



# Journal of Work and Organizational Psychology

<http://journals.copmadrid.org/jwop>



## Predictores Selectivos en el Sector Público: Validez Predictiva y Reacciones de los Candidatos

Alexander Rodríguez y Joseba López-Basterra

*Instituto Vasco de Administración Pública, España*

### INFORMACIÓN DEL ARTÍCULO

#### Historia del artículo:

Recibido el 14 de marzo de 2017  
Aceptado el 24 de noviembre de 2017  
Online el 8 de febrero de 2018

#### Palabras clave:

Selección de personal  
Procedimientos selectivos  
Validez predictiva  
Desempeño laboral  
Justicia procedimental

#### Keywords:

Personnel selection  
Selection procedures  
Predictive validity  
Job performance  
Procedural justice

### R E S U M E N

Esta investigación ha tenido como objetivo profundizar en el conocimiento sobre el funcionamiento de los procedimientos selectivos en el sector público. Para ello, a partir de los datos del concurso-oposición de 2010 para la selección de auxiliares administrativos en el Gobierno Vasco se estudiaron dos cuestiones: a) la validez predictiva de los predictores empleados y b) la justicia percibida por parte de los auxiliares con respecto al proceso selectivo y a las pruebas empleadas y su relación con el resultado en las mismas y el desempeño laboral. Los resultados muestran que la prueba de conocimientos es el mejor predictor del desempeño y que las reacciones de los candidatos presentan una correlación positiva con el desempeño laboral y con el resultado en el test de juicio situacional. Se comentan las limitaciones de la investigación así como las implicaciones prácticas de los resultados.

### Selection predictors in the public sector: Predictive validity and candidate reactions

### A B S T R A C T

The purpose of this research has been to deepen the knowledge about how selection procedures work in the public sector. To this aim two aspects were considered: a) the predictive validity of the predictors employed and b) the perceived justice of the hired workers regarding the selection process and the tests used and their relation with test performance and job performance. The results show that the job knowledge test is the best predictor of performance and that the reactions of the candidates present a positive correlation with job performance and with the situational judgment test results. The limitations of the research as well as the practical implications of the results are discussed.

La selección de personal en el sector público está sometida a factores legales, políticos y sociales que la diferencian del sector privado, dado el elevado interés que suscita en la opinión pública (Austin, Klimoski y Hunt, 1996; Patterson, Lievens, Kerrin, Zibarras y Carette, 2012; Wiesen, Abrams y McAttee, 1994). Esta relevancia social hace que la legislación y la jurisprudencia insten a garantizar que se seleccione a los mejores candidatos desde el punto de vista de predicción del desempeño futuro. Todo ello de una manera justa, objetiva, meritocrática y neutral, desde la igualdad exigida a procesos de estas características (Chaves, 2009; Gorriti y López Basterra, 2014; Ramió, 2011; Van-Biesen, 2006).

La trascendencia de estas cuestiones hizo que en el marco del Proyecto Compartir, puesto en marcha por el Instituto Nacional de Administración Pública (INAP), el grupo de trabajo en selección inte-

grado por representantes de escuelas e institutos de la Administración Pública del Estado se centrara en su estudio empírico. Dentro de ese grupo, el Instituto Vasco de Administración Pública (IVAP) se marcó dos objetivos: estudiar la validez predictiva de las pruebas selectivas y cómo las perciben los candidatos en el ámbito público empleando los datos del último proceso selectivo para la contratación de auxiliares administrativos en el Gobierno Vasco, celebrado en 2010.

### Validez Predictiva de las Pruebas Selectivas en el Sector Público

La función pública se articula sobre la base del sistema de mérito, cuyo objetivo último consiste en la selección de los aspirantes más

Para citar este artículo: Rodríguez, A. y López-Basterra, J. (2018). Predictores selectivos en el sector público: validez predictiva y reacciones de los candidatos. *Journal of Work and Organizational Psychology*, 34, 16-28. <https://doi.org/10.5093/jwop2018a3>

Correspondencia: [alexander-rodriguez@ivap.eus](mailto:alexander-rodriguez@ivap.eus) (A. Rodríguez).

ISSN: 1576-5962/© 2018 Colegio Oficial de Psicólogos de Madrid. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

cualificados (Mauri, 2009). Por ello, el RDL 5/2015 insta a que los procedimientos selectivos cuiden “especialmente la conexión entre el tipo de pruebas a superar y la adecuación al desempeño de las tareas de los puestos de trabajo convocados”. Esta vinculación entre pruebas y desempeño indica que la selección debe hacerse para predecir la futura realización del trabajo y, en consecuencia, que los procedimientos deberán ser aceptables en términos de sus propiedades psicométricas de validez predictiva (Salgado, 2010).

Sin embargo, en el sector público español se han realizado estudios de validez predictiva en contadas ocasiones y se han centrado fundamentalmente en la entrevista conductual estructurada (Berges, 2011; Rodríguez, 2016; Sáez, 2007; Salgado, Gorriti y Moscoso, 2007). Más infrecuente aun ha sido la evaluación de una de las pruebas más utilizadas en el sector público, la prueba de conocimientos, que solo ha sido abordada por el Ayuntamiento de Zaragoza (Berges, 2011) y cuya validez media corregida, según la investigación meta-analítica de Dye, Reck y McDaniel (1993), es de .45 en la predicción del desempeño laboral.

Un tipo de pruebas que pueden medir conocimientos del puesto son los tests de juicio situacional (TJS), que según Lievens (2007) son “instrumentos que plantean a los candidatos situaciones laborales hipotéticas y muestran las posibles soluciones a éstas. Los candidatos tienen que decidir cuál de las respuestas alternativas elegirían”. Estas pruebas, equiparables a las pruebas de carácter práctico requeridas por el RDL 5/2015, se emplean con mucha frecuencia en todo el mundo (Krantowitz, 2014; Salgado, Viswesvaran y Ones, 2001) y se utilizan habitualmente en la selección del sector público europeo (INAP, 2012). En su metaanálisis, McDaniel y Nguyen (2001) hallaron una validez de criterio corregida de .34.

Aunque los TJS se emplean fundamentalmente para medir liderazgo, destrezas interpersonales, personalidad o habilidad cognitiva (Christian, Edwards y Bailey, 2010; McDaniel y Nguyen, 2001) también pueden medir conocimientos del puesto, algo que coincide con la concepción de Schmidt (1994) o Schmidt y Hunter (1993) sobre estas pruebas. Christian et al. (2010), en su metaanálisis basado en los constructos medidos por los TJS, encontraron que los que medían conocimientos del puesto, tanto de tipo declarativo (nombres, términos, fechas) como procedimental (acciones, destrezas y operaciones), presentaron una validez corregida de .19.

Además de estas pruebas, en la selección de empleados públicos se emplea el “curso de méritos”, en el que se valoran aspectos como la experiencia laboral, las titulaciones académicas o el conocimiento de idiomas (Alonso, Táuriz y Choragwicka, 2009; Basoredo y López Basterra, 2013), siendo este sistema el más empleado en el sector público de Estados Unidos (Cook, 1980). En nuestro país se emplea el método de puntos, que consiste en que los candidatos reciben una puntuación en base a criterios y baremos preestablecidos por las bases de convocatoria, que definen la puntuación máxima de cada aspecto a valorar, debiendo los candidatos acreditar el número de meses o años trabajados o las titulaciones que poseen (Alonso et al., 2009).

Según el metaanálisis de Schmidt, Oh y Shaffer (2016) la experiencia, medida a través de los años de permanencia en el puesto, tiene una correlación con el desempeño de .16, mientras que Quiñones, Ford y Teachout (1995) la sitúan en .27.

Con respecto a las titulaciones, el metaanálisis de Dye y Reck (1988) encontró una validez corregida de .13 para el nivel de educación, constructo que puede ser medido mediante años de educación formal o mediante la acreditación de titulaciones, que es la manera usual de puntuar este aspecto en el sector público.

En el caso de la valoración de idiomas, la literatura es escasa y está centrada en el desempeño de trabajadores desplazados al extranjero o empleados en multinacionales (Basoredo y López Basterra, 2013; Kortmann, 2016). Así, cabe mencionar el metaanálisis de Mol, Born, Willemssen y Van der Mollen (2005), que encontraron una correlación corregida de .19 entre conocimiento de la lengua local y el desempe-

ño de trabajadores expatriados, mientras que el estudio primario de Kortmann (2016) investigó la relación entre el desempeño laboral y variables como número de idiomas hablados y dominio de lenguas extranjeras, encontrando correlaciones de .23 y .57 respectivamente con una muestra de trabajadores de multinacionales.

Por todo ello, en base a los estudios previos establecemos las siguientes hipótesis:

H1. Los resultados en la prueba de conocimientos presentarán una correlación positiva con el desempeño del trabajo del personal auxiliar administrativo contratado.

H2. Los resultados en el test de juicio situacional presentarán una correlación positiva con el desempeño del trabajo del personal auxiliar administrativo contratado.

H3. La valoración de la experiencia presentará una correlación positiva con el desempeño del trabajo del personal auxiliar administrativo contratado.

H4. La valoración de las titulaciones académicas presentará una correlación positiva con el desempeño del trabajo del personal auxiliar administrativo contratado.

H5. La valoración del conocimiento de idiomas presentará una correlación positiva con el desempeño del trabajo del personal auxiliar administrativo contratado.

### Reacciones de los Candidatos: Justicia Percibida

Una segunda cuestión muy relevante es la “justicia percibida” por los candidatos a empleos acerca de las pruebas utilizadas. En los últimos años se ha investigado mucho al respecto dada la importancia que tiene para el diseño de procedimientos selectivos por parte de las organizaciones y a una demanda social creciente de justicia, imparcialidad y trato justo en la selección de personal (García-Izquierdo y García-Izquierdo, 2007). Como señalan Zibarras y Patterson (2015), esto es especialmente importante en el ámbito público, ya que las oposiciones son un evento que atrae la atención de la ciudadanía e implican un elevado nivel de escrutinio público y rendición de cuentas (Harris, 2000; Ryan, Greguras y Ployhart, 1996). Consecuentemente, los procedimientos selectivos deben ser justos y defendibles legalmente, por lo que es vital estudiar esta cuestión (Carr y Patterson, 2009; Patterson, Zibarras, Carr, Irish y Gregory, 2011).

Además, el estudio de las reacciones de los candidatos es crucial porque estas pueden influir en la imagen pública de la organización convocante, así como en el resultado en las pruebas selectivas e incluso en el resultado de la evaluación del desempeño (Anderson, Salgado y Hülsheger, 2010; Ryan y Ployhart, 2000; Salgado et al., 2007; McCarthy et al., 2013).

En general, los resultados de diferentes estudios sugieren que existe una relación positiva entre la percepción de justicia procedimental y el resultado en los exámenes, es decir que las personas que tienen reacciones positivas hacia los procedimientos selectivos obtienen puntuaciones superiores en dichos procedimientos y viceversa (Ryan, Sacco, McFarland y Kriska, 2000; Smither, Reilly, Millsap, Pearlman y Stoffey, 1993; Truxillo, Steiner, y Gilliland, 2004), siendo las magnitudes de las relaciones similares a las encontradas por el metaanálisis de Hausknecht, Day y Thomas (2004), que encontró correlaciones corregidas entre .11 y .22 entre las puntuaciones en varios instrumentos selectivos (entrevistas, tests de inteligencia, tests de personalidad, biodata, etc.) y diferentes percepciones de los candidatos (justicia procedimental, relación del instrumento selectivo con el contenido del puesto, motivación hacia los tests y actitudes hacia los tests).

