



Validez predictiva e impacto adverso de la entrevista conductual estructurada en el sector público



Alexander Rodríguez*

Instituto Vasco de Administración Pública, España

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 4 de noviembre de 2015

Aceptado el 7 de abril de 2016

On-line el 9 de junio de 2016

Palabras clave:

Entrevista

Evaluación del desempeño

Selección

Validez predictiva

Diferencias de género

Sector público

Keywords:

Interview

Performance appraisal

Selection

Predictive validity

Gender differences

Public sector

R E S U M E N

En este trabajo se estudia el papel de la entrevista conductual estructurada (ECE) en los procedimientos de selección del sector público, cuya legislación insta a comprobar si los procedimientos selectivos están vinculados con el desempeño y si se cumple el principio de igualdad de oportunidades entre sexos. Para ello se ha analizado en una muestra de 379 candidatos si existían diferencias de género en la puntuación de la entrevista. Por otro lado, se midió el desempeño de las 125 personas contratadas, lo que permitió estudiar la validez predictiva de la ECE. Los resultados son similares a los obtenidos por investigaciones precedentes –la ECE predice de manera significativa el desempeño y no produce impacto de género. Finalmente se discuten las implicaciones de estos resultados.

© 2016 Colegio Oficial de Psicólogos de Madrid. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la CC BY-NC-ND licencia (<http://creativecommons.org/licencias/by-nc-nd/4.0/>).

Predictive validity and adverse impact of the structured behavioral interview in the public sector

A B S T R A C T

This paper studies the role of the structured behavioral interview (SBI) in the selection procedures of the public sector. The Spanish legislation calls on to examine whether selection procedures are linked to performance and whether the principle of equal opportunities regarding gender is respected. Hence, it was analyzed if there were gender differences in the interview scores in a sample of 379 candidates. Furthermore, performance of 125 hired subjects was assessed, allowing us to measure the predictive validity of the SBI. Results are similar to those obtained by previous research –SBIs are a significant predictor of performance and they do not produce adverse gender impact. Finally, implications of these results are discussed.

© 2016 Colegio Oficial de Psicólogos de Madrid. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

La selección de empleados públicos

Tradicionalmente, los métodos de selección utilizados en la mayor parte de los procesos selectivos en la administración pública española han sido las pruebas de conocimientos, la valoración de la experiencia y, en menor medida, las entrevistas (Salgado, 2010).

Según la Ley 7/2007, del Estatuto Básico del Empleado Público (Congreso de Los Diputados, 2007a), que regula la selección en las administraciones públicas españolas, las entrevistas tienen cabida en las oposiciones y deben servir “para asegurar la objetividad y la racionalidad de los procesos selectivos”. El Tribunal Supremo español ha defendido la legalidad de la entrevista y su compatibilidad con los principios de mérito y capacidad, pero también ha dictaminado que debe tener un peso proporcional inferior al de las pruebas de valoración de conocimientos (Chaves, 2009; Marín, 2009; Mauri, 2009). Como señala Chaves (2009), la entrevista se ha introducido con muchas cautelas en el sector público, algo que se puede vincular a su incorrecto uso (Marín, 2009). En ese

* Autor para correspondencia. Instituto Vasco de Administración Pública. C/Donostia-San Sebastián, 1. 01010 Vitoria-Gasteiz.
Correo electrónico: alexander-rodriguez@ivap.es

sentido, Mauri (2009) aclara que no es admisible cualquier tipo de entrevista de selección, y afirma que es preciso diferenciar entre entrevistas convencionales no estructuradas, cuyas características psicométricas no son adecuadas y por tanto no deberían tener lugar en las oposiciones, y las entrevistas conductuales estructuradas, que emplean preguntas elaboradas previamente que se hacen a todos los candidatos, escalas de valoración para cada pregunta y dimensión objeto de valoración y entrevistadores formados adecuadamente. Este autor, en su análisis jurisprudencial, señala que las entrevistas, al igual que el resto de pruebas a emplear en las oposiciones, han de ser objetivas y relacionadas con las funciones a desempeñar y que su utilización en la selección de funcionarios se justifica en tanto puedan servir para elegir al mejor candidato para un puesto de trabajo determinado, principal objetivo del proceso de selección.

La Ley 7/2007 trata de los procesos de selección en el acceso al empleo público en su artículo 61, que insta a que dichos procedimientos cuiden dos aspectos. Por un lado, deben cuidar “especialmente la conexión entre el tipo de pruebas a superar y la adecuación al desempeño de las tareas de los puestos de trabajo convocados”. Sin embargo, como señalan Gorriti y López (2014), son pocas las administraciones públicas españolas que hayan comprobado las predicciones hechas en sus procedimientos selectivos empleando el desempeño como criterio. A modo de ejemplo, es preciso citar las experiencias del Ayuntamiento de Vitoria-Gasteiz (Sáez, 2007), el Ayuntamiento de Zaragoza (Berges, 2011) o el Gobierno Vasco (Salgado, Gorriti y Moscoso, 2007). Estas administraciones han podido demostrar el cumplimiento del mencionado precepto y, por tanto, la eficacia de sus pruebas selectivas, evidenciando su conexión con el desempeño de los trabajadores, es decir, la relación entre el contenido de las pruebas y el comportamiento laboral concreto que predicen. Esto está relacionado con el concepto *validez de criterio*, que se define como “el grado en que las puntuaciones de una prueba están relacionadas con una medida de criterio” (AERA, APA y NCME, 1999). Como señalan Aguinis, Henle y Ostroff (2001), la validez de criterio implica correlación de puntuaciones del predictor (como las pruebas empleadas en la selección) con un criterio (otra medida, por ejemplo de desempeño laboral) para determinar si se pueden tomar decisiones precisas a partir de las puntuaciones en la prueba empleada, y cuando el objetivo es la predicción de una conducta futura se habla de *validez predictiva*. En este caso se trata de estudiar la relación entre dos medidas relativas a los mismos sujetos tomadas en dos momentos distintos: primero, los resultados en el proceso selectivo de las personas contratadas –los predictores– y posteriormente las evaluaciones del desempeño de esas mismas personas en los puestos para los que han sido contratadas –los criterios–. Se trata por tanto, de estudiar en qué medida las puntuaciones de una prueba selectiva predicen el desempeño futuro de los candidatos a un puesto.

Por otro lado, el mencionado artículo 61 afirma que “los órganos de selección velarán por el cumplimiento del principio de igualdad de oportunidades entre sexos”. Como indicaron Alfama y Alonso (2015), la legislación ha previsto tomar medidas para garantizar el acceso de las mujeres al empleo público en condiciones de paridad, por lo que poder comprobar si los métodos de selección producen resultado o impacto adverso en función del género de los candidatos es una cuestión relevante tanto desde un punto de vista aplicado como desde un punto de vista jurídico (Alonso, 2011).

