductas contra-productivas no necesitan este análisis previo porque los estudios factoriales ya las han identificado).

5. METODOLOGIAS DE EVALUACION DEL DESEMPEÑO. CRITERIOS PARA SU ELECCION

Criterios de elección

Existen referencias exhaustivas y buenas para poder conocer las metodologías existentes para realizar una ED. En este sentido y para nuestro país, la referencia del libro de Santiago Quijano (1992) es imprescindible. Para esta ponencia se considera más apropiado exponer aquellos aspectos que deben tenerse en cuenta a la hora de escoger el método. También, aquellos resultados de investigación referidos a la naturaleza del acto de evaluar, a sus limitaciones inherentes y a su fiabilidad y validez.

Los métodos usados para medir la ejecución individual pueden ser clasificados en dos grandes grupos:

- Registros Organizacionales (datos objetivos referidos a número de unidades producidas, número de ausencias al trabajo, horas perdidas, etc.).
- Evaluaciones Subjetivas (cuando una persona evalúa a otra) que a su vez pueden dividirse en:
 - Referidas a criterios preestablecidos: evaluación ("ratings"). Las más famosas son las EVAS que se forman por anclajes conductuales como ejemplificaciones de las conductas que se valoran (Quijano, 1992; Smith y Kendall, 1963).
 - No referidos a una norma concreta, pero con alguna forma de clasificación u orden ("ranking").

Donde lo que se hace es comparar al evaluado con otros (Comparación de Pares, Jerarquización, Distribución Forzada, etc). (Quijano, 1992).

Conceptualmente debe distinguirse entre métodos de medida (el cómo se mide; los de arriba) y tipos de criterio (su naturaleza). Estos últimos pueden clasificarse en:

- Criterio Inmediato: Es la medida de una ejecución de alguien en un momento concreto.
- Criterio Intermedio: Es la medida de la ejecución durante un periodo de tiempo.
- Criterio Último: Es un resumen del valor global o total que tiene un individuo para la organización.

También existe otra forma de dividirlos: ejecución típica o ejecución máxima. Esta última entendida como la ejecución máxima posible sin limitaciones y con máximos niveles de motivación. Tanto para ésta como para las tipologías de criterio inmediato, intermedio y último, se pueden aplicar los métodos Objetivos (registros) y los Subjetivos (juicios de personas).

A la hora de escoger el método deben tomarse tres grandes decisiones dependiendo de los objetivos de la evaluación:

- Dimensionalidad: Medir de forma global o de forma micro.
- Normativas vs. Estándares: Medir comparativamente respecto de una norma o hacerlo mediante una medida absoluta.
- Ejecución Típica vs. Ejecución máxima. En general aquellas que permitan a los sujetos diferenciarse y generar variabilidad.

Limitaciones de las ED

Como toda medida, la ED tiene sus limitaciones y errores más frecuentes. Los conceptos a los que se deben las limitaciones de la ED son los siguientes:

- Contaminación: cuando la puntuación obtenida no depende de la ejecución del evaluado (intereses espurios implicados en la evaluación).
- Deficiencia: cuando lo que se mide no es relevante.
- Sesgo (Bias): cuando se producen errores debidos a la intervención humana (Halo, Benevolencia, Tendencia Central, etc).

Muchas de las soluciones que se han propuesto para solucionar estos errores han sido: la elección forzada como las Mixed Standard Scales (Blanz y Ghiselli, 1972), los "rankings", las EVAS (Quijano, 1992; Smith y Kendall, 1963) con sus anclajes conductuales para valorar la conducta. Seguidamente se exponen de forma concisa los errores más frecuentes en la ED (López Basterra, 2005) aquellos que niegan una de sus condiciones de éxito: la variabilidad de las medidas.

- **Halo:** Ocurre cuando se generaliza a la valoración de todos los aspectos del desempeño de una persona, alguna característica que no tiene que ver con la

ejecución de sus tareas. Por ejemplo: valorar como buen trabajador a alguien por el sólo hecho de ser simpático o agradable. En este caso el error consiste en no diferenciar entre buen trabajador (eficaz, profesional, etc) y ser simpático, y entender que ambas son aspectos distintos^{iv}.

- **Tendencia Central:** En realidad aquí lo que opera es el miedo, la tendencia a no involucrarse en la evaluación por temor a las explicaciones o justificaciones que haya que hacer posteriormente: "si no pongo mal a nadie, no tendré que dar explicaciones". Este error es típico cuando el evaluador interpreta la evaluación desde un punto de vista excesivamente centrado en las repercusiones de la evaluación y se desmarca como elemento importante para la mejora de esta herramienta. Suele responder a la pregunta ¿Y yo qué gano con esto?

- **Benevolencia:** Es una versión del error anterior pero en positivo, más que en neutral. Aquí lo que se piensa es que se evitan líos si todos son buenos. Lo que subyace es la creencia de que nadie preguntará si todos son bien evaluados.