Por otro lado, aunque algunos estudios han señalado la importancia de analizar la relación entre las percepciones de los candidatos y el desempeño laboral (Chan y Schmitt, 2004; Ryan y Ployhart, 2000) esta cuestión se ha investigado en escasas ocasiones. Gilliland

(1994) encontró que los empleados que percibieron que habían sido seleccionados mediante procedimientos injustos tuvieron un desempeño peor que los que tuvieron una percepción positiva. Colquitt, Conlon, Wesson, Porter y Ng (2001), por su parte, hallaron que las percepciones de justicia estaban relacionadas con el desempeño y McCarthy et al. (2013) encontraron evidencias de que las reacciones, cuando son disposicionales, afectan directamente al desempeño laboral y que cuando las reacciones son situacionales afectan al desempeño indirectamente a través de su influencia en las calificaciones obtenidas en pruebas de conocimientos y TJS.

En el sector público se ha investigado la relación entre la justicia percibida y el desempeño en los tests (McCarthy, Hrabliuiky y Jeller, 2009; Smither et al., 1993), aunque no se ha estudiado la relación entre las reacciones de los candidatos y el desempeño laboral, lo que hace que sea necesario investigar sobre esta cuestión.

Dadas las características de la selección en el sector público y la naturaleza multidimensional del constructo *justicia procedimental* (Colquitt et al., 2001; Gilliland y Steiner, 2014) se antoja necesario medir tres dimensiones de la justicia percibida por los participantes en los procesos selectivos y su relación con los resultados en las pruebas y el desempeño laboral: 1) la adecuación de la dificultad de las pruebas (Arvey, Strickland, Drauden y Martin, 1990; Hong, 1999; Puertas y Embid, 2011; Rodríguez Garraza, 2008), 2) su grado de adecuación al temario de la convocatoria, para garantía de las expectativas razonables de los candidatos y ofrecer seguridad jurídica (Chaves, 2009), lo que está vinculado con el principio de transparencia exigido por la legislación y es próximo conceptualmente a la dimensión *justicia informativa* del modelo de justicia procedimental de Gilliland (1993), y 3) su relación con el contenido y las tareas del puesto a cubrir, la dimensión más relevante del modelo de Gilliland (Osca y Salmones, 2004; Ployhart y Ryan, 1998).

Estos tres aspectos reflejan el nivel de satisfacción global con la justicia procedimental del proceso selectivo, es decir, si las pruebas se perciben como justas y adecuadas al proceso y si la toma de decisiones, las reglas y el procedimiento empleados para la decisión selectiva se dan de forma equitativa y justa (Bauer et al., 2001; Hausknecht et al., 2004; Osca y Salmones, 2004).

En base a la investigación previa, establecemos las siguientes hipótesis:

H6. La percepción de los candidatos presentará una correlación positiva con las calificaciones en las pruebas selectivas (prueba de conocimientos y TJS) para el puesto de auxiliar administrativo/a.

H7. La percepción de los candidatos presentará una correlación positiva con el desempeño del trabajo del personal auxiliar administrativo contratado.

## Estudio 1

### Método

**Participantes.** La muestra estuvo compuesta por 101 personas (23 varones y 78 mujeres), con una edad media de 39.5 años ( $SD = 6.03$ ) que superaron el proceso selectivo para la selección de auxiliares administrativos en 2010 y estaban trabajando en el Gobierno Vasco durante el año 2016 y cuyos superiores aceptaron evaluar su desempeño.

**Medidas.** En este estudio se han empleado las siguientes medidas:

**Prueba de conocimientos.** Se empleó un cuestionario de 100 preguntas con cuatro alternativas de respuesta, en las que se penalizaron las respuestas erróneas. Por cada respuesta errónea se descontó la tercera parte del valor de una respuesta correcta. La puntuación máxima alcanzable por los candidatos según las bases de la convocatoria era de 27 puntos. Esta prueba era obligatoria y eliminatória.

Se trató de una prueba de conocimientos relacionados con los contenidos exigidos por el puesto determinados previamente mediante la herramienta de análisis de puestos del Gobierno Vasco (Gorriti y López Basterra, 2010). En su diseño se cuidó especialmente que todos los apartados del temario estuvieran representados en las preguntas. Se ofrecen ejemplos de preguntas de esta prueba en el Apéndice. La fiabilidad de esta prueba (alfa de Cronbach) fue de .81.

**Test de juicio situacional (TJS).** Se empleó un cuestionario de 47 preguntas con cuatro alternativas de respuesta (de las cuales solo una era correcta), en las que se penalizaron las respuestas erróneas del mismo modo que en la prueba de conocimientos. De las 47 preguntas, 36 hacían referencia a 8 casos prácticos y las 11 preguntas restantes contaban con enunciados que planteaban situaciones prácticas de diversa índole. La puntuación máxima alcanzable por los candidatos según las bases de la convocatoria fue de 42 puntos. Esta prueba era obligatoria y eliminatória.

En el diseño de esta prueba destinada a medir los conocimientos de los aspirantes se trató de abarcar el mayor número posible de tareas evaluables mediante una prueba tipo test en función del análisis del puesto, ya que las bases de la convocatoria establecían que debía basarse en el temario y en las tareas del puesto convocado. Se planteaban situaciones a las que se podían enfrentar los auxiliares administrativos en su día a día a las que debían responder aplicando al caso concreto sus conocimientos. Se ofrecen ejemplos de preguntas de esta prueba en el Apéndice. Su fiabilidad (alfa de Cronbach) fue de .51.

**Experiencia.** Se consideró experiencia a los servicios prestados en la Administración Pública en un puesto de contenido igual o similar al de las funciones del puesto convocado. El cómputo de la experiencia se efectuó realizando la suma de los periodos trabajados, valorándose a razón de .12 por mes trabajado hasta un máximo de 22 puntos, lo que equivalía a un máximo de 15 años de experiencia según las bases de la convocatoria. Dada la naturaleza de esta medida, no se pudo computar su fiabilidad.

**Titulaciones académicas.** Se valoraron hasta un máximo de 6 puntos las titulaciones académicas oficiales diferentes a la aportada para acceder al proceso selectivo que se consideraron útiles para desempeñar el puesto de Auxiliar Administrativo: titulaciones de FP II en Administración y Finanzas, Secretariado, Gestión Administrativa, etc. La fiabilidad de esta medida (alfa de Cronbach) fue de .75.

**Conocimientos de informática.** Se valoraron hasta un máximo de 2 puntos los conocimientos en Excel, Access, Internet y PowerPoint acreditados mediante el sistema de certificación IT-Txartela. La fiabilidad de esta medida (alfa de Cronbach) fue de .78.

**Conocimientos de idiomas.** Se valoraron los conocimientos de inglés y francés de nivel igual o superior al B1 del Marco Común Europeo de Referencia hasta un máximo de 1 punto en inglés y francés. Dada la naturaleza de esta medida, no se pudo computar su fiabilidad.

**Medidas del criterio: evaluación del desempeño.** Para la evaluación del desempeño laboral se diseñó una herramienta compuesta por 26 escalas de valoración con anclajes conductuales que medían dos dimensiones: desempeño de tarea y desempeño contextual. Para elaborar los anclajes conductuales se recogieron incidentes críticos proporcionados por tres expertos en el puesto de auxiliar administrativo/a con más de 5 años de experiencia.

Dado que la manifestación conductual del desempeño de tarea varía con el puesto (Borman, Penner, Allen y Motowidlo, 2001; Gorriti y López Basterra, 2009) y que es conveniente que las dimensiones de tarea se identifiquen a partir del análisis de puestos (Bernardin y Beatty, 1984; Díaz-Cabrera et al., 2014; Murphy y Cleveland, 1995; Peiró, Ramos y Cortés, 2012) se optó por elaborar una herramienta específica del puesto objeto de evaluación. Así, la dimensión desempeño de tarea se compuso de 5 subdimensiones: atender consultas y reclamaciones (medida mediante 4 ítems),

comprobar y cotejar datos, registros y documentación (3 ítems), alimentar y actualizar bases de datos y aplicaciones informáticas (3 ítems), elaborar certificaciones, notificaciones, compulsas, etc. (2 ítems) y tramitar expedientes (3 ítems). Las 5 subdimensiones fueron elegidas de entre las tareas más representativas del puesto en base a la información aportada por la herramienta de análisis de puestos del Gobierno Vasco (Gorriti y López Basterra, 2010) por ser aquellas para las que los expertos en el puesto elaboraron más incidentes críticos.

La dimensión desempeño contextual se diseñó en base al modelo de Borman et al. (2001) y estaba compuesta por 4 subdimensiones: ayuda y cooperación (medida mediante 5 ítems), iniciativa (2 ítems), persistencia/tenacidad (2 ítems) y autodesarrollo (2 ítems).

La escala de evaluación de cada ítem contó con cinco niveles (de 1, el nivel de desempeño más negativo o inadecuado, a 5, el nivel de desempeño más positivo o adecuado). La media de las puntuaciones en desempeño de tarea (DT) y desempeño contextual (DC) permitió obtener un valor de desempeño general (DG). El coeficiente alfa de Cronbach de DG, DT y DC fue, respectivamente, de .92, .89 y .86. Los coeficientes alfa de las subdimensiones oscilaron entre .63 y .91.

Por otro lado, además de este cuestionario de evaluación de desempeño convencional, dada la tendencia de los evaluadores a asignar puntuaciones elevadas, lo que se conoce como sesgo de benevolencia (Aguinis, 2013; Díaz Cabrera et al., 2014), o a igualar las puntuaciones de los evaluados (Peiró et al., 2012), es decir, sesgo de tendencia central, se utilizaron otros dos cuestionarios basados en la comparación entre empleados para que los superiores con más de un evaluado comparasen el desempeño de sus subordinados en 8 subdimensiones. Las 4 primeras medían DT y fueron identificadas en base a los criterios de ejecución obtenidos para las tareas del puesto de auxiliar a partir de la herramienta de análisis de puestos del Gobierno Vasco (cantidad de trabajo, satisfacción de demanda de información, garantizar la eficacia e igualdad de actuaciones de la Administración, asegurar la relevancia y corrección de datos/información) y las 4 últimas se correspondían con las dimensiones que medían DC en el cuestionario de desempeño para todos los auxiliares (ayuda y cooperación, iniciativa, persistencia/tenacidad, autodesarrollo). El primer cuestionario se empleó con aquellos superiores con un número de subordinados entre 2 y 4, debiendo ordenarlos de mejor a peor desempeño en cada dimensión. Las puntuaciones en cada dimensión para cada evaluado podían ser, en función del orden, 0 ó 1 punto (con dos evaluados), 0, 1 ó 2 puntos (con tres evaluados) o 0, 1, 2 ó 3 (con 4 evaluados). Este primer cuestionario se empleó para evaluar el desempeño de 20 auxiliares. El segundo cuestionario se empleó con los que tenían 5 o más evaluados, debiendo realizar una distribución forzosa en tres niveles (escala de evaluación de 0 a 2 puntos), pudiendo situar en los niveles superior (2 puntos) e inferior (0 puntos) únicamente a una persona. Este segundo cuestionario se empleó para evaluar a 7 auxiliares.