Alonso (2011) señala que el resultado o impacto adverso “se produce cuando la aplicación de un mismo criterio, método o procedimiento para la selección tiene un efecto negativo sobre uno o más grupos o minorías”. Este concepto está vinculado con la discriminación indirecta por razón de sexo, definida por la legislación española en la Ley 3/2007, para la igualdad efectiva de mujeres y hombres (Congreso de Los Diputados, 2007b), como “la situación en que una disposición, criterio o práctica aparentemente neutros

pone a personas de un sexo en desventaja particular con respecto a personas del otro, salvo que dicha disposición, criterio o práctica puedan justificarse objetivamente en atención a una finalidad legítima y que los medios para alcanzar dicha finalidad sean necesarios y adecuados”. Según Alonso (2011), si se demostrase que las entrevistas de selección producen impacto adverso en contra de las mujeres se podría propiciar la aparición de demandas por discriminación indirecta, al entenderse que vulnerarían los principios de igualdad entre mujeres y hombres establecidos por la Constitución Española y regulados por la Ley 3/2007.

La entrevista

La entrevista de empleo, definida por Dipboye (1992) como un “diálogo entre una o más personas para conseguir información y evaluar las cualidades de un candidato a un puesto”, tiene una destacada presencia en los procesos de selección de empresas y organizaciones de todo el mundo (Alonso, Moscoso y Cuadrado, 2015; Salgado, Viswesvaran y Ones, 2001). Sin embargo, aunque se empleen los términos entrevista de empleo o entrevista de selección de modo genérico, existen diferentes tipos de entrevista, que varían en su frecuencia de uso y sus propiedades psicométricas (Salgado y Moscoso, 2005; Salgado, Moscoso y Gorriti, 2004).

Las entrevistas se pueden clasificar en función de su contenido y de su grado de estructura (Salgado y Moscoso, 2005; Salgado et al., 2004). Con respecto al contenido, es preciso distinguir entre entrevistas convencionales y entrevistas conductuales. Las entrevistas convencionales se basan en preguntas sobre hechos biográficos, credenciales y logros, así como descripciones sobre experiencia e información auto-evaluativa, como virtudes y defectos, agrados y desagradados, metas, etc. Aquí cabe diferenciar entre entrevistas convencionales no estructuradas o semiestructuradas, que duran pocos minutos y donde las preguntas no son las mismas para todos los candidatos, y entrevistas convencionales estructuradas, de mayor duración y que utilizan un repertorio de preguntas similar para todos los entrevistados. Las entrevistas conductuales, por su parte, se centran en preguntas sobre experiencias en el trabajo y, fundamentalmente, en descripciones de conducta. Las entrevistas conductuales más usadas son las entrevistas situacionales, en las que se pregunta a los candidatos sobre su actuación en situaciones hipotéticas, y las entrevistas conductuales estructuradas (ECE), en las que los candidatos deben describir su comportamiento en determinadas situaciones pasadas (Janz, 1982; Salgado y Moscoso, 2005; Salgado et al., 2004).

Por lo que se refiere al grado de estructura, Huffcutt y Arthur (1994) establecieron criterios para clasificar los distintos tipos de entrevista, distinguiendo 4 niveles en cuanto a la estandarización de las preguntas: 1) ausencia de restricciones formales con respecto a las preguntas a formular, algo propio de las entrevistas convencionales no estructuradas o semiestructuradas, 2) hay un esquema de temas a tratar, pero no preguntas concretas, como en las entrevistas convencionales estructuradas, 3) especificación previa de preguntas, aunque los entrevistadores pueden adaptar la entrevista a los diferentes candidatos, elegir entre cuestiones alternativas y sondear a los solicitantes sobre cuestiones específicas, como en la entrevista de descripción de conducta de Janz (1982) y 4) mismas cuestiones para todos los candidatos, sin posibilidad de reformular las preguntas ni de realizar preguntas para profundizar en las respuestas –repreguntas– como en las entrevistas conductuales estructuradas.

Huffcutt y Arthur (1994) también distinguieron tres niveles progresivamente más estandarizados para la valoración de las respuestas del candidato: (1) una única evaluación global en función de la información obtenida en el conjunto de la entrevista, como sucede en las entrevistas convencionales no estructuradas o

semiestructuradas, (2) valoraciones múltiples en base a las dimensiones o rasgos medidos por la entrevista, como en la entrevista de Janz (1982) o las entrevistas convencionales estructuradas y (3) valoración independiente de cada una de las respuestas dadas por los candidatos a las preguntas en diferentes escalas adaptadas a cada una de esas cuestiones, como en las entrevistas conductuales estructuradas.

Por su parte, Campion, Palmer y Campion (1997), tras estudiar la literatura en materia de estructuración sobre entrevistas, identificaron 15 componentes característicos: (1) basar las preguntas en un análisis del puesto, (2) hacer las mismas preguntas a todos los candidatos, (3) limitar las repreguntas o la reformulación de las preguntas, (4) usar mejores tipos de preguntas –como preguntas relativas a conductas realizadas en el pasado o bien preguntas situacionales–, (5) emplear un mayor número de preguntas –en los estudios analizados por Campion et al. la media de preguntas fue de 16.5 ($SD = 8.71$)– o entrevistas más largas –dos tercios de las entrevistas estudiadas por Campion et al. duraron entre 30 y 60 minutos–, (6) controlar la información complementaria, como resultados de otras pruebas, entrevistas previas, curriculum, etc., (7) no permitir preguntas del entrevistado hasta la finalización de la entrevista, (8) evaluar cada respuesta o emplear escalas múltiples en vez de realizar una evaluación global del conjunto de la entrevista, (9) utilizar escalas de evaluación basadas en anclajes detallados, (10) tomar notas detalladas, (11) utilizar varios entrevistadores, (12) emplear el mismo entrevistador –o los mismos entrevistadores– para todos los candidatos, (13) no hablar de los candidatos o de sus respuestas entre entrevistas, (14) proporcionar formación amplia a los entrevistadores y (15) utilizar predicciones estadísticas que impliquen poca subjetividad o juicio, en vez de predicciones clínicas basadas en la intuición o la percepción.

Según Salgado y Moscoso (2005), estos 15 componentes diferenciables se pueden resumir en 6 características: preguntas desarrolladas a partir de un análisis de puestos, mismas preguntas para todos los candidatos, uso de escalas de valoración basadas en la conducta, mismo proceso con todos los candidatos, formación de entrevistadores y toma de decisión sobre la contratación tras realizar todas las entrevistas.

