



## Artículo

---

### La Evaluación del Desempeño en las Administraciones Públicas Españolas

---

### Performance Evaluations in the Spanish Public Administration

MIKEL GORRITI BONTIGUI\*

---

#### RESUMEN

Este trabajo demuestra la necesidad, no sólo técnica sino ya legal (Ley 7/2007, de 12 de abril), de acometer la evaluación de desempeño como práctica común de las actividades de Recursos Humanos en las Administraciones Públicas españolas. Identifica las ventajas de no hacer evaluación de desempeño de forma sistemática ya que, su premisa es que alguna forma de evaluación de desempeño se realiza siempre en las AAPP. También justifica y describe las tres categorías de este tipo de evaluaciones: Evaluación de Tarea, Desempeño Contextual y Contra-productividad, y aporta los criterios para escoger cómo hacerlo, solventar los errores clásicos de estas medidas, tanto desde el punto de vista de la literatura como desde un proceso más exhaustivo de análisis de trabajo. Por último aporta los resultados más recientes de la investigación meta-analítica que permiten afirmar, con evidencias, que existen criterios, contenidos y técnicas para hacer buenas Evaluaciones de Desempeño en las Administraciones Públicas.

#### ABSTRACT

This paper shows the need, not only technical but also legal (Act 7/2007), of carrying out performance evaluations as a normal human resources practice in the Spanish Public Administration. The article identifies the advantages of not doing this kind of evaluations systematically in the Spanish public administration, as some kind of performance evaluation is somewhat made. Three categories in the performance domain are described –task

---

\* Responsable de Recursos Humanos. Dirección de Función Pública. Departamento de Hacienda y Administración Pública del Gobierno Vasco.

performance, contextual performance and counter-productive behaviors. Some criteria are also offered with respect to how to proceed with performance evaluation and how to overcome typical measurement errors –both from literature and from the task analysis process. Finally, the paper shows recent meta-analysis outcomes about reliability and construct validity of performance evaluations –an evidence for the existence of criteria, contents and techniques to guarantee good performance evaluations in the Spanish public administrations.

### **PALABRAS CLAVE**

Evaluación de Desempeño, Administraciones Públicas, Desempeño Contextual, Contraproductividad, Meta-análisis, Fiabilidad, Validez, Análisis de Trabajo.

### **KEY WORDS**

Performance Evaluation, Public Administration, Contextual Performance, Counter-productive Behaviors, Meta-analysis, Reliability, Construct Validity, Task Analysis.

## INTRODUCCIÓN

El presente trabajo pretende facilitar a los técnicos y gestores de los Recursos Humanos (RRHH) de las administraciones públicas españolas, conceptos, criterios y métodos para la más que probable obligatoriedad de acometer Evaluaciones del Desempeño (ED) de los empleados públicos en las administraciones en las que vaya a ser vinculante el futuro Estatuto Básico del Empleado Público (EBEP), recientemente aprobado (Ley 7/2007, de 12 de abril).

Se estudia en primer lugar el informe de expertos que ha sido la base técnica sobre la que se ha elaborado el Estatuto. Él es el referente más cercano para explicar la necesidad de cómo diseñar la ED, en él se han basado las exigencias del Estatuto. Seguidamente se extiende el posible uso de la ED, más allá de una frecuente concepción fiscalizadora de esta práctica, y se reconoce la importancia capital de la ED para otros usos típicos de las actividades de RRHH. También se explica qué es conceptualmente, cuáles son las dimensiones y la naturaleza de sus contenidos, su imprescindible origen en el análisis del trabajo, sus limitaciones y, por último, sus propiedades psicométricas.

Con todo ello, se pretende dar confianza técnica para abordar esta inevitable rea-

lidad cuya trascendencia, probablemente, sólo ha sido atisbada. La ED exige un difícil cambio cultural en las administraciones públicas españolas que no aborda este trabajo. Este cambio y la posibilidad técnica que sí se aborda, son un binomio imprescindible para hacer de la ED una práctica sistemática, rigurosa, provechosa y justa.