En estos dos cuestionarios de comparaciones forzosas, la puntuación de cada auxiliar en DG, DT y DC se obtuvo a partir de la proporción entre los puntos acumulados en cada dimensión y el máximo de puntos que podían recibir en cada dimensión (8 puntos en DT y DC en el caso de tres evaluados, 4 puntos en DT y DC en el caso de 2 evaluados). La fiabilidad (entendida como alfa de Cronbach) para las medidas de DG, DT y DC obtenidas mediante comparaciones forzadas con el primer cuestionario ( $n = 20$ ) fue de .96, .94 y .93 respectivamente.

Se facilitó a los evaluadores un manual con información precisa acerca los criterios de valoración del desempeño y el procedimiento para puntuarlo según el tipo de cuestionario, los sesgos a considerar, las definiciones conceptuales de cada dimensión y los anclajes conductuales a emplear. Además, se ofreció un teléfono de atención de consultas para que los evaluadores pudieran solucionar sus dudas con respecto a la evaluación.

Los superiores directos, jefes de servicio que tienen la oportunidad de observar continuamente el desempeño de sus subordinados, evaluaron el desempeño de los sujetos que conformaron la muestra durante el último año.

**Procedimiento.** En el procedimiento selectivo convocado por el Gobierno Vasco para seleccionar 155 auxiliares administrativos se inscribieron 13,700 personas. De estas, 2,141 aprobaron los ejercicios eliminatorios (conocimientos y TJS) y por tanto podían ser contratados, en función de la nota final obtenida (la suma de las puntuaciones en todos los instrumentos selectivos mencionados en el apartado Medidas), para desempeñar bien un puesto fijo como funcionarios de carrera (lo que lograron los 155 con mejores notas finales) o bien un puesto temporal como funcionarios interinos.

Los resultados del procedimiento selectivo ya obraban en manos del IVAP (registros informatizados de las 2,141 personas que superaron el proceso selectivo de la oferta pública de empleo al cuerpo de auxiliares administrativos/as del Gobierno Vasco celebradas en 2010). De esas 2,141 personas, 339 estaban trabajando como auxiliares administrativos en el Gobierno Vasco en 2016 y para obtener los datos de desempeño era necesario contar con la colaboración de sus superiores directos. Para ello se solicitó a los distintos departamentos del Gobierno su colaboración. Finalmente, se consiguió la colaboración de 77 superiores directos (43 mujeres y 34 varones que ocupaban jefaturas de servicio) que evaluaron el desempeño de 101 de las 339 personas contratadas tras haber superado la oposición; 10 de estos superiores evaluaron también el desempeño de 27 auxiliares mediante cuestionarios basados en la comparación entre empleados.

Habida cuenta de que contar con apenas 27 cuestionarios de comparación forzosa implicaba un problema de potencia estadística, es decir, que aumente la probabilidad de que un test estadístico lleve al rechazo de la hipótesis nula, afirmar que no hay relación entre las variables cuando en realidad la hay (Cohen, 1988), se procedió de la siguiente manera: para poder agrupar en una escala única las puntuaciones obtenidas mediante comparaciones y distribuciones forzadas con las puntuaciones obtenidas mediante el cuestionario convencional (el cuestionario basado en anclajes conductuales) se transformaron en puntuaciones  $z$  los distintos tipos de puntuaciones para obtener una puntuación única en desempeño para el total de 101 auxiliares. Además, para estudiar con precisión la relación entre predictores y criterios, habiendo empleado instrumentos distintos en una misma muestra de sujetos, se empleó la correlación parcial para eliminar el efecto del tipo de instrumento. Los valores obtenidos se corrigieron para controlar los artefactos estadísticos de la misma manera que al emplear las medidas de criterio no basadas en comparaciones y poder calcular los valores de validez operativa.

Se evaluó el desempeño a lo largo del año 2016 evaluando únicamente a aquellas personas con al menos un año de experiencia en el puesto. Los cuestionarios fueron cumplimentados en el lugar de trabajo mediante una aplicación que garantizaba el anonimato de las respuestas.

**Corrección de artefactos estadísticos.** Para contrarrestar los errores debidos a artefactos estadísticos (errores no debidos a la varianza sistemática de los candidatos sino a errores de los propios instrumentos de medida), los coeficientes de validez se corrigieron primero por falta de fiabilidad del criterio –empleando el valor de .52 estimado como fiabilidad meta-analítica para pruebas de desempeño realizadas por superiores (Sackett, 2014; Salgado, 2015; Salgado et al., 2003; Salgado, Moscoso y Anderson, 2016; Salgado y Táuriz, 2014; Schmidt, Le e Ilies, 2003; Shen, Cucina, Walmsley y Seltzer, 2014; Viswesvaran, Ones y Schmidt, 1996; Viswesvaran, Ones, Schmidt, Le y Oh, 2014)– y después por restricción del rango directa –la restricción que se produce en la muestra como efecto de la propia selección, ya que sólo es posible evaluar el criterio en las personas contratadas–, teniendo en cuenta el valor de  $U_x$ , es decir, el cociente entre desviación típica para el rango total –el total de 2141 opositores aprobados– y la

desviación típica para el rango restringido –la muestra de 101 personas seleccionadas cuyo desempeño se pudo evaluar– (Thorndike, 1949). Los valores de Ux oscilaron entre 0.78 y 1.95, indicándose en la Tabla 1. Para llevar a cabo las pruebas de significación de las correlaciones corregidas se siguieron las indicaciones de Salgado (1997) y Bobko y Rieck (1980).

### Resultados

En las Tablas 1 y 2 se observan las medias y desviaciones típicas de los predictores y criterios empleados. En el caso de los procedimientos selectivos la media de la experiencia es de 6.42 (equivalente a 4.37 años de servicios en puestos similares). Por lo que se refiere a las medidas de desempeño, las medias son muy elevadas, estando cerca de la puntuación máxima de la escala, y las desviaciones típicas de DG, DT y DT han oscilado entre .54 y .70. La fiabilidad de estas medidas de desempeño fue adecuada (Nunnally, 1978), presentando valores superiores a .70 en todos los casos, con la excepción de persistencia/tenacidad.

**Tabla 1.** Medias, desviación típica, fiabilidad, valores de la restricción del rango (Ux), asimetría y curtosis de los predictores del estudio 1

Predictor	Alfa de Cronbach	Media	DT muestra	DT población	Ux	Asimetría	Curtosis
Conocimientos	.81	19.62	2.73	5.33	1.95	-0.30	-0.65
TJS	.51	27.52	3.75	5.23	1.40	0.49	-0.30
Experiencia	N/A	6.42	7.64	5.94	0.78	1.09	-0.14
Titulaciones	.75	2.81	2.00	1.80	0.90	0.90	-1.08
Informática	.78	1.80	0.49	0.56	1.13	-2.77	7.10
Idiomas	N/A	0.27	0.43	0.46	1.06	1.00	-0.95

Nota. N = 101. DT = desviación típica; Ux = valores de la restricción del rango directa en el predictor; TJS = Test de Juicio Situacional; N/A = no aplicable.

**Tabla 2.** Medias, desviación típica, fiabilidad, asimetría y curtosis de las medidas de criterio del estudio 1

Criterio	Alfa de Cronbach	Media	DT	Asimetría	Curtosis
Desempeño general	.92	4.39	0.54	-1.17	1.04
Desempeño de tarea	.89	4.44	0.55	-1.17	0.60
Atender consultas	.79	4.37	0.64	-0.79	-0.38
Comprobar datos	.73	4.48	0.55	-1.27	0.82
Alimentar bases de datos	.83	4.52	0.64	-1.47	1.50
Elaborar certificaciones	.74	4.63	0.53	-1.53	1.71
Tramitar expedientes	.82	4.28	0.72	-0.82	-0.02
Desempeño contextual	.86	4.18	0.70	-1.10	1.12
Ayuda y cooperación	.91	4.43	0.72	-1.85	3.88
Iniciativa	.90	3.77	0.96	-0.60	0.21
Persistencia/tenacidad	.63	4.30	0.76	-1.60	3.53
Autodesarrollo	.75	4.19	0.89	-1.02	0.03

Nota. N = 101. DT = desviación típica.

**Tabla 3.** Matriz de correlaciones observadas de las variables del estudio 1

Variable	Conocimientos	TJS	Experiencia	Titulación	Idiomas	DT	DC	DG
Conocimientos	1							
TJS	.40** (.22)	1						
Experiencia	-.31** (-.12)	-.13 (.07)	1					
Titulación	-.18 (.02)	.03 (-.16)	.14 (-.06)	1				
Idiomas	.15 (-.05)	-.12 (.08)	-.23** (-.04)	-.12 (.08)	1			
DT	.08 (-.12)	.03 (-.16)	.07 (-.13)	-.02 (.17)	.01 (-.18)	1		
DC	.03 (-.16)	.02 (-.17)	-.18 (.02)	-.03 (.17)	-.04 (.16)	.67** (.55)	1	
DG	.09 (-.11)	.06 (-.14)	-.01 (.18)	-.05 (.15)	-.01 (.18)	.98** (.97)	.80** (.71)	1

Nota. N = 101. Los límites inferiores del Intervalo de confianza del 95% aparecen entre paréntesis a la derecha de las correlaciones. TJS = Test de Juicio Situacional; DT = desempeño de tarea; DC = desempeño contextual; DG = desempeño general.

\*\* p < .01 (bilateral).

Los predictores del concurso de méritos presentan asimetría y/o curtosis, siendo el caso más destacado el de la variable conocimientos de informática. Lo mismo sucede con las variables que median el desempeño laboral. En este caso, la mayor parte de variables presentan distribuciones asimétricas negativas, es decir, la mayor parte de los evaluadores han empleado las puntuaciones más altas de la escala, las que indicaban un mejor desempeño. Asimismo, se observan distribuciones leptocúrticas –puntuaciones concentradas en torno a la media de la escala– en algunos de los criterios empleados, siendo los casos más pronunciados ayuda y cooperación y persistencia/tenacidad. En cualquier caso, todas las variables, con la excepción de conocimientos de informática, presentan valores inferiores a los límites que indican problemas de normalidad, 2 para asimetría y 7 para curtosis (Curran, West y Finch, 1996). Dados estos valores, no se han calculado los valores de las correlaciones para este predictor.

En las Tablas 3 y 4 se observan, respectivamente, la validez observada y la validez operativa (validez corregida por falta de fiabilidad de criterio y por restricción del rango) de los predictores empleados en la selección de auxiliares administrativos. Todos los límites de los intervalos de confianza de los predictores incluyeron el 0, por lo que no se obtuvieron correlaciones significativas. Por consiguiente, estos datos no permiten apoyar las hipótesis 1, 2, 3, 4 y 5.