Estos componentes están presentes en las entrevistas conductuales estructuradas, término en el que Salgado y Moscoso (2002) agruparon entrevistas muy estudiadas en la literatura y centradas en preguntas sobre descripciones de conducta, como la entrevista de descripción de conducta, la entrevista situacional, la entrevista multimodal y la entrevista basada en análisis de puestos (Janz, 1982, 1989; Latham, Saari, Pursell y Campion, 1980; Motowidlo et al., 1992; Schuler y Funke, 1989). Las ECE incorporan contenidos relativos a la búsqueda de conductas relevantes para el desempeño laboral, por lo que el proceso de decisión está vinculado a factores relacionados con el puesto de trabajo, y además contribuyen a reforzar la objetividad demandada por la legislación haciendo hincapié en la estandarización en la aplicación del instrumento (mismas instrucciones y condiciones para todos los participantes) y en la evaluación de los resultados, dando como resultado una entrevista con un alto grado de estructuración, lo que a su vez está vinculado con una mayor validez de criterio (Levashina, Hartwell, Morgeson y Campion, 2014; McDaniel, Whetzel, Schmidt y Maurer, 1994; Sáez, 2007; Wiesner y Cronshaw, 1988). Las ECE son uno de los mejores predictores del desempeño laboral y presentan coeficientes de validez superiores a los de las entrevistas convencionales estructuradas, que muestran coeficientes de validez aceptables, y a los de las entrevistas convencionales no estructuradas, que no cuentan con niveles de validez suficientes (Salgado y Moscoso, 2005; Taylor y Small, 2002). No obstante, el grado de utilización de las ECE es menor que el de las entrevistas convencionales no estructuradas, que son las entrevistas más empleadas en las organizaciones (Alonso et al., 2015).

Una organización pública que ha utilizado frecuentemente la entrevista conductual estructurada es el Gobierno Vasco. En esta administración, la ECE, cuyo uso está respaldado por la Sentencia del Tribunal Supremo (STS) de 28 de marzo de 2011, se ha empleado con resultados alentadores en procesos de promoción interna para la cobertura de puestos de jefatura (Salgado et al., 2007; Salgado et al., 2004). Siguiendo la estela de esta experiencia, el Instituto Vasco de Administración Pública (IVAP) ha utilizado con frecuencia la ECE en oposiciones públicas, con los siguientes rasgos principales: los candidatos fueron entrevistados por un único entrevistador que utilizó preguntas de descripción de conductas laborales realizadas en el pasado, en la línea del modelo de Janz (1982), y empleó escalas de valoración con anclajes conductuales (Motowidlo et al., 1992; Smith y Kendall, 1963).

Las ECE se basan en el principio de que el mejor predictor de la conducta futura es la conducta pasada en similares circunstancias. En cada pregunta, primero se plantea la situación sobre la que se quiere preguntar para localizar un ejemplo concreto en el pasado del candidato y que éste se centre en él. Una particularidad de la entrevista de descripción de conducta de Janz (1982) es la posibilidad de reformular las preguntas –componente nº 3 de los identificados por Campion et al. (1997)– adaptando la entrevista para candidatos sin experiencia.

Las preguntas de las ECE miden las dimensiones conductuales que debe poseer el futuro empleado público que son identificadas por el análisis del puesto. Como señalan Salgado y Moscoso (2005), cada puesto tiene su propia constelación de dimensiones específicas que han de ser evaluadas mediante la entrevista, midiendo, según el caso, relaciones con los compañeros, relaciones con usuarios o clientes, orden y minuciosidad en el trabajo, capacidad de persuasión e influencia, motivación laboral, habilidades sociales, capacidad analítica, capacidad de decisión, adaptabilidad, habilidades de comunicación, creatividad, etc.

Las escalas de valoración con anclajes conductuales –en inglés, *Behaviorally Anchored Rating Scales* o BARS– sirven para valorar las respuestas (Smith y Kendall, 1963). Las BARS permiten fijar un marco de referencia común para todos los evaluadores y se basan en conductas que indican diferentes niveles de ajuste a una dimensión determinada, lo que posibilita estandarizar qué es hacer bien o mal las tareas de un puesto (Gorriti y López, 2009).

Investigación sobre la validez predictiva, la fiabilidad y el impacto de género de las entrevistas

La validez predictiva de las entrevistas conductuales estructuradas ha recibido una considerable atención en diferentes publicaciones en todo el mundo. En España cabe citar los trabajos realizados en administraciones públicas por parte de Berges (2011), Sáez (2007) y Salgado et al. (2007), en los que se evaluó el grado en que las entrevistas pronosticaban el desempeño laboral en puestos de policía en los dos primeros casos y puestos de mando en el último.

El procedimiento de evaluación del desempeño llevado a cabo en dichos estudios estuvo guiado por las recomendaciones de la investigación en materia de desempeño: evaluación de las conductas bajo control del trabajador (Campbell, 1990; Viswesvaran, 2001), medición referida a un periodo de tiempo concreto (Motowidlo y Kell, 2013), evaluación realizada por superiores, ya que su fiabilidad interjueces (.52) es superior a la de los compañeros (.42), según Viswesvaran, Ones y Schmidt (1996), y empleo de escalas de valoración con anclajes conductuales para evaluar las respuestas de los candidatos, teniendo en cuenta que las BARS se han mostrado como un moderador de la validez de las ECE (Taylor y Small, 2002).

Por lo que se refiere a la naturaleza del desempeño laboral, Viswesvaran y Ones (2000) o Rotundo y Sackett (2002) concluyeron

que existen tres dimensiones de desempeño para cualquier organización y para cualquier trabajo: el desempeño de tarea, definido por Borman y Motowidlo (1993) como “la pericia con la que los ocupantes de un puesto realizan actividades formalmente reconocidas como parte de sus trabajos, actividades que contribuyen al núcleo técnico de la organización, bien directamente como parte de su proceso tecnológico o indirectamente proveyéndolo con materiales necesarios o servicios”, el desempeño contextual, definido por Borman y Motowidlo (1993) como “conductas que mantienen el ambiente organizacional, social y psicológico en el que el núcleo técnico tiene que funcionar” y que incluyen actividades como la ayuda a otros en el trabajo, el desempeño voluntario de tareas o asignaciones extra o el apoyo a la organización y las conductas contraproducentes, definidas por Sackett y DeVore (2001) como “cualquier conducta intencional por parte del empleado y que es percibida por la organización como contraria a sus legítimos intereses” y que incluyen comportamientos como absentismo, robos, abusos de alcohol y drogas, etc.

En los trabajos de Berges (2011), Sáez (2007) y Salgado et al. (2007) la ECE se mostró como un predictor válido del desempeño laboral, lo que coincide con los hallazgos de investigaciones realizadas a nivel mundial. Como ejemplo cabe citar los metaanálisis de Taylor y Small (2002) y Huffcutt, Conway, Roth y Klehe (2004), que encontraron que las entrevistas conductuales estructuradas basadas en conductas previas tenían un coeficiente de validez sin corregir de .31. Sin embargo, como señalaron Hunter y Schmidt (2004), los valores de las correlaciones corregidas ofrecen valores más próximos a la validez verdadera. Así, diferentes metaanálisis realizados en torno a las entrevistas estructuradas (Huffcutt y Arthur, 1994; Huffcutt et al., 2004; McDaniel et al., 1994; Salgado y Moscoso, 1995; Taylor y Small, 2002; Wiesner y Cronshaw, 1988) han encontrado coeficientes de validez corregidos por fiabilidad del criterio y restricción del rango en el predictor con valores entre .44 y .63.