## INFORME DE LA COMISIÓN PARA EL ESTUDIO Y PREPARACIÓN DEL ESTATUTO BÁSICO DEL EMPLEADO PÚBLICO

Con fecha 16 de septiembre de 2004 se constituyó la Comisión para el Estudio y Preparación del Estatuto Básico del Empleado Público<sup>1</sup>, 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 Recursos Humanos en las Administraciones Públicas (AAPP) a las que afectará la entrada en vigor de dicha Ley.

Por lo que respecta a la ED, muchas son las 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<sup>2</sup>. Este apartado pretende repasar de forma

<sup>1</sup> BOE, número 228, de 21 de Septiembre de 2004, Orden APU/3018/2004. Página 31644.

<sup>2</sup> El EBEP en su Artículo 20 referido exclusivamente a la Evaluación del Desempeño dice lo siguiente:

1. Las Administraciones Públicas establecerán sistemas que permitan la evaluación del desempeño de sus empleados. La evaluación del desempeño es el procedimiento mediante el cual se mide y valora la conducta profesional y el rendimiento o el logro de resultados.

2. Los sistemas de evaluación del desempeño se adecuarán, en todo caso, a criterios de transparencia, objetividad, imparcialidad y no discriminación y se aplicarán sin menoscabo de los derechos de los empleados públicos.

3. Las Administraciones Públicas determinarán los efectos de la evaluación en la carrera profesional horizontal, la formación, la provisión de puestos de trabajo y en la percepción de las retribuciones complementarias previstas en el artículo 24 del presente Estatuto.

4. La continuidad en un puesto de trabajo obtenido por concurso quedará vinculada a la evaluación de acuerdo con los sistemas de evaluación que cada Administración Pública determine, dándose audiencia al interesado, y por la correspondiente resolución motivada.

5. La aplicación de la carrera profesional horizontal, de las retribuciones complementarias derivadas del apartado c) del artículo 24 del presente Estatuto y el cese del puesto de trabajo obtenido por el procedimiento de concurso requerirán la aprobación previa, en cada caso, de sistemas objetivos que permitan evaluar el desempeño de acuerdo con lo establecido en los apartados 1 y 2 de este artículo.

sucinta estas referencias y analizar la trascendencia que desde un punto de vista práctico pueden tener para nosotros como profesionales de la Psicología del Trabajo y de las Organizaciones (PTO) y en general para todos los profesionales susceptibles de tener la responsabilidad de diseñarla.

- 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 tareas; 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 desmotivador y más formal que eficaz, ligar la carrera a la antigüedad. Mejorar en el puesto de trabajo sin moverse de él, 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 EBEP 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 novedoso 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).

### **POSIBLES USOS DE LA EVALUACIÓN DEL DESEMPEÑO**

Es posible que la primera percepción de la ED sea negativa en el sentido de que se utilice como un instrumento de fiscalización de la actividad del/la empleado/a público. Sin embargo esta visión es limitada y culturalmente condicionada. 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 (Cleveland, Murphy y Williams 1989):

- Evaluación entre personas: salarios, promociones.
- Evaluación de personal: identificar potencialidades y flaquezas (formación, motivación).
- Mantenimiento del sistema: Investigación (conocimiento del criterio), planificación de recursos humanos (RRHH), organización.
- Documental: para justificar decisiones e intervenciones ante recursos o cualquier otro proceso administrativo o legal implicado.