En cuanto a la magnitud de las correlaciones corregidas encontradas, la prueba de conocimientos presentó un valor de .24 con DG como criterio y de .22 con DT a partir de los valores de las correlaciones observadas (correlaciones entre .08 y .09). EL TJS presentó correlaciones corregidas muy bajas (entre .05 y .12), la experiencia osciló entre -.25 y .07, las titulaciones académicas presentaron valores entre -.06 y -.02 y el conocimiento de idiomas osciló entre -.16 y .02.

En cuanto a las correlaciones corregidas entre las pruebas eliminatorias y las subdimensiones de las medidas de criterio, la prueba de conocimientos presenta correlaciones corregidas más elevadas con tramitar expedientes, .39, y autodesarrollo, .36, y los valores de las correlaciones observadas respectivas fueron .15 y .14. Del mismo modo, el TJS presenta las correlaciones corregidas más elevadas con tramitar expedientes, .35, y autodesarrollo, .28, y los valores de las correlaciones observadas respectivas fueron .18 y .14 (Tabla 5).

Ante la escasa variabilidad y las elevadas puntuaciones observadas en las medidas de criterio se midió el desempeño empleando cuestionarios de comparaciones y distribuciones forzadas y se obtuvieron puntuaciones en DG, DT y DC para los 27 auxiliares evaluados cuyos superiores tenían más de un evaluado, de manera que se pudiera observar el efecto de las comparaciones y distribuciones forzadas en las correlaciones entre predictores y criterios. En la Tabla 6 se observa cómo la validez operativa de la prueba de conocimientos asciende a .45 con DG como criterio, si bien no es significativa (p = .08). El resto de predictores presentaron límites del intervalo de confianza que incluían el 0.

**Tabla 4.** Validez operativa de los predictores del estudio 1

Predictor	Criterios		
	DT	DC	DG
Conocimientos	.22 (-.14)	.09 (-.29)	.24 (-.11)
TJS	.05 (-.21)	.04 (-.23)	.12 (-.16)
Experiencia	.07 (-.08)	-.25 (.02)	-.01 (.14)
Titulaciones	-.02 (.15)	-.03 (.14)	-.06 (.11)
Idiomas	.02 (-.19)	-.06 (.15)	-.16 (.19)

Nota.  $N = 101$ . Los límites inferiores del Intervalo de confianza del 95% aparecen entre paréntesis a la derecha de las correlaciones. TJS = Test de Juicio Situacional; DT = desempeño de tarea; DC = desempeño contextual; DG = desempeño general.

**Tabla 5.** Correlaciones de la prueba de conocimientos y el TJS con las subdimensiones de desempeño del estudio 1

Criterio	Predictor			
	Conocimientos		TJS	
	$r_o$	$r_{RR}$	$r_o$	$r_{RR}$
<i>Desempeño de tarea</i>				
Atender consultas	.01 (-.18)	.03 (-.35)	-.04 (.16)	-.08 (.19)
Comprobar datos	-.01 (.18)	-.02 (.35)	-.06 (.14)	-.11 (.15)
Alimentar bases de datos	.11 (-.09)	.28 (-.07)	.07 (-.13)	.13 (-.13)
Elaborar certificaciones	.03 (-.17)	.07 (-.29)	.05 (-.15)	.10 (-.17)
Tramitar expedientes	.15 (-.05)	.39 (-.02)	.18 (-.02)	.35 (-.02)
<i>Desempeño contextual</i>				
Ayuda y cooperación	.01 (-.18)	.02 (-.35)	.02 (-.18)	.03 (-.23)
Iniciativa	-.03 (.17)	-.08 (.29)	-.04 (.16)	-.08 (.19)
Persistencia tenacidad	.00 (-.19)	.00 (-.38)	-.05 (.15)	-.09 (.17)
Autodesarrollo	.14 (-.06)	.36 (-.04)	.15 (-.05)	.28 (-.03)

Nota.  $N = 101$ . Los límites inferiores del intervalo de confianza del 95% aparecen entre paréntesis a la derecha de las correlaciones;  $r_o$  = validez observada;  $r_{RR}$  = validez operativa (validez observada corregida por fiabilidad del criterio y restricción directa en el predictor).

**Tabla 6.** Validez observada y operativa de los predictores del estudio 1 empleando medidas de criterio con comparaciones y distribuciones forzosas

Predictor	Criterios					
	DG		DT		DC	
	$r_o$	$r_{RR}$	$r_o$	$r_{RR}$	$r_o$	$r_{RR}$
Conocimientos	.18 (-.02)	.45 (-.01)	.12 (-.08)	.31 (-.02)	.07 (-.13)	.19 (-.18)
TJS	.13 (-.07)	.24 (-.04)	.14 (-.06)	.26 (-.03)	.05 (-.15)	.10 (-.17)
Experiencia	-.10 (.10)	-.11 (.04)	-.06 (.14)	-.07 (.09)	-.16 (.04)	-.17 (.02)
Idiomas	-.04 (.16)	-.05 (.15)	-.02 (.18)	-.08 (.18)	.03 (-.17)	.04 (-.16)
Titulaciones	.10 (-.10)	.15 (-.09)	-.06 (.14)	-.12 (.10)	-.06 (.14)	-.07 (.10)

Nota.  $N = 101$ . Los límites inferiores del Intervalo de confianza del 95% aparecen entre paréntesis a la derecha de las correlaciones. TJS = Test de Juicio Situacional; DT = desempeño de tarea; DC = desempeño contextual; DG = desempeño general;  $r_o$  = validez observada;  $r_{RR}$  = validez operativa (validez observada corregida por fiabilidad del criterio y restricción directa en el predictor).

## Estudio 2

### Método

**Participantes.** La muestra estuvo compuesta por 101 personas (32 varones y 69 mujeres), con una edad media de 39.39 años ( $SD = 6.29$ ) que superaron el proceso para la selección de auxiliares administrativos en 2010, estaban trabajando en el Gobierno Vasco durante el año 2016 y aceptaron cumplimentar el cuestionario de evaluación de las reacciones de los opositores. Esta muestra no coincide con la del estudio 1 más que en 34 sujetos, los de aquellos auxiliares cuyos superiores directos, en este caso 10 jefes de servicio, aceptaron evaluar el desempeño laboral de sus subordinados.

**Medidas.** En este estudio se han empleado las siguientes medidas:

**Predictores.** De los instrumentos selectivos empleados en la oposición del Gobierno Vasco para la plaza de auxiliar administrativo, si-

guiendo el modelo de McCarthy et al. (2013), únicamente se tuvieron en cuenta las pruebas con carácter eliminatorio, es decir, la prueba de conocimientos y el TJS, ambos descritos en el estudio 1.

**Reacciones de los candidatos.** Las reacciones de los candidatos se midieron mediante un cuestionario de 5 ítems que evaluaron respectivamente la adecuación de la dificultad de las pruebas eliminatorias (conocimientos y TJS), el grado de adecuación al temario de las pruebas eliminatorias, el grado de adecuación a las tareas del puesto de la prueba de conocimientos, el grado de adecuación a las tareas del puesto del TJS y el nivel de satisfacción global con respecto a la justicia procedimental del proceso selectivo y la adecuación de las pruebas empleadas. Los 4 primeros ítems se evaluaron mediante una escala de 5 puntos (de 1 a 5), reflejando 1 un *nivel muy inadecuado* y 5 *muy adecuado*, y el ítem relativo a la justicia procedimental se evaluó de 0 a 10. La consistencia interna de la escala fue de .83.

Además, para poder interpretar de manera adecuada las valoraciones cuantitativas de los auxiliares, estos tuvieron la oportunidad de manifestar sus impresiones mediante un ítem de respuesta abierta. Para analizar con mayor precisión las respuestas cualitativas de los auxiliares, se realizó un análisis de contenido de sus valoraciones dadas en dicho ítem (Graneheim y Lundman, 2004).

**Evaluación del desempeño.** Se evaluó el desempeño de cada participante mediante el cuestionario de evaluación del desempeño convencional basado en anclajes conductuales descrito en el estudio 1. No se emplearon los cuestionarios de comparaciones y distribuciones forzosas.

**Procedimiento.** Se solicitó a las 339 personas que estaban trabajando en el Gobierno Vasco tras haber aprobado el proceso selectivo de 2010, es decir, tras haber superado la prueba de conocimientos y el TJS, que respondiesen al cuestionario para evaluar sus percepciones con respecto a la justicia de las pruebas y del procedimiento selectivo del año 2016. El cuestionario fue cumplimentado en su lugar de trabajo por los 101 auxiliares administrativos que aceptaron participar mediante una aplicación que garantizaba el anonimato en sus respuestas. Aproximadamente tres meses después su desempeño laboral fue evaluado por parte de sus superiores directos siguiendo el procedimiento descrito en el estudio 1.

Las reacciones de los candidatos se pudieron correlacionar con sus resultados en el proceso selectivo, que ya figuraban en el registro del IVAP. De los 101 auxiliares que rellenaron el cuestionario de reacciones 34 fueron evaluados por sus superiores mediante el cuestionario de evaluación de desempeño.

## Resultados

Del análisis de las reacciones de los auxiliares cabe concluir que su valoración de la adecuación de la dificultad, adecuación al temario, adecuación a las tareas de la prueba de conocimientos y nivel de satisfacción general con la equidad del proceso es positiva, ya que las puntuaciones medias en estos aspectos están en el punto medio de las respectivas escalas o ligeramente por encima. No sucede lo mismo en cuanto a la valoración de la prueba TJS. En este caso la media es de 2.74 (Tabla 7).

Por lo que se refiere a las correlaciones entre las variables empleadas en el estudio 2, como se puede comprobar en la Tabla 8, el resultado en la prueba TJS presenta una correlación significativa con la adecuación de la dificultad percibida del proceso selectivo, .28, la adecuación al temario, .33, y la adecuación de la prueba de TJS a las tareas del puesto, .34. Sin embargo, el resultado en la prueba de conocimientos no presenta ninguna correlación significativa con las reacciones de los candidatos. En consecuencia, la hipótesis 6 se cumple parcialmente solo para el TJS.

En cuanto a la relación entre reacciones y evaluación del desempeño, DG presenta correlaciones significativas con la adecuación de la dificultad percibida del proceso selectivo y la adecuación al temario, con un valor de .41 en ambos casos. Por consiguiente, la hipótesis 7 se cumple parcialmente.

**Tabla 7.** Medias, desviaciones típicas, tamaño muestral, asimetría y curtosis de las medidas empleadas en el estudio 2

Variable	N	Media	DT	Asimetría	Curtosis
<i>Reacciones</i>					
Adecuación de la dificultad	101	3.27	0.98	-0.56	-0.35
Adecuación al temario	101	3.40	0.92	-0.40	-0.66
Adecuación Prueba de conocimientos a tareas	101	3.00	0.94	-0.22	-0.40
Adecuación Test Juicio Situacional a tareas	101	2.74	1.17	0.14	-0.84
Nivel de satisfacción global	101	5.61	2.14	-0.44	-0.49
<i>Pruebas selectivas</i>					
Prueba de conocimientos	101	19.46	2.76	-0.26	-0.74
Test de juicio situacional	101	27.59	3.74	0.32	-0.69
<i>Desempeño laboral</i>					
Desempeño de tarea	34	5.99	7.61	-1.16	0.71
Desempeño contextual	34	63.18	7.67	-1.41	2.16
Desempeño general	34	4.42	0.53	-1.38	2.01

Nota. DT = desviación típica.