Salgado y Moscoso (2005) señalaron que las revisiones metaanalíticas coinciden al señalar que las ECE presentan coeficientes de validez equiparables a los de los instrumentos de selección con una validez más alta, que el formato de la entrevista (individual, con un único entrevistador, o de panel, con más de un entrevistador para cada entrevistado) parece que apenas tiene efecto diferencial sobre las ECE y que los coeficientes de validez son significativos para diferentes puestos.

En cuanto a la fiabilidad de las ECE, de los tres métodos existentes para estimar la fiabilidad de las entrevistas, test-retest, consistencia interna y fiabilidad interevaluadores, Dipboye (1992) señala que el primero no es adecuado por producir estimaciones infladas de la fiabilidad. La consistencia interna, medida mediante el coeficiente alfa de Cronbach, puede ser útil para saber si añadir ítems puede suponer diferencias en la ordenación de las puntuaciones totales de los candidatos y para establecer el número de preguntas a realizar (Conway, Jako y Goodman, 1995) y la fiabilidad interevaluador o interjueces se refiere al grado de acuerdo entre dos o más entrevistadores cuando evalúan independientemente al mismo individuo (Salgado y Moscoso, 2005).

Con respecto a este último método para estimar la fiabilidad, tanto el metaanálisis de Conway et al. (1995) como el de Salgado et al. (2004) encontraron una fiabilidad interjueces de .92 para entrevistas en formato panel, mientras que para entrevistas separadas el valor máximo obtenido por los primeros fue de .66 en entrevistas con alto grado de estructuración. Por lo que respecta a la consistencia interna de las entrevistas, el metaanálisis de Conway et al. (1995) indicó que en entrevistas con repreguntas la media de los coeficientes alfa de Cronbach para entrevistas de entre 6 y 10 preguntas y de entre 11 y 20 fue respectivamente de .81 y .82.

Por lo que se refiere a la investigación en materia de impacto de género de las entrevistas, el metaanálisis llevado a cabo por Alonso

(2011) ha mostrado que las ECE no producen resultados adversos para las mujeres. El tamaño del efecto del sexo sobre las puntuaciones en las entrevistas (d de Cohen) fue de -0.18 para las ECE, lo que implica puntuaciones más bajas para el grupo de los hombres que para el grupo de mujeres. Tanto este valor como el obtenido por el metaanálisis de Levashina et al. (2014) para entrevistas estructuradas de diferentes tipos, .09, son pequeños en términos de Cohen (1988), por lo que cabe concluir que las diferencias entre mujeres y hombres son mínimas o inexistentes en cuanto a su puntuación en la entrevista.

La muestra del metaanálisis de Alonso (2011) incluía datos de los trabajos realizados en administraciones públicas españolas (Berges, 2011; Sáez, 2007; Salgado et al., 2007), aunque se basaba fundamentalmente en dos estudios (McCarthy, Van Iddekinge y Campion, 2010; Sacco, Scheu, Ryan y Schmitt, 2003), por lo que la autora recomendaba la realización de más estudios sobre este tema en diferentes países, tipos de organizaciones y para puestos de diferentes categorías.

Aunque hay resultados metaanalíticos tanto en materia de validez predictiva como de impacto adverso de género hacia las mujeres de las ECE, los trabajos de validación de instrumentos predictivos siguen siendo escasos en España, especialmente en el sector público. Hasta ahora, las publicaciones referidas a resultados de investigación sobre validez predictiva de la entrevista conductual estructurada se han realizado en el acceso o provisión en policía local y en la provisión de puestos de mando en la administración general, por lo que el análisis de validez predictiva y de impacto de género en oposiciones para el ingreso en otras categorías o puestos no ha sido suficientemente estudiado. Además, se antoja necesario contribuir a romper los prejuicios acerca de la entrevista y su aplicación en las oposiciones, comprobando que cumple las exigencias demandadas por la legislación.

A tal fin se ha diseñado la presente investigación, cuyo objetivo es estudiar en qué medida las entrevistas conductuales estructuradas que se utilizan en el IVAP son válidas a la hora de predecir el desempeño de los candidatos a puestos de empleados públicos y no producen impacto adverso de género. Las hipótesis de este trabajo son:

H1. La ECE predecirá de manera significativa el desempeño laboral.

H2. No habrá diferencias significativas entre hombres y mujeres en cuanto a su puntuación en la ECE.

Método

Participantes

La muestra estuvo formada por 379 personas –195 varones y 184 mujeres– que realizaron una entrevista en 45 oposiciones a entidades públicas vascas (ayuntamientos y empresas públicas). De estas 379 personas, 125 –90 varones y 35 mujeres– lograron plaza de funcionario tras realizar la entrevista en los mencionados 45 procesos selectivos. Se trata, por tanto, de una muestra incidental.

De las personas contratadas, 6 pertenecían al grupo de titulación A (técnicos superiores y arquitectos), 15 al grupo B (técnicos medios y arquitectos técnicos), 29 al grupo C (administrativos y técnicos auxiliares), 68 al D (auxiliares administrativos, bomberos y personal de oficios) y 7 al E (asistentes domiciliarios, personal de oficios y conserjes).

Medidas

Entrevista. Se empleó una entrevista conductual estructurada (Salgado y Moscoso, 2005) para valorar la adecuación de las personas aspirantes al perfil de la plaza y a las funciones a desempeñar.

De los 15 criterios establecidos por [Campion et al. \(1997\)](#), las entrevistas analizadas en este trabajo cumplieron 11 y no fue posible cumplir cuatro criterios: no se pudo controlar el acceso a información complementaria, cada candidato solo pudo ser evaluado por un único entrevistador, no se pudo emplear al mismo entrevistador con todos los candidatos en oposiciones cuyas entrevistas se realizaron a lo largo de varios días y, en estos casos, no se instruyó a los entrevistadores para que evitaran discutir sobre las respuestas de los aspirantes entre entrevistas.

En todos los casos se empleó un análisis del puesto en el que, en base a la información aportada por expertos en el mismo, se definieron las dimensiones relevantes del puesto y las preguntas que se derivarían de esas dimensiones, centradas en las conductas previas de los candidatos que se consideraron eficaces para predecir el futuro rendimiento del candidato en el puesto. Esta vinculación a las características de cada puesto hizo que el número y el tipo de dimensiones sobre las que se preguntó fueran diferentes en cada oposición, siempre en función del análisis realizado previamente.

Se definió previamente tanto la manera en que se podían reformular las preguntas como las repreguntas que se podían hacer, destinadas a aclarar la información expresada por el candidato. Así, el guión de las entrevistas del IVAP permitió reformular mínimamente las preguntas para candidatos sin experiencia laboral previa, en el sentido de darles la posibilidad de que extrapolasen sus respuestas a situaciones similares a las laborales y que requiriesen conductas semejantes –en el entorno académico o de trabajo asociativo– y que se pudieran evaluar con los mismos criterios. Esto situaría a las entrevistas del presente trabajo en el nivel 3 de estructuración de [Huffcutt y Arthur \(1994\)](#), si bien las escalas de valoración de las respuestas fueron las mismas para todos los candidatos, lo que es un aspecto del nivel más elevado en la clasificación de estos autores.