Hasta lo que el autor sabe, muy pocas son las AAPP que han realizado algún tipo de ED y menos las que han publicado algo al respecto independiente de las referidas a la valoración puestos o los costes de personal. ¿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 cri-

terios para salarios, para promociones, para diseñar la formación, etc., el problema es que es necesario hacerla de forma sistemática y con un método contrastado. Esta es una de las premisas de este artículo: 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, y 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* (Jenkins, 1946; Ghiselli, 1956; Guion, 1961; Dunnette, 1963; Wallace, 1965; Morales, 1988; Binning y Barrret, 1989; Austin y Villanova, 1992; Guion, 1997; SIOP, 1987 y 2004; Viswesvaran, 2001; Viswesvaran y Ones, 2005), como una variable lo suficientemente importante como para ser predicha, mejorada o corregida. Una variable que necesita ser explicada y definida. 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 este conocimiento, ¿qué justificación existe para inferir perfiles

selectivos o contenidos formativos, para definir salarios, para diseñar la carrera, etc., si no se tiene claro qué significa hacer bien un trabajo? ¿Cómo se puede saber qué exigir o medir en un proceso selectivo, si no se sabe qué significa hacerlo bien? ¿Cómo se pueden proponer mejoras en la formación o reconducir actuaciones de los trabajadores/as, si no se sabe 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 nuevos o prorrogados, exime de la necesidad de demostrar la eficacia de la Administración y por ende, de diferenciar entre quien hace bien o mal su trabajo. En realidad todos sabemos quién trabaja bien y quién lo hace peor. 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. Desde el conocimiento no sistemático del desempeño, se toman decisiones sobre las personas que no se quieren hacer evidentes porque evita actuaciones desagradables:

- Evita tener que ser asertivos cuando hay que pasar el mal trago de decir a alguien que hace mal su trabajo y que algo le va a pasar por ello.
- Evita tener que establecer compromisos de cumplimiento que estructuren el tiempo.
- Evita tener que tomar decisiones retributivas que paguen más a los que mejor hacen y romper el “café para todos”.

- 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.
- 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.
- Permite raseros generales a la hora de la movilidad entre puestos, lo cual tiene más posibilidades de éxito social pero un coste de eficacia.
- 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.
- Evita cuestionar el status quo, al no tener un referente de cuáles son las características de producto o servicio que exigen los ciudadanos.
- Etc.

Cualquier Administración que no mide su desempeño, dificulta el natural cambio gradual y está abocada a cambios bruscos, cualitativos y traumáticos producidos por cambios políticos, tecnológicos, etc. 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 circunscribir cualquier realidad política surgida de un cambio electoral, a su auténtica naturaleza

de estrategia política, orientación del servicio público, etc.

### **LA EVALUACIÓN DEL DESEMPEÑO DESDE EL PUNTO DE VISTA CONCEPTUAL. LOS PLANTEAMIENTOS TAXONÓMICOS DE LOS TRES GRANDES ENFOQUES**

Pero, ¿qué es una evaluación del desempeño? Existen distintas concepciones de lo que es la ED: eficacia y valor de una conducta del trabajo y sus efectos (SIOP, 1987). Un comportamiento organizacional evaluable y que está bajo el control del propio trabajador (Viswesvaran 2001). Medida de una ejecución o conducta laboral relevante (SIOP, 2004). El procedimiento mediante el cual se mide y valora la conducta profesional y el rendimiento o el logro de resultados (EBEP, 2007).

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, técnicamente, Evaluación del Desempeño (ED) es el proceso mediante el que se mide y valora un comportamiento organizacional relevante, bajo el dominio del propio ejecutor.

En cuanto a los criterios para saber si una ED es válida, hay acuerdo general (Viswesvaran, 2001; SIOP, 2004) 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 trascendentes 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.

Aunque existen otras taxonomías de las dimensiones que constituyen el desempeño, las más general y aceptada es la de Viswesvaran (Viswesvaran, Ones y Schmidt, 1996; Viswesvaran, 2001) 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. Viswesvaran et al. (1996) concluyeron que el desempeño es una estructura multidimensional formada por un factor general de desempeño en el ápice y por otros subfactores menores, jerárquicamente relacionados.

En la última década y fruto de distintas investigaciones que tienen que ver con la relevancia de la ED (Borman y Motowidlo, 1993; Conway, 1999; Sackett y DeVore, 2001), 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 con-



ductas contraproduktivas. 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 contraproduktivas 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.).