**Tabla 8.** Matriz de correlaciones de las medidas empleadas en el estudio 2

Variable	N	1	2	3	4	5	6	7	8	9	10
<i>Reacciones</i>											
1. AdDif	101	1									
2. AdTem	101	.74**	1								
3. ConTar	101	.54**	.42**	1							
4. TJS Tar	101	.63**	.65**	.56**	1						
5. NivSat	101	.78**	.71**	.58**	.71**	1					
<i>Pruebas selectivas</i>											
6. Conocimientos	101	-.03	-.08	.02	-.17	-.11	1				
7. TJS	101	.28**	.33**	.16	.34**	.19	.52**	1			
<i>Desempeño laboral</i>											
8. DT	34	.33	.27	.13	-.05	-.02	.07	.03	1		
9. DC	34	.33	.31	.24	.08	.02	.02	.03	.82**	1	
10. DG	34	.41*	.41*	.30	.08	.06	.08	.07	.94**	.97**	1

Nota. AdDif = adecuación de la dificultad de las pruebas selectivas; AdTem = adecuación al temario de las pruebas empleadas; ConTar = adecuación de la prueba de conocimientos a las tareas del puesto; TJS Tar = adecuación del TJS a las tareas del puesto; NivSat = nivel de satisfacción con el proceso selectivo; TJS = Test de Juicio Situacional; DT = desempeño de tarea; DC = desempeño contextual; DG = desempeño general.

Por último, la mayor parte de las valoraciones (más del 46% del total) expresadas por los auxiliares en el ítem de respuesta libre reflejan desacuerdo con las características del TJS, especialmente en cuanto a su desajuste con las funciones y tareas del puesto, así como en cuanto a la información dada sobre sus características antes de la celebración de la misma.

### Discusión General

Se ha encontrado una baja validez predictiva para la mayor parte de los aspectos valorados en el proceso selectivo, siendo la prueba de conocimientos la que presenta magnitudes más altas, con un valor corregido idéntico, .24, al encontrado por Berges (2011) con una muestra de policías, pero muy inferior al apuntado por el metaanálisis de Dye et al. (1993), .45. No obstante, se aproxima al valor hallado por dicho metaanálisis para puestos de baja complejidad y pruebas de moderada similitud al contenido del puesto, aspecto este último que coincide con las reacciones indicadas por los auxiliares de la muestra del estudio 2.

Esta diferencia entre el valor hallado y el del metaanálisis puede deberse a que, como señala Táuriz (2014), los exámenes de oposición muchas veces incluyen contenidos de tipo jurídico que no están directamente relacionados con el puesto en sí, por lo que cabe esperar que su validez predictiva sea inferior a .45. Esta cuestión apunta a la

necesidad de realizar estudios de validez de contenido para mejorar la adecuación de las pruebas al contenido de los puestos ofertados según las directrices de Lawshe (1975), Choragwicka (2007) o Choragwicka y Moscoso (2007).

Esto es relevante ya que la similitud o la correspondencia conceptual entre el dominio del constructo de los predictores y el dominio del constructo de los criterios aumenta la magnitud del coeficiente de validez (Binning y Barrett, 1989; Christian et al., 2010; Lievens, Buyse y Sackett, 2005). En esta investigación tanto la prueba de conocimientos como el TJS presentan una correlación de tamaño medio a la hora de predecir tramitación de expedientes, la medida de criterio más relacionada con el importante peso concedido en el temario al conocimiento de la legislación sobre procedimiento administrativo.

Asimismo, tanto la prueba de conocimientos como el TJS presentan correlaciones de tamaño medio con autodesarrollo, criterio que medía el interés por la formación y la adquisición de conocimientos y podría vincularse con el aprovechamiento formativo. La magnitud de la correlación corregida encontrada con la prueba de conocimientos como predictor se aproxima a la encontrada por el metaanálisis de Dye et al. en 1993 para predecir el aprovechamiento formativo, .47.

La prueba de TJS no presenta correlaciones distintas de 0 con DG, DT o DC en nuestra muestra, alcanzando valores corregidos entre .05 y .12, muy inferiores a los encontrados para el conjunto de TJS por el metaanálisis de McDaniel y Nguyen (2001), .34. Una posible explicación es que el TJS empleado en esta investigación medía exclusivamente conocimientos y, como hallaron Christian et al. (2010), la validez de criterio de los TJS varía en función del constructo medido. En dicho metaanálisis los TJS que medían constructos como destrezas interpersonales, liderazgo o personalidad presentaron valores que oscilaron entre .25 y .43, mientras que los que medían conocimientos tuvieron el valor más bajo de todos los constructos medidos, .19.

La experiencia, por su parte, presenta valores muy bajos en cuanto a su validez predictiva del desempeño general y de tarea, presentando unos valores cercanos a cero, lo que coincide con lo apuntado por Sturman (2003) o McDaniel y Nguyen (2001) cuando los seleccionados son en gran parte los que tienen mucha experiencia en el puesto, como en este caso, y además se corresponde con los datos hallados por el estudio de Berges (2011) en una muestra de empleados públicos.

Por otro lado, las evaluaciones de desempeño con comparaciones forzadas han ayudado a elevar el tamaño de la validez operativa de la prueba de conocimientos hasta .45 y la del TJS hasta .24, algo que cabía esperar en base a lo apuntado por Berger, Harbring y Sliwka (2013), que encontraron que este tipo de evaluación del desempeño contribuía a evitar el sesgo de benevolencia y a discriminar mejor entre los evaluados, algo que contribuye a aumentar al tamaño de la correlación (Smith y Smith, 2005).

La magnitud de las correlaciones de las titulaciones académicas ha sido baja, si bien era algo que cabía esperar en base a los datos metaanalíticos. Asimismo, se han obtenido valores bajos en el conocimiento de idiomas, por lo que, dada la escasez de investigaciones, es preciso seguir profundizando en el conocimiento de las relaciones entre esta variable y el desempeño laboral. En ambos casos, como señala Schneider (1994), cabe argumentar que no se puede presumir que todas las personas que hayan acreditado un título o certificado formativo tengan el mismo nivel de destrezas laborales y, como señalan Dye y Reck (1988), el nivel educativo, más allá de cierto punto no produce mejoras en el desempeño.

En cuanto a los resultados del estudio 2, cabe destacar que la puntuación en TJS presenta una correlación significativa con adecuación de la dificultad, adecuación al temario y adecuación al contenido del puesto, con magnitudes superiores a las encontradas por el metaanálisis de Hausknecht et al. (2004). No ha sucedido lo mismo con la prueba de conocimientos, que no presenta correlaciones significativas con ninguno de los ítems que medían las reacciones de los auxiliares.

Una posible explicación la encontramos en los trabajos de Bauer

y Truxillo (2006) o Chan, Schmitt, Sacco y DeShon (1998), que indican que las reacciones negativas ante pruebas que miden desempeño máximo, es decir, pruebas que exigen el máximo esfuerzo posible a los candidatos (Klehe y Anderson, 2007), pueden provocar un descenso en la motivación para rendir adecuadamente. La media de las reacciones ante el TJS es inferior a la de la prueba de conocimientos, por lo que cabría plantear que las percepciones más negativas a esta prueba hayan tenido un efecto perjudicial en los resultados.

Esta reacción negativa, expresada por un 46% de auxiliares en el ítem de respuesta abierta, es algo que investigaciones como la de Patterson et al. (2011) han comprobado al abordar la selección de profesionales de medicina general en el National Health Service (NHS) británico. Al introducir los TJS a finales de la década pasada, fueron inicialmente percibidos por los candidatos como pruebas de menor validez aparente que las pruebas de conocimientos, aunque, tras ofrecer más información con respecto a los TJS en años posteriores, estos fueron evaluados de manera más positiva en cuanto a su relación con el puesto.

No obstante, también hay que considerar el posible sesgo de *interés personal* [self-serving bias] señalado por Chan y Schmitt (1997). Es decir, cabe considerar que sean los resultados en la prueba los que determinen las percepciones acerca de la prueba, teniendo en cuenta que la medición de las reacciones de los auxiliares se hizo una vez conocidos los resultados. Sin embargo, el hecho de que las correlaciones únicamente hayan sido significativas en el caso de la prueba TJS hace que cobre más fuerza la hipótesis que indica que son las reacciones las que provocan los resultados.

Más relevante si cabe es la correlación significativa que presentan adecuación de la dificultad y adecuación al temario con el desempeño general, lo que indica una posible relación directa entre reacciones y desempeño, posibilidad ya señalada por Colquitt et al. (2001) o McCarthy et al. (2013). Como señalan estos últimos autores, una posible interpretación es que las personas que no creen en la validez de las pruebas selectivas puedan desarrollar desconfianza hacia las evaluaciones, las políticas o los procedimientos de la organización, lo que a su vez está relacionado con la *teoría de la equidad heurística* [fairness heuristic theory], en la que Lind (2011) sugiere que las personas se forman impresiones de equidad de manera temprana y que estas impresiones moldean las percepciones y las conductas posteriores, por lo que las reacciones a las pruebas pueden tener consecuencias en las conductas posteriores a la contratación.

### Limitaciones y Futuras Investigaciones

Como limitación del presente trabajo es necesario mencionar que la muestra se ha formado exclusivamente con auxiliares administrativos, por lo que es preciso realizar estudios similares para otros puestos del sector público para confirmar lo encontrado en esta investigación.

El tamaño de la muestra en el estudio 2 es pequeño y consecuentemente los resultados encontrados pueden variar si la muestra aumenta, aunque la magnitud de las correlaciones halladas es un factor que apunta a seguir investigando en la línea de lo planteado por este trabajo en futuras investigaciones de índole similar.

Por otro lado, el hecho de haber recogido las reacciones de los auxiliares, tras haber realizado los tests y teniendo conocimiento de sus resultados, aunque permite que los participantes puedan evaluar de una manera adecuada la relación entre los contenidos del puesto y el test, también puede provocar el sesgo de interés personal, algo que se puede aclarar siguiendo el enfoque de Chan et al. (1998) y midiendo también las reacciones antes de realizar los tests. Además, el lapso de tiempo transcurrido entre la celebración del proceso selectivo y la obtención de los datos del estudio 2, 6 años, hace que no sea posible descartar la acción de variables extrañas que hayan podido afectar a los resultados, como satisfac-

ción laboral, relaciones con los superiores, etc. Por ello, en futuros estudios convendría reducir el tiempo entre la celebración de las pruebas y la obtención de las reacciones de los empleados.

### Implicaciones Prácticas

Este trabajo es una de las escasas investigaciones en el ámbito público en cuanto a validez predictiva y reacciones de los candidatos, y tiene interesantes repercusiones prácticas que pueden llevar a la reflexión en cuanto al tipo de procedimientos selectivos a emplear en el sector público y en cuanto a la importancia relativa que se debe conceder a cada prueba, habida cuenta de la mayor capacidad predictiva de la prueba de conocimientos frente a la prueba de carácter más práctico, algo que coincide por lo señalado por el metaanálisis (Schmidt et al., 2016).