Las entrevistas tuvieron una duración que osciló entre los 30 minutos y la hora y media, en función de las dimensiones a medir en cada caso, y el número de preguntas osciló entre 7 y 18 ($M = 11.62$, $SD = 2.33$).

Se emplearon escalas de valoración con anclajes conductuales para valorar las respuestas de los candidatos. En todas las entrevistas se utilizó una escala de 5 puntos, de 0, mal ajuste al perfil del puesto, a 4, excelente ajuste, para cada dimensión o pregunta, y la puntuación final se calculó a partir de la puntuación media en las dimensiones evaluadas. En cada dimensión se definieron, con ayuda del análisis de puestos, conductas observables ejemplo de conductas negativas –punto 0 de la escala–, aceptables –punto 2– o muy buenas –punto 4– que sirvieron de referencia para otorgar las puntuaciones.

Se evaluaron un total de 38 dimensiones distintas, siendo las 5 dimensiones más evaluadas trabajo en equipo, responsabilidad, resistencia al estrés y autocontrol, amabilidad y comunicación. Por grupos de titulación las dimensiones más evaluadas fueron, para el grupo A, resistencia al estrés y autocontrol, responsabilidad, innovación y creatividad, orden y comunicación. En el grupo B, orden, trabajo en equipo, responsabilidad, comunicación y seguridad en sí mismo y asertividad. En el grupo C, responsabilidad, meticulosidad, trabajo en equipo, comunicación y flexibilidad para asumir diferentes tareas. En el grupo D, amabilidad, resistencia al estrés y autocontrol, trabajo en equipo, comunicación y responsabilidad. En el grupo E, trabajo en equipo, responsabilidad, disponibilidad, meticulosidad y capacidad de mejora.

La puntuación media establecida para las entrevistas fue de 4.3 puntos, mientras que la puntuación máxima media que se podía obtener en el conjunto de pruebas de las oposiciones según las bases de convocatoria fue de 48.76 puntos, por lo que la entrevista tuvo un peso medio del 8.83% del total de la oposición. En todos los casos la entrevista fue “no eliminatoria” y fue el último ejercicio realizado por los opositores. Pese a que se recibieron 16 reclamaciones en vía

administrativa acerca de la puntuación en la entrevista, en ningún caso fueron objeto de recursos judiciales.

Los 4 entrevistadores que realizaron las entrevistas que forman parte de este trabajo superaron la formación en entrevistas conductuales estructuradas diseñada por el IVAP y tenían experiencia de al menos dos años haciendo entrevistas. Dicha formación incluyó role-playing, entrevistas de práctica, elaboración de guión de preguntas y anclajes conductuales y feedback por parte del profesor y compañeros de clase.

Medidas del criterio: evaluación del desempeño. Para la evaluación del desempeño laboral, se diseñó una herramienta con 6 escalas de valoración con anclajes conductuales que midieron la competencia de los individuos en 6 características del desempeño en base al instrumento empleado por [Salgado et al. \(2007\)](#). Dicho instrumento contaba con 7 dimensiones generales de desempeño de tarea (solución de problemas, toma de decisiones, dirección/supervisión, conocimiento del puesto, organización, resistencia a la presión y productividad) y tres dimensiones generales de desempeño contextual (competencia interpersonal, iniciativa y desempeño extra-rol).

Para el presente estudio únicamente se tuvieron en cuenta las dimensiones consideradas como dimensiones “puras” de tarea ([Gorriti y López, 2010](#); [Ramírez, 2013](#)). Las tres dimensiones empleadas, resolución de problemas, organización y productividad, son generalizables a todos los puestos estudiados y están vinculadas a dimensiones propias de desempeño de tarea mencionadas por diferentes autores, como calidad, eficacia y cantidad. Del mismo modo, se emplearon las tres dimensiones de desempeño contextual, competencia interpersonal, iniciativa y desempeño extra-rol, por entenderlas compatibles con la taxonomía de [Borman y Motowidlo \(1993\)](#) y por su aplicabilidad a cualquier puesto.

La escala de evaluación de cada dimensión contó con cinco niveles (de 0 a 4), siendo 0 el nivel de desempeño más negativo o inadecuado y 4 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).

Se facilitó a los evaluadores un manual con información precisa acerca de los propósitos de la investigación y su vinculación con los imperativos legales, los criterios de valoración del desempeño a emplear en este estudio junto con un procedimiento para puntuarlo, los sesgos a considerar, las definiciones conceptuales de cada dimensión, los anclajes conductuales a emplear, además de una breve exposición sobre el marco teórico de la evaluación del desempeño y su relación con la selección en base a las indicaciones de [Gorriti y López \(2009\)](#).

Procedimiento

Se aplicó una ECE a todos los candidatos que realizaron una entrevista en 45 oposiciones celebradas a partir de 2006. Sólo se incluyeron en el estudio aquellas personas contratadas que seguían en el mismo puesto para el que opositaron.

Las entrevistas se realizaron de manera individual, es decir, un entrevistador entrevistó a un candidato en todos los casos. Todos los sujetos fueron entrevistados por un entrevistador homologado del IVAP, que fue nombrado asesor técnico del tribunal calificador de la oposición y se encargó de analizar el puesto en cuestión y diseñar la entrevista.

Los resultados de este instrumento de selección ya obraban en manos del IVAP –informes de resultados y registros informatizados–, pero para obtener los datos de desempeño fue necesario contar con la colaboración de las administraciones a las cuales el IVAP ha prestado servicio y donde trabajan las personas que habían realizado las entrevistas diseñadas por este Instituto. Para ello se solicitó a las 25 entidades públicas en las que se llevaron

Tabla 1
Estadísticos descriptivos: predictores

Predictor	Opositores contratados		Total de opositores	
	Media N = 125	DT	Media N = 379	DT
Entrevista	3.08	0.41	3.00	0.50

a cabo pruebas con las mencionadas características su colaboración, aceptando finalmente 12. Estas entidades consiguieron la colaboración de 45 superiores –14 mujeres y 31 varones– para evaluar el desempeño de las 125 personas contratadas tras haber aprobado una oposición en la que había que realizar una entrevista. Se evaluó el desempeño a lo largo del último año laboral –evaluando únicamente a aquellas personas con al menos un año de experiencia en el puesto–.

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 metaanalítica para pruebas de desempeño realizadas por superiores (Salgado et al., 2003; Salgado y Táuriz, 2014; Viswesvaran et al., 1996)– 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 379 opositores–, y la desviación típica para el rango restringido –la muestra de 125 personas seleccionadas– (Thorndike, 1949). El valor de U_x en el presente caso fue de 1.22 ($u = 0.82$).