### 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 (Schmidt y Hunter, 1998; Ones, Viswesvaran y Dilchert, 2005) formada por los conocimientos del puesto, las destrezas entendidas como formas estandarizadas de resolver problemas (Campbell y Kuncel, 2001) y los hábitos entendidos como respuestas características (Borman, Penner, Allen, Motowidlo, 2001).

En la tabla 1 se exponen resumidas las dimensiones que fruto de estudios factoriales y otras metodologías formales, se han concluido relevantes para evaluar el desempeño de tarea<sup>3</sup>.

A diferencia de las otras tipologías de

desempeño, la manifestación conductual de las referidas a tareas varían con el puesto, sin embargo las referidas al desempeño contextual y las conductas Contraproduktivas son generales (Borman et al., 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, d) porque aunque deseables, son más implícitas que explícitas.

La principal taxonomía de desempeño contextual es de Borman et al. (2001) y consta de tres dimensiones con sus subdimensiones (tabla 2).

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 (Borman, et al., 2001; Penney y Borman, 2005). Así, un estudio de Miller, Griffin y Hart

<sup>3</sup> 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.

**Tabla 1.**

<b>Autor/es</b>	<b>Naturaleza de la medida</b>	<b>Dimensiones</b>
Toops (1944)	Distinguiendo entre adecuación: calidad y ausencia de errores, y cantidad: volumen de producto	<ul style="list-style-type: none"> <li>• <b>Unidades de producción</b></li> <li>• <b>Calidad del trabajo.</b></li> <li>• Tiempo de permanencia</li> <li>• Destrezas de supervisión y liderazgo.</li> </ul>
Wherry (1957)		<ul style="list-style-type: none"> <li>• <b>Resultados</b></li> <li>• <b>Calidad</b></li> <li>• Tiempo perdido en la ejecución</li> <li>• Cambio de trabajos</li> <li>• Tiempo de formación o promoción</li> <li>• Satisfacción</li> </ul>
Bernardin y 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"> <li>• <b>Calidad</b></li> <li>• <b>Cantidad</b></li> <li>• <b>Tiempo</b></li> <li>• <b>Costo-eficacia</b></li> <li>• Necesidad de supervisión</li> <li>• Impacto interpersonal</li> </ul>
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"> <li>• <b>Eficacia en tareas específicas.</b></li> <li>• <b>Eficacia en tareas no específicas (para todos)</b></li> <li>• Comunicación oral y escrita.</li> <li>• Esfuerzo demostrado (en acabar una tarea)</li> <li>• Disciplina personal sostenida</li> <li>• Facilitación de ejecución pares y equipos</li> <li>• Supervisión</li> <li>• <b>Gestión o administración.</b></li> </ul>
Murphy (1989)	Entiende la ejecución como el compendio de cuatro dimensiones	<ul style="list-style-type: none"> <li>• Conductas que hacen perder el tiempo</li> <li>• Ejecución de las tareas o roles definidos para el puesto.</li> <li>• Interpersonales: de ayuda y de equipo.</li> <li>• Conductas destructivas: violencia, robos, etc.</li> </ul>

(1999) realizado con empleados públicos encontró una correlación de .42 (N=104) entre el factor Conciencia<sup>4</sup> de personalidad (disciplina, sostenimiento de esfuerzo, responsabilidad, etc.) y el desempeño contextual.

### **Conductas Contraproductivas**

Una conducta contraproductiva (CCP) es todo comportamiento intencional de un miembro de la organización, contrario a los legítimos intereses de ésta (Sackett y

<sup>4</sup> 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 (Hough y Ones, 2001).

Tabla 2.

<i>Dimensión</i>	<i>Definición</i>	<i>Sub-dimensiones</i>
<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"> <li>• Cooperación</li> <li>• Cortesía</li> <li>• Motivación: (aplaudir los éxitos ajenos, bromear en tiempos de adversidad, mostrar confianza en sus capacidades para triunfar, ayudarles para superar las dificultades).</li> </ul>
<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"> <li>• Representación</li> <li>• Lealtad</li> <li>• Compromiso</li> </ul>
<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"> <li>• Persistencia.</li> <li>• Iniciativa.</li> <li>• Auto-desarrollo.</li> </ul>

DeVore, 2001). 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 (Grays 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, que aparecen en la tabla 3.