Además, se antoja necesario reflexionar sobre el contenido de los temarios y los exámenes a aplicar, que deberían presentar una mayor similitud con las tareas de los puestos, algo que lleva a recomendar el uso de estudios de validez de contenido, y también sobre la manera de evaluar el desempeño, primando herramientas con comparaciones y distribuciones forzadas que ayuden a combatir los sesgos habituales en su uso en la Administración Pública, puntuaciones elevadas e iguales para todo el personal evaluado.

Por último, la relación observada entre las reacciones de los participantes en los procesos selectivos y sus resultados en las pruebas y sobre todo en el desempeño es una cuestión muy a tener en cuenta en el diseño de los mismos, por lo que sería positivo poder realizar estudios similares en el futuro.

### Extended Summary

The Basque Institute of Public Administration (IVAP) studied two crucial matters in the selection procedures for the public sector with a sample of administrative assistants: the predictive validity of selection tests and the way applicants perceive them.

The Spanish legislation urges that selection should be made to predict future job performance and that procedures should be acceptable in terms of their psychometric properties of predictive validity, an issue which has not been frequently investigated in Spain with usual predictors such as knowledge tests or job-experience evaluations.

Furthermore, results of different studies suggest there is a positive relationship between perception of procedural justice and test performance, i.e., people who have positive reactions to selection procedures obtain higher scores in those procedures and vice versa (Ryan, Sacco, McFarland, & Kriska, 2000; Smither, Reilly, Millsap, Pearlman & Stofey, 1993; Truxillo, Steiner, & Gilliland, 2004). Nevertheless, although some studies have emphasized the importance of analyzing the relationship between candidate perceptions and job performance (Chan & Schmitt, 2004; Ryan & Ployhart, 2000), this issue has seldom been studied. McCarthy et al. (2013) found that situational reactions affect job performance indirectly through their influence on the scores obtained.

A literature review led to the following hypotheses:

H1. Job knowledge test (JKT) scores will show a positive correlation with job performance: task performance (TP), contextual performance (CP), and overall performance (OP) of hired personnel.

H2. The Situational Judgment Test (SJT) scores will show a positive correlation with job performance of hired personnel.

H3. Job-experience evaluation will have a positive correlation with job performance of hired personnel.

H4. The evaluation of academic degrees attained will have a positive correlation with job performance of hired personnel.

H5. The evaluation of foreign language knowledge will have a positive correlation with job performance of hired personnel.

H6. Candidates' reactions (adequacy of tests' difficulty, adjustment to the tests' list of subjects, job-relatedness and procedural justice) will have a positive correlation with selection tests' scores (JKT and SJT) of administrative staff.

H7. Candidates' reactions will have a positive correlation with job performance (TP, CP and OP) of administrative staff.

## Study 1

### Method

The sample consisted of 101 people (23 men and 78 women), with a mean age of 39.5 years ( $SD = 6.03$ ) who passed a competition to hire administrative assistants and were working for the Basque Government during 2016 and whose superiors agreed to evaluate their performance.

The predictors used were two pass/fail tests (a JKT and a SJT), job experience, academic degrees, computer knowledge, and foreign languages' knowledge. Job performance was evaluated with a behaviourally-anchored rating scale (BARS) that measured TP (answering queries and complaints; data, records and documentation checking; feed and update databases and applications; elaboration of certifications, notifications, compulsions, etc.; processing administrative files) and CP (aid and cooperation; initiative; persistence-tenacity; self-development). Aside from this conventional questionnaire, two other types of job performance questionnaires based on rankings and forced distributions were also used to counter evaluators' biases: one for superiors with two to four subordinates, and the other for superiors with more than five subordinates. The average scores for TP and CP allowed obtaining an Overall Performance (OP) value. To accurately study the relationship between predictors and criteria when using different instruments to assess job performance in the same sample of subjects, scores were transformed into z-scores and partial correlation was used to eliminate the effect of the type of instrument.

### Results

No support was found for H1, H2, H3, H4 and H5, as all predictors had non-significant correlations. The JKT showed the highest positive corrected correlation with both OP, .24, and TP, .22, the SJT had corrected correlations between .05 and .12, and job-experience had values close to 0 (Table 1).

As for the corrected correlations between eliminatory tests and the subdimensions of criterion measures, the highest correlations were found between JKT and processing administrative files, .39, and self-development, .36, and between SJT and processing administrative files, .35, and self-development, .28 (Table 5).

Given the low variability and the high scores observed in the criterion measures (Table 2) performance was measured using rankings and forced distributions' questionnaires and scores were obtained in OP, TP and CP for 27 workers. These results were considered along with performance assessed with the conventional, BARS-based questionnaire for the other 74 workers. Table 6 shows that the operational validity of the JKT rose to .45 with OP as a criterion.

## Study 2

### Method

The sample consisted of 101 administrative assistants (32 men and 69 women), with a mean age of 39.39 years ( $SD = 6.29$ ) who passed a competition and were working for the Basque Government during 2016 and agreed to evaluate their reactions. Job performance of 34 of these 101 administrative assistants was assessed by 10 superiors

who agreed to utilize the conventional questionnaires mentioned in study 1.

Candidates' reactions were measured with a 5-item questionnaire, which evaluated the adequacy of the difficulty of the screening phase tests, their degree of adjustment to the tests' subject list, job relatedness of the JKT, job relatedness of the SJT, and overall satisfaction regarding the procedural fairness of the selection process and the adequacy of the tests used. The first four items were evaluated with a 5-point scale, while the item related to procedural justice was evaluated from 0 to 10. The internal consistency of the scale was .83.

### Results

Partial support for H6 was found as SJT scores presented significant correlations with adequacy of the perceived difficulty of the selective process, adjustment to the subjects' list and job relatedness of the SJT. However, scores in the JKT did not show significant correlations with candidates' reactions.

Regarding the relationship between reactions and performance, OP presents significant correlations with the adequacy of the perceived difficulty of the selection process and the adjustment to the subjects' list, which partially supports H7 (Table 8).

## Discussion

The JKT's corrected value found for predictive validity, .24, is identical to that found by Berges (2011), but much lower than the one pointed out by the meta-analysis of Dye, Reck, & McDaniel (1993), .45. However, it is close to the value found by this meta-analysis for low complexity jobs and tests of moderate job-relatedness, which is in line with the reactions of the participants of study 2.

An argument to explain the difference between the value found and that of meta-analysis of JKTs, as Tauriz (2014) points out, is that these tests often include content that is not job-related, i.e., constitutional or administrative law, so it can be expected that its predictive validity is less than .45. This points to the need of carrying out content validity studies to improve the job-content adjustment of tests.

Nevertheless, low predictive validity has been found for all the other predictors used. The SJT does not present correlations different from 0 with the criteria in our sample, with values far from the magnitude found by the meta-analysis by McDaniel & Nguyen (2001), .34. A possible explanation is that SJTs that measure knowledge have criterion validity estimates lower than those measuring constructs such as interpersonal skills, personality, or leadership (Christian et al., 2010).

Furthermore, the predictive validity of experience regarding OP and TP is close to zero, which is consistent with Sturman's (2003) or McDaniel & Nguyen's (2001) findings, since applicants selected were largely those with more job experience, and it matches the findings of Berges (2011) with a sample of public employees.

Both academic degrees and foreign language presented low correlations with criteria, which is in line with previous research (Dye & Reck, 1988). It could be argued that it cannot be presumed that all people who have attained a degree or certificate of training have the same level of work skills, and, as pointed out by Dye and Reck (1988), the educational level, beyond a certain point, does not lead to improvements in performance (Schneider, 1994).

An aspect that may have influenced the predictive validity of the predictors employed is the similarity or conceptual correspondence between the predictor construct domain and the construct domain of the criteria (Binning & Barrett 1989; Christian, Edwards, & Bradley, 2010; Lievens, Buyse, & Sackett, 2005). We must emphasize that both the JKT and the SJT have positive correlations when predicting processing administrative files, the criterion measure most related to the important weight given in the subjects' list to the knowledge of administrative procedure.

On the other hand, using performance appraisals with forced comparisons and rankings helped to increase the operational validity of the JKT and the SJT to .45 and .24 respectively, something that could be expected based on what was pointed out by Berger, Harbring, and Sliwka (2013). These results indicate that this type of appraisals could help to overcome the biases often present in conventional performance appraisals and should encourage the use of questionnaires with forced comparisons and rankings in future research.

As for study 2, it is worth noting that scores in the SJT present significant correlations with adequacy of the difficulty, adjustment to the subjects' list and job-relatedness, with magnitudes higher than those found by Hausknecht, Day, and Thomas' (2004) meta-analysis. Even more relevant is the significant correlation between adequacy of the difficulty and adjustment of the subject list with OP, indicating a possible direct relationship between reactions and performance, a possibility already pointed out by Colquitt, Conlon, Wesson, Porter, and Ng (2001) or McCarthy et al. (2013).

Nevertheless, the findings of this research, given the small size of the sample used in study 2, should be confirmed with later studies. This limitation, along with the time elapsed from the selection process to the implementation of study 2, 6 years, makes it impossible to rule out the action of extraneous variables that may have affected the results, something that must be considered in the interpretation of this results.

This paper is one of the few studies in the Spanish public sector on predictive validity and candidates' reactions, and has interesting practical repercussions: it should foster reflection on the selection procedures to be used as well as on the content of tests to be utilized, which should be more job-related, and at the same time, it should contribute to stress the relevance of both transparency of selection and participation of job incumbents in the design of tests in order to enhance their content validity.

### Conflicto de Intereses

Los autores de este artículo declaran que no tienen ningún conflicto de intereses.