Resultados

Análisis descriptivo de las medidas empleadas

Las tablas 1 y 2 muestran los estadísticos descriptivos (medias y desviaciones típicas) de las variables incluidas en el estudio para la muestra de opositores contratados y para el total de 379 personas que realizaron la entrevista. En la tabla 2 se recogen las medidas de criterio empleadas en esta investigación, que se sitúan por encima del punto medio de la escala. Se aprecia también que las puntuaciones medias son muy semejantes entre sí.

En la tabla 3 se muestra que en las variables que midieron desempeño se produjeron distribuciones asimétricas –asimetría negativa– ya que la mayor parte de las puntuaciones otorgadas se situaron a la derecha de la distribución –los evaluadores han

Tabla 3
Asimetría y curtosis de las variables empleadas

Variable	Asimetría		Curtosis	
	Estadístico	Error típico	Estadístico	Error típico
Entrevista	-0.15	.22	0.32	.43
Desempeño de tarea	-1.09	.22	1.76	.43
Resolución de problemas	-0.73	.22	1.04	.43
Organización del trabajo	-0.66	.22	0.57	.43
Productividad	-0.92	.22	1.06	.43
Desempeño contextual	-1.01	.22	1.32	.43
Competencia interpersonal	-0.77	.22	0.80	.43
Desempeño más allá del puesto	-0.53	.22	-0.47	.43
Iniciativa	-0.75	.22	0.71	.43
Desempeño general	-1.20	.22	2.03	.43

Tabla 2
Estadísticos descriptivos: criterios

Criterio	Media N = 125	DT
<i>Desempeño de tarea</i>	3.00	0.65
Resolución de problemas	2.97	0.79
Organización del trabajo	2.92	0.76
Productividad	3.12	0.83
<i>Desempeño contextual</i>	2.93	0.72
Competencia interpersonal	2.95	0.83
Desempeño más allá del puesto	3.01	0.88
Iniciativa	2.84	0.88
Desempeño general	2.97	0.65

utilizado en la mayoría de los casos las puntuaciones más altas de la escala–, además de curtosis –distribución leptocúrtica–, ya que las puntuaciones tienden a concentrarse en torno a la media de la escala, lo que contribuye al apuntamiento de la distribución.

Fiabilidad de las medidas empleadas

Se calculó el coeficiente alfa de Cronbach en aquellas oposiciones en las que el número de opositores –contratados y no contratados– que realizaron las entrevistas superó una N de 20. De las 45 oposiciones estudiadas en este trabajo, sólo son cinco las que presentaron estas características, para un total de 190 de las 379 entrevistas realizadas en total. En un caso se midieron 6 dimensiones, en dos casos se midieron 7 dimensiones y en otros dos 12. Los coeficientes alfa oscilaron entre .75 y .93. Por lo que respecta a los criterios, el coeficiente alfa del desempeño general, desempeño de tarea y desempeño contextual en este estudio fue .87, .75 y .79 respectivamente.

Efecto de la variable género

Al comparar los resultados de hombres y mujeres (tabla 4), se observa que las mujeres contratadas presentan una media ligeramente superior ($M = 3.15, SD = 0.34$) a la de los hombres contratados ($M = 3.06, SD = 0.44$). Esta diferencia no es significativa para $t(123) = -1.18, p = .24$ y el tamaño del efecto ($d = -0.23$) es pequeño en términos de Cohen (1988). Lo mismo sucede cuando se estudian las diferencias en la muestra de 379 candidatos, para $t(377) = -.90, p = .37$, con un tamaño del efecto ($d = -0.09$), pequeño según Cohen (1988) e irrelevante desde el punto de vista práctico.

Coefficientes de validez

La tabla 5 ofrece los datos de las correlaciones observadas y corregidas entre la entrevista y las medidas de desempeño. La ECE presenta una correlación observada significativa tanto con el

Tabla 4
Calificación en la entrevista según el género de los entrevistados

Género	Opositores contratados			Total de opositores		
	N	Media	DT	N	Media	DT
Hombres	90	3.06	0.44	195	2.97	0.49
Mujeres	35	3.15	0.34	184	3.02	0.51

desempeño general ($r = .27$) como con el desempeño de tarea ($r = .29$) y con el desempeño contextual ($r = .23$). Por lo que se refiere a los coeficientes de validez corregidos (validez operativa) entre la entrevista y el desempeño general, la ECE presenta una correlación significativa con las tres medidas de desempeño: ($\rho_{xy} = .44$) con DG, ($\rho_{xy} = .47$) con DT, y ($\rho_{xy} = .37$) con DC.

Según Cohen (1988), los coeficientes de correlación del orden de .10 son “pequeños”, las de .30 son “medianos” y los de .50 son “grandes” en términos de magnitud de los tamaños del efecto. A la hora de evaluar coeficientes de validez predictiva, se considera que correlaciones entre .20 y .34 son adecuadas, entre .35 y .44 son razonables, entre .45 y .54 buenas y si son superiores a .55, excelentes (Smith y Smith, 2005).

Discusión

De manera acorde a los hallazgos previos de otros autores (Berges, 2011; Sáez, 2007; Salgado et al., 2007), la ECE se ha mostrado como un predictor válido de los tres criterios de desempeño empleados, lo que confirma la primera hipótesis formulada.

Por lo que se refiere a los valores de las correlaciones observadas, son similares a los obtenidos por Salgado et al. (2007), .32, .34 y .22 para DG, DT y DC, respectivamente, ligeramente superiores a las Sáez (2007), .22, .22 y .13, e inferiores a las de Berges (2011), .37, .35 y .37. Las correlaciones no corregidas con DT y con DG alcanzan valores muy próximos a la cuantía que han registrado diferentes metaanálisis para las entrevistas basadas en conductas previas: $r = .31$ (Huffcutt et al., 2004; Taylor y Small, 2002). Además estos valores observados son similares a los que presentan los tests de muestras de trabajo, .26, según Roth, Bobko y McFarland (2005). Asimismo, las correlaciones no corregidas de este estudio se aproximan a las del mejor predictor individual del desempeño laboral, la inteligencia general, que según el metaanálisis de Salgado et al. (2003) con muestras españolas se sitúa en $r = .35$ y según el metaanálisis de Ones, Dilchert, Viswesvaran y Salgado (2009) oscila entre .25 y .27 en función del grado de complejidad del puesto.

Más interesante resulta prestar atención a las correlaciones corregidas, al tratarse de valores más próximos a la validez verdadera (Hunter y Schmidt, 2004), lo que es el objetivo principal de toda investigación. En este caso, los valores obtenidos son inferiores a los hallados en los tres estudios más similares a este trabajo. Sáez (2007) obtuvo correlaciones corregidas cuyos valores fueron de .61, .64 y .42 con DG, DT y DC, respectivamente. Salgado et al. (2007) encontraron las siguientes correlaciones: .60, .62 y .47. Berges (2011), por su parte, aplicando únicamente corrección por falta de fiabilidad de criterio, obtuvo los siguientes valores: .51, .48 y .50. Sin embargo, tomando como referencia las estimaciones

meta-analíticas, los valores de las correlaciones corregidas entre la entrevista y DG y DT son similares a los indicados por el metaanálisis de McDaniel et al. (1994), .44 para el conjunto de entrevistas estructuradas, y se aproximan a los valores hallados por Huffcutt et al. (2004) para las entrevistas de descripción de conducta, .51, mostrándose las correlaciones corregidas de la ECE con las tres medidas de criterio superiores a la cuantía hallada por este último metaanálisis para las entrevistas de descripción de conducta con diseño predictivo, .33.