Toda evaluación de desempeño implica variabilidad simplemente por que la gente se diferencia en cómo hace las cosas. Algunos son buenos en algunas tareas y peores en otras. Caer en los tres errores arriba citados implica negar esta evidencia. Incluso aunque la gente sea buena, existe un gradiente entre ser bueno y muy bueno.

Fiabilidad y Validez de la ED.

Desde principios de los años noventa, la aplicación de las técnicas meta-analíticas a la Psicología del Trabajo y de las Organizaciones (Hunter y Schmidt, 1990b), ha concluido una serie de avances en el conocimiento que permiten no empezar de cero a la hora de acometer una ED. Esto es de especial importancia caso de usarse esta técnica para las AAPP ya que justificaría una serie de intervenciones metodológicas y sus

resultados. Estas son las principales conclusiones de este tipo de investigaciones.

La fiabilidad es uno de los principales handicaps a la hora de implantar una ED en las AAPP. La fiabilidad habla de la consistencia, estabilidad, equivalencia y grado de acuerdo entre medidas y evaluadores. Si ya es un cambio cultural hacer una ED en las AAPP, más lo es hacerlo dos veces, preguntar

a varios, preguntar lo mismo en distintos periodos de tiempo o a compañeros de trabajo (pares). Por suerte los datos meta-analíticos para superiores y pares son los siguientes^v (Viswesvaran, 2001):

^{iv} Para ver posibles soluciones y controles formales (estadísticos) del error de Halo, ver los trabajos de Hoyt, W. T. (2000) y de Hoyt, W.T. & Kerns, M.D. (1999)

^v La fiabilidad se expresa en términos de correlación. Sus valores van de 0 a 1. La validez de -1 a +1

Por dimensiones:

Dimensión	Inter-jueces	α^{vi}
Productividad	.57	.82
Liderazgo	.53	.77
Esfuerzo	.55	.79
Competencia Interpersonal	.47	.77
Conocimiento	.53	

Por tipo de evaluador:

Fiabilidad	Superiores	Pares
α	.86	.85
Inter-jueces	.52 ^{vii}	.42

^{vi} Cuando se cita el símbolo α , se indica el coeficiente Alpha de Cronbach.

^{vii} Estos datos son trascendentes si se quieren realizar estudios de validez en los que se corrijan los coeficientes de validez observados. Así, en un estudio realizado en la Administración General Vasca respecto de la Entrevista Conductual Estructurada, se ha usado el coeficiente de fiabilidad de .52 para corregir la validez observada y, junto con la corrección de la restricción del rango (no se puede evaluar el desempeño de los que no han sido seleccionados), obtener la validez operativa. Para más detalles, ver López Basterra, J. Y Gorriti, M (2005).

O sea, que los superiores evalúan de forma más consistente que los pares y sobre todo respecto del grado de acuerdo entre diferentes evaluadores del mismo rango (fiabilidad inter-jueces). En general y respecto de la fiabilidad se puede concluir:

- Por defecto, el coeficiente que mide el tamaño del efecto de la fiabilidad inter-jueces, es decir, el grado de acuerdo entre evaluadores es: .52
- Hay más acuerdo entre evaluadores del mismo o distinto rango cuando lo que se mide es una dimensión global del desempeño, que una dimensión concreta (.62 entre superiores y pares).

- Las evaluaciones más fiables son las de superiores sobre dimensiones de tarea.

Hablar de validez en las medidas de desempeño puede parecer un círculo vicioso porque se pretende validar aquello con lo que se valida (la variable dependiente), sin embargo sí es posible y pertinente validar los contenidos de la ED (el criterio) como constructo. En realidad las propias taxonomías expuestas en el punto 3 de este documento y sus metodologías para probarse, son estudios de validez de constructo en sí mismos. La definición de qué, cuáles y sus relaciones nomológicas, son formas de probar qué debe medirse. Los contenidos referidos a los distintos ámbitos de desempeño (tarea,

contextual y contra-productividad) son resultados de estudios de validez de constructo utilizando metodologías factoriales. Otras metodologías utilizadas son la validez convergente y la matriz multi-rago, multi-método (Campbell y Fiske, 1959). Como la validez de constructo no se puede medir directamente lo que se hace es ver con qué correlaciona, o si varios evaluadores coinciden en cuanto a lo que están midiendo. Luego se les pone etiquetas.

Respecto a los métodos que usaban la validez convergente, la teoría dice que cuando la validez convergente es alta es porque hay un constructo subyacente que está siendo medido. Y la pregunta que surge aquí es la siguiente: de todo lo que se mide cuando se mide ¿cuánto es lo que se refiere al desempeño y cuánto lo que contamina el propio evaluador al medir y cuánto lo que es debido al inevitable azar. Schmidt, Viswesvaran y Ones (2000) en respuesta a una polémica que afirmaba que el tamaño del efecto de la idiosincrasia del evaluador era .54 concluyeron que este dato confundía fiabilidad con validez e hicieron un meta-análisis que obtuvo las siguientes conclusiones:

- Tamaño del efecto debido al desempeño por sí mismo: .52 (lo importante para los fines originales)
- Tamaño del efecto debido a los errores del evaluador, a la interacción entre el evaluado y evaluador: .29 (mejorable por formación, EVAS, control de errores, etc.)
- Tamaño del efecto debido al error aleatorio, al azar: .19 (no controlable)

Todos estos datos nos permiten abordar con garantías de relevancia, fiabilidad y discriminación una posible ED en las Administraciones Públicas. Esta ponencia ha demostrado la necesidad, no sólo técnica sino pronto legal, de acometer esta práctica de RRHH en ellas, también ha identificado qué medir y ha dado los criterios para escoger cómo hacerlo. Por último estos datos nos permiten rechazar uno de los componentes del miedo a implantar estas técnicas y afirmar, con evidencias, que sí existen criterios, contenidos y técnicas para hacer buenas ED en las AAPP.