Las 10 CCP consideradas como más susceptibles de incurrir son las que se muestran en la tabla 4.

El estudio, además, concluyó que la mayoría de las CCP son comunes a la mayoría de las organizaciones. Cuando una persona se ve envuelta en un CCP es

**Tabla 3.**

<i>Categorías de CCP</i>	<i>N</i>	<i>a</i>
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

N= número de conductas contraproductivas incluidas en la categoría. a=Alpha de Cronbach

**Tabla 4.**

<i>Conductas Contraproductivas</i>	<i>Media (1-7)</i>	<i>S<sub>x</sub></i>
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

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.

#### **EL ANÁLISIS DEL TRABAJO COMO FUENTE DE CONTENIDOS PARA LA EVALUACIÓN 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 estas fueran fiables y discriminaran lo suficiente entre quienes desempeñan bien o mal el traba-

jo. 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 (Viswesvaran, 2001; 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 distancias formativas como condición de la carrera, etc.

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. 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 y lo aprobado en el EBEP.

El 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 a través de su Intranet, previa identificación y verificación. 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 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/la experto/a 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 hacerlo 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 en 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.

- *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.

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 contraproductivas no necesitan este análisis previo porque los estudios factoriales ya las han identificado).

## **METODOLOGÍAS DE EVALUACIÓN DEL DESEMPEÑO. CRITERIOS PARA SU ELECCIÓN**

### **Criterios de elección**

Existen referencias exhaustivas para

poder conocer las metodologías existentes para realizar una ED. En este sentido, para nuestro idioma, la referencia del libro de Santiago de Quijano (1992 y 1997) es imprescindible. Para este artículo 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 BARS (Behaviorally Anchored Rating Scales) que se forman por anclajes conductuales como ejemplificaciones de las conductas que se valoran (Smith y Kendall, 1963).
  - No referidas 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 y 1997).

Conceptualmente debe distinguirse

entre métodos de medida (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 (Sackett, Zedeck y Fogli, 1988; Klehe y Anderson, 2005). 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 exhaustiva.
- **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 (SIOP, 2004):

- **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:** 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 BARS (Smith y Kendall, 1963; Quijano, 1992) 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 enten-

der que ambas son aspectos distintos<sup>5</sup>.

- **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 porque 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, 1990), 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 en 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 ( $\alpha$ ), estabilidad, equivalencia y grado de acuerdo entre medidas y evaluadores (fiabilidad interjueces). 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<sup>6</sup> (Viswesvaran, Ones y Schmidt, 1996; Viswesvaran, 2001) (Tablas 5 y 6).

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 interjueces). 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.

<sup>5</sup> 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. y Kerns, M.D. (1999).

<sup>6</sup> La fiabilidad se expresa en términos de correlación. Sus valores van de 0 a 1. La validez de  $-1$  a  $+1$ .



**Tabla 5.**  
Por dimensiones

<i>Dimensión</i>	<i>Interjueces</i>	<i>α</i>
Productividad	.57	.82
Liderazgo	.53	.77
Esfuerzo	.55	.79
Competencia Interpersonal	.47	.77
Conocimiento	.53	

**Tabla 6.**  
Por tipo de evaluador

<i>Fiabilidad</i>	<i>Superiores</i>	<i>Pares</i>
<i>α</i>	.86	.85
Interjueces	.52 <sup>7</sup>	.42

- 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 este artículo 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 desem-

peño (tarea, contextual y contraproductividad) son resultados de estudios de validez de constructo utilizando metodologías factoriales. Otras metodologías utilizadas son la validez convergente y la matriz multi-rasgo-multimétodo (Campbell y Fiske, 1959). Como la validez de constructo no se puede medir directamente (Binning y Barrett, 1989), 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?