En ausencia de datos de fiabilidad interjueces –en esta investigación cada persona contratada fue evaluada únicamente por un evaluador–, se empleó la estimación metaanalítica de la fiabilidad interjueces, $r_{xx} = .52$, como estimación de la fiabilidad del criterio (Viswesvaran et al., 1996). Pese a la visión de autores como Lebreton, Scherer y James (2014), que están en desacuerdo con esta corrección, que a su juicio infla artificialmente las correlaciones obtenidas tras la corrección, existe un consenso amplio basado en numerosas y potentes evidencias en que cuando se emplean escalas de evaluación el coeficiente de fiabilidad a considerar sea el de fiabilidad interevaluadores, cuya estimación es .52 en el caso de un solo evaluador (Sackett, 2014; Salgado, Moscoso y Anderson, 2016; Shen, Cucina, Walmsley y Seltzer, 2014; Viswesvaran, Ones, Schmidt, Le y Oh, 2014). Emplear el coeficiente alfa de Cronbach como estimación de la fiabilidad de la evaluación del desempeño no es una alternativa adecuada ya que solo controla una fuente de error (el error aleatorio), dejando de lado el error transitorio y el error debido a los diferentes evaluadores. El coeficiente test-retest tampoco lo es, ya que no controla la falta de acuerdo entre evaluadores. Sin embargo, los juicios del supervisor son una media útil y adecuada del desempeño y pueden ser usados con confianza como criterio, especialmente cuando los datos de desempeño se recogen en el intervalo temporal de un año, como en esta investigación (Salgado, 2015; Salgado et al., 2016; Schmidt, Le e Ilies, 2003).

Cabría pensar si un mayor grado de estructuración de la entrevista –aspectos considerados en trabajos similares a este, como añadir un segundo entrevistador o realizar las mismas preguntas sin reformularlas– podría suponer un aumento de la validez de las entrevistas realizadas por el IVAP, algo que tendrán que confirmar futuras investigaciones. En cualquier caso, su coeficiente de validez corregido al predecir DG y DT se acerca al estimado por metaanálisis para otras pruebas empleadas habitualmente en oposiciones, como pruebas de conocimientos, .48, o pruebas prácticas, .54, y es superior al de la experiencia, .18. (Gorriti, 2010; Salgado, 2010; Schmidt y Hunter, 1998).

Por otro lado, la puntuación en la ECE no se ha visto afectada por el género de los candidatos, tal y como apuntaban los trabajos de Alonso (2011), Choragwicka (2009) o Levashina et al. (2014), lo cual confirma la segunda hipótesis de este trabajo. Tal y como sugieren Levashina et al. (2014), la estructura puede contribuir a reducir el impacto de potenciales sesgos como el género en la calificación de las entrevistas, provocando que las diferencias entre grupos sean generalmente insignificantes, si bien su metaanálisis se basaba en un número pequeño de estudios.

Según Levashina et al. (2014), un aspecto clave que puede evitar que el género afecte a la puntuación es puntuar después de cada

Tabla 5
Correlaciones entre la entrevista y las medidas de desempeño

	Desempeño de tarea	Desempeño contextual	Desempeño general
Validez observada de la entrevista	.29**	.23*	.27**
Desempeño de tarea	1	.78**	.94**
Desempeño contextual	.78**	1	.95**
Desempeño general	.94**	.95**	1
Validez corregida por fiabilidad del criterio y por restricción del rango	.47**	.37**	.44**

* $p < .05$ (bilateral), ** $p < .01$ (bilateral)

pregunta y no al finalizar la entrevista, algo que coincide con la visión de [Campion et al. \(1997\)](#) que, al hablar del componente nº 8 de la estructura de las entrevistas, señalaron que emplear escalas específicas puede reducir la contaminación porque sólo se evalúan conductas relevantes.

En consecuencia, esta investigación es una evidencia más de que las ECE no producen impacto adverso de género, pues no se han encontrado diferencias en las puntuaciones de los candidatos en función de su sexo. Por ello, cabe concluir que este instrumento de selección no produce discriminación indirecta y que no incumple la legislación en materia de igualdad de oportunidades en el acceso al empleo público.

En cualquier caso, para valorar la aportación de este trabajo es necesario mencionar sus limitaciones. La limitación más importante es la ausencia de datos acerca de la fiabilidad entendida como consistencia interna del total de las entrevistas analizadas o acerca de la fiabilidad interjueces –estadístico recogido en las entrevistas realizadas en panel, las habituales en estudios de validez–, como consecuencia de haber contado en todos los casos con un único entrevistador. Como señalan [Campion et al. \(1997\)](#), añadir entrevistadores puede cancelar errores aleatorios, el recuerdo puede ser mejor y el rango de información y los juicios desde diferentes perspectivas pueden incrementar la precisión. Con todo, los resultados obtenidos son similares a los obtenidos en investigaciones anteriores realizadas sobre el conjunto de las entrevistas estructuradas (de panel o individuales), por lo que cabría pensar que la estandarización de la entrevista contribuye a mejorar la eficacia de las entrevistas individuales, además de a mejorar la validez interna.

Tampoco se debe obviar el hecho de que una persona consiga plaza y por tanto sea posible evaluar su desempeño no depende exclusivamente de su puntuación en la entrevista, sino de la suma de las puntuaciones en todas las pruebas de la oposición y el concurso –que suele consistir en la suma de puntos en función del tiempo de experiencia en puestos similares, titulaciones diferentes a la exigida, o cursos realizados. Como indicó [Sáez \(2007\)](#), los sesgos relacionados con la selección de la muestra, como los requisitos para participar en las oposiciones (edad, titulación, conocimiento de euskera) o el filtro progresivo que se produce en la realización de cada prueba del proceso selectivo, merman la deseable aleatoriedad muestral.

Asimismo, es necesario mencionar que en todos los criterios había asimetría negativa, las puntuaciones medias han sido muy altas, y curtosis, lo cual puede haber mermado la intensidad de la asociación entre predictores y criterios. Como indican [Beck, Beatty y Sackett \(2014\)](#), en evaluaciones con fines administrativos es común que todos los empleados reciban evaluaciones elevadas para que los evaluadores puedan evitar las dificultades interpersonales que podrían derivarse de una evaluación baja y las evaluaciones con un propósito de investigación, como la empleada en este caso, no están libres de esta presión, si bien es previsible que la motivación para distorsionar las puntuaciones sea menor. Según [Bartram \(2007\)](#), estos sesgos en la medida del criterio invitan a contemplar la posibilidad de emplear cuestionarios de elección forzada.