REFERENCIAS

- Antón, P. (2005): *Informe de Análisis de Puestos*. Vitoria: Dirección de Función Pública. Gobierno Vasco.
- Bernardin, H.J. & Beatty, R. (1984). *Performance Appraisal: Assessing human behavior at work*. Boston: Kent-PWS.
- Blanz, F. & Ghiselli, E.E. (1972). The mixed standard scale: A new rating system. *Personnel Psychology*, 25, 185-199
- Borman, W.C., & Motowidlo, S.J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt, & W.C. Borman (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco, CA: Jossey-Bass.
- Borman, W.C., Penner, L.A., Allen, T.D. & Motowidlo, S.J. (2001). Personality Predictors of Citizenship Performance. *International Journal of Selection and Assessment*, Vol. 9, Numbers 1/2. (pp. 52-69)
- Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by means of the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105
- Campbell, J.P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. Dunnette & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, 2nd ed., pp. 687-731). Palo Alto, CA: Consulting Psychologist Press.
- Cascio, W.F. (1991). *Applied psychology in personnel management* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Comisión para el Estudio y Preparación del Estatuto Básico del Empleado Público (Miguel Sánchez Morón, Presidente) (2005). *Informe*. Madrid. Ministerio de Administración Públicas. Estatuto Básico del Empleado Público.
- Gruys, M.L. & Sackett, P.R. (2003). Investigating the Dimensionality of Counterproductive Work Behavior. *International Journal of Selection and Assessment*. Vol. 11, number 1.
- Guion, R. (1997). Criterion Measurement and the Criterion Dilemma. En N. Anderson y P. Herriot (Eds.) *International Handbook of Selection and Assessment*. Cap. 13 (pp.267-286). Londres: John Wiley & Sons.
- Hoyt, W.T. & Kerns, M.D. (1999). Magnitude and moderators of bias in observer ratings: A meta analysis. *Psychological Methods*, 4, 403-424.
- Hoyt, W.T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods*, 5, 64-86.
- Hunter, J.E. & Schmidt, F.L. (1990). *Methods of meta-analysis: Correcting for error and bias in research findings*. Newbury Park, CA: Sage.
- López Basterra, J y Gorriti, M. (2005). *Validación de la entrevista conductual estructurada (ECE) en la Administración General del País Vasco*. Vitoria: Dirección de Función Pública, Gobierno Vasco.
- López Basterra, J. (2005). *Errores comunes en la Evaluación de Desempeño*. Vitoria: Dirección de Función Pública. Gobierno Vasco.
- Miller, R.L., Griffin, M.A., & Hart, P.M. (1999) Personality and organizational health: The role of conscientiousness. *Work and Stress*, 13, 7-19
- Murphy, K. R. (1989) Dimensions of job performance. In R. Dillon & J. Pelligrino (Eds.), *Testing: Applied and theoretical perspectives* (pp. 218-247). New York: Praeger.

- Quijano, S. (1992). *Sistemas efectivos de evaluación y rendimiento: Resultados y desempeños*. Barcelona: PPU.
- Schmidt, F.L., Viswesvaran, C., & Ones, D.S. (2000). Reliability is not validity and validity is not reliability. *Personnel Psychology*, 53, 901-912.
- Smith, P.C. & Kendall, L.M. (1963). Retranslation of expectation. *Journal of Applied Psychology*, 47, 149-155.
- Society for Industrial and Organizational Psychology, Inc. (1987). *Principles for the validation and use of personnel selection procedures*. (3rd.ed.) College Park, MD: Author.
- Society for Industrial and Organizational Psychology, Inc. (2004). *Principles for the validation and use of personnel selection procedures*. (Four Edition). www.siop.org
- Toops, H.A. (1944). The criterion. *Educational and Psychological Measurement*, 4, 271-297
- Viswesvaran, C. (2001) Assessment of individual Job Performance: A review of the Past Century and a Look Ahead. In Neil Anderson, Deniz S. Ones, Handan Kepir Sinangil & Chockalingam Viswesvaran, Editors. *Handbook of Industrial and Organizational Psychology. Vol.1: Personnel Psychology. Chapter 6*, pp: 110-126. SAGE
- Viswesvaran, C., Ones, D.S., Schmidt, F.L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 81, 557-574
- Wherry, R.J. (1957). The past and future of criterion evaluation. *Personnel Psychology*, 10, 1-5