<sup>7</sup> 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).

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, BARS, control de errores, etc.).
- Tamaño del efecto debido al error aleatorio, al azar: .19 (no controlable)<sup>8</sup>.

Todos estos datos nos permiten abordar con garantías de relevancia, fiabilidad y discriminación una posible ED en las Administraciones Públicas a las que es de aplicación el recientemente aprobado Estatuto Básico del Empleado Público. Este trabajo 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 que sí existen criterios, contenidos y técnicas para hacer buenos diseños técnicos de Evaluación de Desempeño en las Administraciones Públicas.

## REFERENCIAS BIBLIOGRÁFICAS

Austin, J.T. y Villanova, P. (1992). The criterion problem: 1917-1992. *Journal of Applied Psychology*, 77, 836-874.

Bernardin, H.J. y Beatty, R. (1984). *Performance Appraisal: Assessing human behavior at work*. Boston: Kent-PWS.

Binning, J.F. y Barrett, G.V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, 74, 478-494.

Blanz, F. y Ghiselli, E.E. (1972). The mixed standard scale: A new rating system. *Personnel Psychology*, 25, 185-199.

Borman, W.C., y Motowidlo, S.J. (1993). Expanding the criterion domain to include elements of contextual performance. En N. Schmitt, y W.C. Borman (Ed.), *Personnel selection in organizations* (71-98). San Francisco, CA: Jossey-Bass.

Borman, W.C., Penner, L.A., Allen, T.D. y Motowidlo, S.J. (2001). Personality predictors of citizenship performance. *International Journal of Selection and Assessment*, 9, 52-69.

Campbell, D.T., y 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. En M. Dunnette y L.M. Hough (Ed.), *Handbook of industrial and organizational psychology 1*, 2<sup>nd</sup> ed. (687-731). Palo Alto, CA: Consulting Psychologist Press.

<sup>8</sup> Nótese que la suma de los tres efectos da 1.

- Campbell, J.P. y Kuncel, N.R. (2001). Individual and team training. En N. Anderson, D. S. Ones, H. K. Sinangil y C. Viswesvaran, (Editores). *Handbook of Industrial, Work and Organizational Psychology. 1*, 278-312. Londres: SAGE.
- Cascio, W.F. (1991). *Applied psychology in personnel management* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Cleveland, J.N., Murphy, K.R., y Williams, R.E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. *Journal of Applied Psychology*, 74, 130-135.
- 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.
- Conway, J.M. (1999). Distinguishing contextual performance from task performance for managerial jobs. *Journal of Applied Psychology*, 84, 3-13.
- Dunnette, M.D. (1963). A note on the criterion. *Journal of Applied Psychology*, 47, 251-254.
- Ghiselli, E.E. (1956). Dimensional problems of criteria. *Journal of Applied Psychology*, 40, 1-4.
- Gruys, M.L. y Sackett, P.R. (2003). Investigating the dimensionality of counterproductive work behavior. *International Journal of Selection and Assessment*, 11, (1) 30-42.
- Guion, R.M. (1961). Criterion measurement and personnel judgments. *Personnel Psychology*, 14, 141-149.
- Guion, R. M. (1997). Criterion measurement and the criterion dilemma. En N. Anderson y P. Herriot (Ed.) *International Handbook of Selection and Assessment* (267-286). Londres: John Wiley y Sons.
- Hough, L.M. y Ones, D.S. (2001). The structure, measurement, validity and use of personality variables in industrial, work and organizational psychology. En N. Anderson, D. S. Ones, H. K. Sinangil y C. Viswesvaran, (Edits.). *Handbook of Industrial, Work and Organizational Psychology. 1*, 233-277. Londres: SAGE.
- Hoyt, W.T. y 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. y Schmidt, F.L. (1990). *Methods of meta-analysis: Correcting for error and bias in research findings*. Newbury Park, CA: Sage.
- Jenkins, J.G. (1946). Validity for what? *Journal of Consulting Psychology*, 10, 93-98
- Klehe, U.C., y Anderson, N. (2005). The prediction of typical and maximum performance in employee selection. En A. Evers, N. Anderson y O. Voskuilj (Editores). *The Blackwell Handbook of Personnel Selection* (331-353). Londres: Blackwell.