### Referencias

- Aguinis, H. (2013). *Performance management*. Upper Saddle River, NJ: Pearson.
- Alonso, P., Táuriz, G. y Choragwicka, B. (2009). Valoraciones de méritos (Training and experience) en la Administración Pública y la empresa: fiabilidad, validez y discriminación de género. *Revista de Psicología del Trabajo y de las Organizaciones*, 25, 245-258.
- Anderson, N., Salgado, J. F. y Hülshager, U. R. (2010). Applicant Reactions in Selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment*, 18, 291-304. <https://doi.org/10.1111/j.1468-2389.2010.00512.x>
- Arvey, R. D., Strickland, W., Drauden, G., y Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, 43, 695-716. <https://doi.org/10.1111/j.1744-6570.1990.tb00679.x>
- Austin, J. T., Klimoski, R. J. y Hunt, S. T. (1996). Dilemmatics in public sector assessment: A framework for developing and evaluating selection systems. *Human Performance*, 9, 177-198. [https://doi.org/10.1207/s15327043hup0903\\_1](https://doi.org/10.1207/s15327043hup0903_1)
- Basoredo, C. y López Basterra, J. (2013). Bases de un modelo de evaluación de méritos profesionales en la administración pública, centrado en el desempeño. *Revista Vasca de Gestión de Personas y Organizaciones Públicas*, 5, 8-19.
- Bauer, T. N. y Truxillo, D. M. (2006). Applicant reactions to situational judgment tests: Research and related practical issues. En J. A. Weekley y R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 233-249). Mahwah, NJ: Erlbaum.
- Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P. y Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel psychology*, 54, 387-419. <https://doi.org/10.1111/j.1744-6570.2001.tb00097.x>
- Berger, J., Harbring, C. y Sliwka, D. (2013). Performance appraisals and the impact of forced distribution-An experimental investigation. *Management Science*, 59, 54-68. <https://doi.org/10.1287/mnsc.1120.1624>
- Berges, A. (2011). Validación de procedimientos de selección de policías locales. *Anuario de Psicología Jurídica*, 21, 83-96. <https://doi.org/10.5093/jr2011v21a8>
- Bernardin, H. J. y Beatty, R. W. (1984). *Performance appraisal: Assessing human behavior at work*. Boston, MA: Kent Pub. Co.
- 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. <https://doi.org/10.1037/0021-9010.74.3.478>
- Bobko, P. y Rieck, A. (1980). Large sample estimators for standard errors of functions of correlation coefficients. *Applied Psychological Measurement*, 4, 385-398.
- 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(1-2), 52-69. <https://doi.org/10.1111/1468-2389.00163>
- Carr, V. y Patterson, F. (2009). *Developing assessment centre methodologies for high-stakes selection: Implications for policy*. Conferencia presentada en British Psychological Society, Division of Occupational Psychology Conference. Blackpool, UK.
- Chan, D. y Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82, 143-159. <https://doi.org/10.1037/0021-9010.82.1.143>
- Chan, D. y Schmitt, N. (2004). An agenda for future research on applicant reactions to selection procedures: A construct-oriented approach. *International Journal of Selection and Assessment*, 12, 9-23. <https://doi.org/10.1111/j.0965-075X.2004.00260.x>
- Chan, D., Schmitt, N., Sacco, J. M. y DeShon, R. P. (1998). Understanding pretest and posttest reactions to cognitive ability and personality tests. *Journal of Applied Psychology*, 83, 471-485. <https://doi.org/10.1037/0021-9010.83.3.471>
- Chaves, J. R. (2009). *Control de concursos y oposiciones en la jurisprudencia*. Madrid, España: Editorial Reus.
- Choragwicka, B. (2007). *Comparación de dos métodos de estimación de validez de contenido*. Comunicación presentada en el X Congreso Nacional de Psicología Social. Cádiz, España.
- Choragwicka, B. y Moscoso, S. (2007). Validez de contenido de una entrevista conductual estructurada. *Revista de Psicología del Trabajo y de las Organizaciones*, 23, 75-92.
- Christian, M. S., Edwards, B. D. y Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, 63, 83-117. <https://doi.org/10.1111/j.1744-6570.2009.01163.x>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Mahwah, NJ: LEA.
- Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. y Ng, K. Y. (2001). Justice at the millennium: a meta-analytic review of 25 years of organizational justice research. *Journal of applied psychology*, 86, 425-445. <https://doi.org/10.1037/0021-9010.86.3.425>
- Cook, C. L. (1980). *Rating education, training and experience in the public sector*. Artículo presentado en la conferencia anual de la "International Personnel Management Association". Boston, Estados Unidos.
- Curran, P. J., West, S. G. y Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1, 16-29. <https://doi.org/10.1037/1082-989X.1.1.16>
- Díaz Cabrera, M. D., Hernández, E., Isla, M. R., Delgado, N., Díaz Vilela, L. y Rosales, C. (2014). Factores relevantes para aumentar la precisión, la viabilidad y el éxito de los sistemas de evaluación del desempeño laboral. *Papeles del psicólogo*, 35, 115-121.
- Dye, D. A. y Reck, M. (1988). *A literature review and meta-analysis of education as a predictor of job performance*. Springfield, VA: US Office of Personnel Management.
- Dye, D. A., Reck, M. y McDaniel, M. A. (1993). The validity of job knowledge measures. *International Journal of Selection and Assessment*, 1, 153-157. <https://doi.org/10.1111/j.1468-2389.1993.tb00103.x>
- García-Izquierdo, A. L., y García-Izquierdo, M. (2007). Discriminación, igualdad de oportunidades en el empleo y selección de personal en España. *Revista de Psicología del Trabajo y de las Organizaciones*, 23, 111-138.
- Gilliland, S.W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review*, 18(4), 694-734. <https://doi.org/10.5465/AMR.1993.9402210155>
- Gilliland, S.W. (1994). Effects of procedural and distributive justice reactions to a selection system. *Journal of Applied Psychology*, 79, 691-701. <https://doi.org/10.1037/0021-9010.79.5.691>
- Gilliland, S. W., y Steiner, D. D. (2014). Applicant reactions to testing and selection. En N. Schmitt, (Ed.), *The Oxford handbook of personnel assessment and selection* (pp.629-666). New York, NY: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199732579.013.0028>
- Gorriti, M. y López Basterra, J. (2009). *Identificación de criterios de desempeño a partir del análisis de puestos*. Oñati, España: IVAP
- Gorriti, M. y López Basterra, J. (2010). *Análisis de puestos en la Administración Pública*. Oñati, España: IVAP.
- Gorriti, M. y López Basterra, J. (2014). *Guía para la validación de sistemas de acceso a la Función Pública*. Oñati, España: IVAP.

- Graneheim, U. H. y Lundman, B. (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse education today*, 24, 105-112. <https://doi.org/10.1016/j.nedt.2003.10.001>
- Harris, L. (2000). Procedural justice and perceptions of fairness in selection practice. *International Journal of Selection and Assessment*, 8, 148-157. <https://doi.org/10.1111/1468-2389.00143>
- Hausknecht, J. P., Day, D. V. y Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57, 639-683. <https://doi.org/10.1111/j.1744-6570.2004.00003.x>
- Hong, E. (1999). Test anxiety, perceived test difficulty, and test performance: Temporal patterns of their effects. *Learning and Individual Differences*, 11, 431-447. [https://doi.org/10.1016/S1041-6080\(99\)80012-0](https://doi.org/10.1016/S1041-6080(99)80012-0)
- Instituto Nacional de Administración Pública - INAP. (2012). *Modelos comparados de selección de empleados públicos*. Recuperado de [http://bci.inap.es/alfresco\\_file/30309fad-762b-43f4-8c87-9c0877de629b](http://bci.inap.es/alfresco_file/30309fad-762b-43f4-8c87-9c0877de629b)
- Klehe, U. C. y Anderson, N. (2007). El rol del desempeño típico y máximo en selección de personal. *Revista de Psicología del Trabajo y de las Organizaciones*, 23, 151-178.
- Kortmann, A. K. (2016). *The relationship between multicultural competence, foreign language mastery and job performance* (tesis doctoral). Radboud University Nijmegen, Países Bajos.
- Krantowitz, T. M. (2014). *Global assessment trends report*. Surrey, UK: CEB-SHL.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563-575.
- Lievens, F. (2007). Tests de Juicio Situacional: introducción y revisión de la investigación. *Revista de Psicología del Trabajo y de las Organizaciones*, 23, 93-110.
- Lievens, F., Buysse, T. y Sackett, P. R. (2005). The operational validity of a video-based situational judgment test for medical college admissions: illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology*, 90, 442-452. <https://doi.org/10.1037/0021-9010.90.3.442>
- Lind, E. A. (2001). Fairness heuristic theory: Justice judgments as pivotal cognitions in organizational relations. En J. Greenberg y R. Cropanzano (Eds.), *Advances in organizational justice* (pp 56-88). Stanford, CA: Stanford University Press.
- Mauri, J. (2009) *La selecció dels funcionaris al servei de les entitats locals de Catalunya*. Barcelona, España: Generalitat de Catalunya.
- McCarthy, J., Hrablucik, C. y Jelley, R. B. (2009). Progression through the ranks: assessing employee reactions to high-stakes employment testing. *Personnel Psychology*, 62, 793-832. <https://doi.org/10.1111/j.1744-6570.2009.01158.x>
- McCarthy, J. M., Van Iddekinge, C. H., Lievens, F., Kung, M.-C., Sinar, E. F. y Campion, M. A. (2013). Do candidate reactions relate to job performance or affect criterion-related validity? A multistudy investigation of relations among reactions, selection test scores, and job performance. *Journal of Applied Psychology*, 98, 701-709. <https://doi.org/10.1037/a0034089>
- McDaniel, M. A. y Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*, 9, 103-113. <https://doi.org/10.1111/1468-2389.00167>
- Mol, S. T., Born, M. P., Willemsen, M. E. y Van Der Molen, H. T. (2005). Predicting expatriate job performance for selection purposes: A quantitative review. *Journal of Cross-Cultural Psychology*, 36, 590-620.
- Murphy, K. R. y Cleveland, J. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage.
- Nunnally, J. C. (1978). *Psychometric theory* (2<sup>nd</sup> ed.). New York, NY: McGraw-Hill.
- Osca, A. y Salmones, L. G. (2004). ¿Cómo perciben los candidatos los procesos de selección? Una aproximación desde el modelo de Justicia Procedimental de Gilliland (1993). *Revista de Psicología del Trabajo y de las Organizaciones*, 20, 225-247.
- Patterson, F., Lievens, F., Kerrin, M., Zibarras, L. y Carrette, B. (2012). Designing Selection Systems for Medicine: The importance of balancing predictive and political validity in high-stakes selection contexts. *International Journal of Selection and Assessment*, 20, 486-496. <https://doi.org/10.1111/ijsa.12011>
- Patterson, F., Zibarras, L., Carr, V., Irish, B. y Gregory, S. (2011). Evaluating candidate reactions to selection practices using organisational justice theory. *Medical education*, 45, 289-297. <https://doi.org/10.1111/j.1365-2923.2010.03808.x>
- Peiró, J. M., Ramos, J. y Cortés, J. V. (2012). *Productividad y gestión de recursos humanos en las Administraciones Públicas*. Madrid, España: Funcas.
- Ployhart, R. E. y Ryan, A. M. (1998). Applicants' reactions to the fairness of selection procedures: The effects of positive rule violations and time of measurement. *Journal of Applied Psychology*, 83, 3-16. <https://doi.org/10.1037/0021-9010.83.1.3>
- Puertas, E. y Embid, Ó. (2011). Las novedades en la selección de personal como palanca de cambio para las Administraciones Públicas. *Revista Aragonesa de Administración Pública*, 13, 171-195.
- Quiñones, M. A., Ford, J. K. y Teachout, M. S. (1995). The relationship between work experience and job performance: A conceptual and meta-analytic review. *Personnel Psychology*, 48, 887-910. <https://doi.org/10.1111/j.1744-6570.1995.tb01785.x>
- Ramió, C. (2011). Estrategias para la modernización del acceso a la función pública. *Nuevas Políticas Públicas: Anuario multidisciplinar para la modernización de las Administraciones Públicas*, 7, 129-144.
- Rodríguez, A. (2016). Validez predictiva e impacto adverso de la entrevista conductual estructurada en el sector público. *Revista de Psicología del Trabajo y de las Organizaciones*, 32, 75-85. <https://doi.org/10.1016/j.rpto.2016.04.003>
- Rodríguez Garraza, T. (2008). *La selección por puestos como alternativa a la selección por cuerpos*. Ponencia presentada en el XIII Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública. Buenos Aires, Argentina.
- Ryan, A. M., Greguras, G. J. y Ployhart, R. E. (1996). Perceived job relatedness of physical ability testing for firefighters: Exploring variations in reactions. *Human Performance*, 9, 219-240. [https://doi.org/10.1207/s15327043hup0903\\_3](https://doi.org/10.1207/s15327043hup0903_3)
- Ryan, A. M. y Ployhart, R. E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, 26, 565-606. <https://doi.org/10.1177/014920630002600308>
- Ryan, A. M., Sacco, J. M., McFarland, L. A. y Kriska D. S. (2000). Applicant self-selection: Correlates of withdrawal from a multiple hurdle process. *Journal of Applied Psychology*, 85, 163-179. <https://doi.org/10.1037/0021-9010.85.2.163>
- Sackett, P. R. (2014). When and why correcting validity coefficients for interrater reliability makes sense. *Industrial and Organizational Psychology*, 7, 501-506. <https://doi.org/10.1111/iops.12185>
- Sáez, J. (2007). Diseño y validación de una entrevista conductual estructurada para la selección de agentes de policía local. *Revista de Psicología del Trabajo y de las Organizaciones*, 23, 151-178.
- Salgado, J. F. (1997). VALCOR: A program for estimating standard error, confidence intervals, and probability of corrected validity. *Behavior Research Methods*, 29, 464-467.
- Salgado, J. F. (2010). *Selección de recursos humanos por competencias en las Administraciones públicas: evidencias basadas en la investigación*. Barcelona, España: CEMICAL.
- Salgado, J. F. (2015) Estimating coefficients of equivalence and stability for job performance ratings: The importance of controlling for transient error on criterion measurement. *International Journal of Selection and Assessment*, 23, 37-44. <https://doi.org/10.1111/ijsa.12093>
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., De Fruyt, F. y Rolland, J. P. (2003). GMA measures and the prediction of Job Performance and training success for different occupations in the European Community. *Journal of Applied Psychology*, 88, 1068-1081. <https://doi.org/10.1037/0021-9010.88.6.1068>
- Salgado, J. F., Gorriti M. y Moscoso, S. (2007). La entrevista estructurada y el desempeño laboral en la administración pública española: propiedades psicométricas y reacciones de justicia. *Revista de Psicología del Trabajo y de las Organizaciones*, 23, 151-178.
- Salgado, J. F., Moscoso, S. y Anderson, N. (2016). Corrections for criterion reliability in validity generalization: The consistency of Hermes, the utility of Midas. *Revista de Psicología de Trabajo y de las Organizaciones*, 32, 17-23. <https://doi.org/10.1016/j.rpto.2015.12.002>
- Salgado, J. F. y Táuriz, G. (2014). The five-factor model, forced-choice personality inventories and performance: A comprehensive meta-analysis of academic and occupational validity studies. *European Journal of Work and Organizational Psychology*, 23, 3-30. <https://doi.org/10.1080/1359432x.2012.716198>
- Salgado, J. F., Viswesvaran, C. y Ones, D. S. (2001). Predictors used for personnel selection: An overview of constructs. En N. R. Anderson, D. S. Ones, H. K. Sinangil y C. Viswesvaran (Eds.), *Handbook of industrial, word and organizational psychology* (pp.165-199). London, UK: Sage.
- Schmidt, F. L. (1994). The future of personnel selection in the US. Army. En M. G. Rumsey, C. B. Walker y J. H. Harris (Eds.), *Personnel selection and classification* (pp. 333-350). Hillsdale, NJ: Erlbaum.
- Schmidt, F. L. y Hunter, J. E. (1993). Tacit knowledge, practical intelligence, general mental ability, and job knowledge. *Current Directions in Psychological Science*, 2, 8-9. <https://doi.org/10.1111/1467-8721.ep10770456>
- Schmidt, F. L., Le, H. y Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual-differences constructs. *Psychological methods*, 8, 206-224. <https://doi.org/10.1037/1082-989x.8.2.206>
- Schmidt, F. L., Oh, I. S. y Shaffer, J. A. (2016). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 100 Years of Research Findings. *Fox School of Business Research Paper*. Retrieved from <https://ssrn.com/abstract=2853669>
- Schneider, R. T. (1994). *The rating of experience and training: A review of the literature and recommendations on the use of alternative E & T procedures*. Alexandria, VA: International Personnel Management Association.
- Shen, W., Cucina, J. M., Walmsley, P. T. y Seltzer, B. K. (2014). When correcting for unreliability of job performance ratings, the best estimate is still. *52. Industrial and Organizational Psychology*, 7, 519-524. <https://doi.org/10.1111/iops.12187>
- Smith, M. y Smith, P. (2005). *Testing people at work*. Malden, MA.: Wiley-Blackwell.

- Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K. y Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel psychology*, 46, 49-76. <https://doi.org/10.1111/j.1744-6570.1993.tb00867.x>
- Sturman, M. C. (2003). Searching for the inverted u-shaped relationship between time and performance: Meta-analyses of the experience/performance, tenure/performance, and age/performance relationships. *Journal of Management*, 29, 290-316. [https://doi.org/10.1016/S0149-2063\\_03\\_00028-X](https://doi.org/10.1016/S0149-2063_03_00028-X)
- Táuriz, G. (2014). *Selección de empleados públicos y oposiciones: ¿qué nos dice la evidencia?* Recuperado de <http://politikon.es/2014/11/19/seleccion-de-empleados-publicos-y-oposiciones-que-nos-dice-la-evidencia/>
- Thorndike, R. L. (1949). *Personnel selection*. New York, NY: Wiley.
- Truxillo, D. M., Steiner, D. D. y Gilliland, S. W. (2004). The importance of organizational justice in personnel selection: Defining when selection fairness really matters. *International Journal of Selection and Assessment*, 12, 39-53. <https://doi.org/10.1111/j.0965-075X.2004.00262.x>
- Van-Biesen, G. (2006, marzo). *Overview of civil service selection procedures in EU context*. Seminar on Civil Service Recruitment Procedures. Vilnius, Lituania. Recuperado de <http://www.oecd.org/dataoecd/52/39/36762418.pdf>
- Viswesvaran, C. y Ones, D. S. (2001). Integrity test and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment*, 9, 31-39. <https://doi.org/10.1111/1468-2389.00161>
- Viswesvaran, C., Ones, D. S. y Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 81, 557-574. <https://doi.org/10.1037/0021-9010.81.5.557>
- Viswesvaran, C., Ones, D. S., Schmidt, F. L., Le, H. y Oh, I. (2014). Measurement error obfuscates scientific knowledge: Path to cumulative knowledge requires corrections for unreliability and psychometric meta-analyses. *Industrial and Organizational Psychology*, 7, 507-518. <https://doi.org/10.1111/iops.12186>
- Wiesen, J. P., Abrams, N. y McAttee, S. A. (1994). Employment testing: a public sector viewpoint. In *Policy Issues in Employment Testing* (pp. 77-130). Dordrecht, The Netherlands: Springer.
- Zibarras, L. D. y Patterson, F. (2015). The Role of Job Relatedness and Self-efficacy in Applicant Perceptions of Fairness in a High-stakes Selection Setting. *International Journal of Selection and Assessment*, 23, 332-344. <https://doi.org/10.1111/ijsa.12118>

## Apéndice

Ejemplo de pregunta de la prueba de conocimientos

27. ¿QUÉ DICE LA LEGISLACIÓN DE PROCEDIMIENTO ADMINISTRATIVO SOBRE LOS ACTOS ADMINISTRATIVOS QUE SE SEPARAN DEL CRITERIO SEGUIDO EN ACTUACIONES PRECEDENTES?

- a) Que no se pueden dictar.
- b) Que, de dictarse, se reputarán inválidos.
- c) Que, de dictarse, se reputarán ineficaces.
- d) Que requieren motivación.

Ejemplos de preguntas de la prueba TJS

CASO Nº 2. UNA PERSONA VIENE AL REGISTRO GENERAL CENTRAL DE LA ADMINISTRACIÓN DE LA COMUNIDAD AUTÓNOMA DE EUSKADI EN LAKUA CON EL ORIGINAL DE SU TÍTULO DE BACHILLER EXPEDIDO POR EL DEPARTAMENTO DE EDUCACIÓN DEL GOBIERNO VASCO. SOLICITA QUE SE LE HAGA UNA COPIA COMPULSADA PARA ENVIAR AL AYUNTAMIENTO DE SANTANDER A FIN DE PARTICIPAR EN UN PROCESO DE SELECCIÓN DE PERSONAL.

168. ¿CUÁL ES LA TAREA A REALIZAR PARA LA COMPULSA?

- a) Reproducir literalmente el original.
- b) Cotejar minuciosamente la copia del documento con su original.
- c) Certificar el contenido del documento original.
- d) Poner un sello de registro de entrada.

CASO Nº 6.- Begoña es una administrativa que trabaja en una sección del Servicio de Bibliotecas de una Administración Pública. Para realizar una de las tareas que tenía encomendada, debía alcanzar unos expedientes archivados a una altura a la que llegaba con dificultad. Asier, uno de sus compañeros de trabajo seguía los esfuerzos que estaba haciendo Begoña para alcanzar los expedientes a través de una puerta acristalada de grandes dimensiones. Un movimiento descontrolado de Begoña hizo que todos los expedientes cayesen sobre la cara de ésta, produciéndole un corte en la nariz. Asier, al percatarse del accidente, se dirigió a donde estaba Begoña y empujó la puerta de cristal con tanta fuerza, que aquella se rompió en pedazos. Asier sufrió un gran corte en el brazo. El resto del personal del Servicio acudió a auxiliarles, pero se asustaron y no sabía cómo actuar: Begoña se quejaba mucho y Asier sangraba abundantemente, aunque aparentaba tranquilidad. Inicialmente todos atendieron a Begoña y, como sangraba por la nariz, le pusieron un trapo mojado en la nuca y le aconsejaron que pusiera la cabeza hacia atrás. Al mismo tiempo se percataron de que Asier se iba poniendo pálido y presentaba indicios de que iba a perder el conocimiento. En ese momento, un compañero aplicó un trapo sobre la herida con la intención de parar la hemorragia. Cada poco tiempo y al objeto de conocer si la hemorragia había cedido retiraba el trapo. El compañero al darse cuenta que la hemorragia no paraba le practicó un torniquete. Esto pareció solucionar el problema. Posteriormente alguien pidió ayuda a los servicios de urgencias. Antes de la llegada de los servicios sanitarios, Asier se quejaba de fuertes dolores en el brazo herido.

183. EN EL CASO QUE NOS OCUPA, ¿CUÁL SERÍA LA SECUENCIA CORRECTA DE ACTUACIONES PARA DETENER LA HEMORRAGIA DE ASIER?:

- a) Compresión arterial – torniquete – compresión directa.
- b) Torniquete - compresión directa – compresión arterial.
- c) Compresión directa – compresión arterial – torniquete.
- d) Sólo compresión arterial.