- Ley 7/2007, de 12 de abril, del *Estatuto Básico del Empleado Público*. Boletín Oficial de Estado. Núm. 89 (16270-16299).
- López Basterra, J y Gorriti, M. (2005). *Validación de la entrevista conductual estructurada (ECE) en la Administración General del País Vasco*. Vitoria (España): Dirección de Función Pública, Gobierno Vasco.
- López Basterra, J. (2005). *Errores comunes en la Evaluación de Desempeño*. Vitoria (España): Dirección de Función Pública. Gobierno Vasco.
- Miller, R.L., Griffin, M.A., y Hart, P.M. (1999) Personality and organizational health: The role of conscientiousness. *Work & Stress, 13*, 7-19.
- Morales, P. (1988). *Medición de actitudes en psicología y educación. Construcción de escalas y problemas metodológicos*. San Sebastián. Tártalo.
- Murphy, K. R. (1989) Dimensions of job performance. En R. Dillon y J. Pelligrino (Edits.), *Testing: Applied and theoretical perspectives* (218-247). Nueva York: Praeger.
- Ones, D.S., Viswesvaran, C. y Dilchert, S. (2005). Cognitive ability in personnel selection decisions. En A. Evers, N. Anderson y O. Voskuijl (Ed.). *The Blackwell Handbook of Personnel Selection*. 143-173. London: Blackwell.
- Penney, L.M. y Borman, W.C. (2005). The prediction of contextual performance. En A. Evers, N. Anderson y O. Voskuijl (Ed.). *The Blackwell Handbook of Personnel Selection* (376-396). Londres: Blackwell.
- Quijano, S. (1992 y 1997). *Sistemas efectivos de evaluación del rendimiento: Resultados y desempeños*. Barcelona: EUB.
- Sackett, P.R., Zedeck, S., y Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology, 73*, 482-486.
- Sackett, P.R. y DeVore, C. J. (2001). Counterproductive behaviors at work. En N. Anderson, D. S. Ones, H. K. Sinangil y C. Viswesvaran (Ed.). *Handbook of Industrial, Work and Organizational Psychology, 1*, 145-164. Londres: SAGE.
- Schmidt, F.L. y Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124* (2), 262-274.
- Schmidt, F.L., Viswesvaran, C., y Ones, D.S. (2000). Reliability is not validity and validity is not reliability. *Personnel Psychology, 53*, 901-912.
- Smith, P.C. y 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* (3<sup>rd</sup>.ed.) College Park, MD: Author.
- Society for Industrial and Organizational Psychology, Inc. (2004). *Principles for the validation and use of personnel selection procedures* (4<sup>a</sup> Edición). [www.siop.org](http://www.siop.org)
- Toops, H.A. (1944). The criterion. *Edu-*

*cational and Psychological Measurement*, 4, 271-297.

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.

Viswesvaran, C. (2001). Assessment of individual Job Performance: A review of the Past Century and a Look Ahead. En N. Anderson, D. S. Ones, H. K. Sinangil y C. Viswesvaran, (Ed.). *Handbook of Industrial, Work and Organizational Psychology*. 1, (110-126). Londres: SAGE.

Viswesvaran, C. y Ones, D.S. (2005). Job performance: Assessment issues in personnel selection. En A. Evers, N. Anderson y O. Voskuijl (Eds.). *The Blackwell Handbook of Personnel Selection* (354-375). Londres: Blackwell.

Wallace, S.R. (1965). Criteria for what?. *American Psychologist*, 20, 411-417.

Wherry, R.J. (1957). The past and future of criterion evaluation. *Personnel Psychology*, 10, 1-5.