Estas limitaciones invitan a la prudencia con respecto a la posibilidad de generalizar los resultados. Pese a ello, los resultados obtenidos permiten afirmar con un amplio margen de confianza que la entrevista conductual estructurada individual es un predictor válido del desempeño para diferentes puestos del ámbito de la selección en el sector público y que no produce impacto adverso en base al género, algo que confirma los datos obtenidos por trabajos anteriores de índole similar. Esta investigación es una prueba más de que la ECE contribuye a elegir a los aspirantes más idóneos y adecuados a los puestos convocados, lo que está ligado a los principios de mérito y capacidad, según la STS de 15 de diciembre de 2011, y además lo hace respetando el principio de igualdad, según la STS de 28 de marzo de 2011.

Por último, además de las propiedades psicométricas de la entrevista conductual estructurada, es necesario considerar otro aspecto relevante en el entorno de la Administración Pública: la percepción de los candidatos acerca de la justicia en los procesos selectivos, algo esencial para evitar efectos negativos para las organizaciones, como acciones legales. ([Osca y López-Araujo, 2009](#)). La ECE se percibe por parte de los entrevistados como un instrumento objetivo, justo, exhaustivo, igualitario y relacionado con el trabajo, gracias a su elevado grado de estructuración y al contenido de las preguntas, que están vinculadas con las funciones a desempeñar en cada puesto evaluado, y como una herramienta que facilita la toma de decisiones en la selección ([Dipboye, Macan y Shahani-Denning, 2012](#); [Salgado et al., 2004](#)). Todo lo cual es un argumento más a favor de que tenga una mayor presencia en las oposiciones a todo tipo de plazas de empleados públicos y un peso similar al de las pruebas tradicionalmente empleadas, ya que cumple con los imperativos de la legislación y la jurisprudencia y presenta coeficientes de validez similares.

Extended Summary

The Public Employee Basic Statute gives interviews an important role in civil service examinations. This law requires that selection procedures used in civil service examinations cannot discriminate against women and that they must be linked to job performance ([Congreso de Los Diputados, 2007a](#)). Nevertheless, its objectivity and its weight in the selection process of public employees when compared to knowledge tests have often been questioned because of the extensive use of unstructured conventional interviews ([Chaves, 2009](#); [Mauri, 2009](#)).

In contrast with conventional interviews, structured interviews have tried to strengthen objectivity by emphasizing standardization in the implementation of this selection procedure and the evaluation of results, resulting in a more structured interview. Several studies over the past decades have shown that interviews, more specifically structured behavioral interviews, can predict job performance and do not produce gender adverse impact.

In the 1990s, several structured interview frameworks were developed ([Campion, Palmer, & Campion, 1997](#); [Huffcutt & Arthur, 1994](#)). According to [Salgado & Moscoso \(2005\)](#), the interview components utilized most frequently in structured interviews are basing questions on a job analysis, using the same questions and process with all the applicants, using behaviorally anchored rating scales (BARS), training of interviewers, and decision-taking after having conducted all the interviews.

These components are featured in the structured interview modalities that have received more attention by researchers, which were grouped by [Salgado and Moscoso \(2002\)](#) under the label of structured behavioral interview (SBI). The SBI group includes structured interviews such as behavior description interview, situational interview, multi-modal interview, and job analysis based interview ([Janz, 1982, 1989](#); [Latham, Saari, Pursell, & Campion, 1980](#); [Motowidlo et al., 1992](#); [Schuler & Funke, 1989](#)).

Meta-analysis has shown that structured interviews are valid predictors of job performance ([Huffcutt & Arthur, 1994](#); [Huffcutt, Conway, Roth, & Klehe, 2004](#); [McDaniel, Whetzel, Schmidt, & Maurer, 1994](#); [Salgado & Moscoso, 1995](#); [Taylor & Small, 2002](#); [Wiesner & Cronshaw, 1988](#)). In Spain, studies conducted in the public sector indicate that this is also the case for SBIs in civil service examinations ([Berges, 2011](#); [Sáez, 2007](#); [Salgado, Gorriti, & Moscoso, 2007](#)). The corrected validity coefficients of structured behavioral interviews are similar to those obtained by other selection methods used in the public sector, namely knowledge tests and work sample tests, and superior to the values obtained by the assessment of job experience ([Gorriti, 2010](#); [Salgado,](#)

2010). Besides, another issue researched by meta-analysis is whether structured interviews produce gender adverse impact. [Alonso's \(2011\)](#) and [Levashina, Hartwell, Morgeson, & Campion's \(2014\)](#) meta-analyses indicate that this is not the case, since there are no differences in scores for men and women in structured selection interviews.

Despite these meta-analytic results, the analysis of predictive validity and impact of gender in civil service examinations for categories or positions other than local police and command posts in the general administration has not been sufficiently studied in Spain. Furthermore, it seems necessary to help break prejudices about the interview and its application in civil service examinations checking that it meets the requirements demanded by the legislation.

Taking into account the issues highlighted by previous research and by legislation, the aim of this paper was to study to what extent SBIs are valid in predicting the performance of applicants for the public sector and to determine whether they produce gender adverse impact. The hypotheses of this paper are:

H1. The interview will significantly predict job performance.

H2. There will be no significant differences between men's and women's interview scores.

Method

Participants

The sample consisted of 379 applicants –195 men and 184 women– who took part in 45 selection processes to Basque public entities, which included an interview carried out by the Basque Institute of Public Administration (IVAP). One hundred and five of these candidates –90 men and 35 women– were employed in the public sector after being interviewed. It is, therefore, an incidental sample.

Measures

Interview. A structured behavioral interview was used to assess the adequacy of candidates to the requirements and functions of various positions ([Salgado & Moscoso, 2005](#)). Among the criteria established by [Campion et al. \(1997\)](#), interviews analyzed in this study met the criteria No. 1, 2, 3, 4, 5, 7, 8, 9, 10, 14, and 15.

It was previously defined how to rephrase behavioral description questions and follow-up questions to clarify the information expressed by the candidate, which would place this job interviews at level 3 of the [Huffcutt and Arthur \(1994\)](#) framework, although the rating scales of the answers are the same for all candidates, as in SBIs, which is an aspect of level 4 in the classification of these authors.

For each respondent, each dimension or question received a score from 0 to 4, and the final score in the interview was calculated from the average score in the dimensions evaluated.

Criterion measures: performance appraisal. A 6-item questionnaire based on the instrument used by [Salgado et al. \(2007\)](#) was utilized to evaluate job performance. The first three items, grouped around the task performance construct, 1) problem solving, 2) work organization, 3) productivity), and the last three related to the contextual performance construct, 4) interpersonal abilities, 5) performance beyond the position, and 6) initiative.

Direct supervisors evaluated the performance of the subjects that formed the sample using a 5 point rating scale that ranged from 0 – inadequate performance level – to 4 – excellent performance level. Mean scores on task performance (TP) and contextual performance (CP) yielded a value of overall performance (OP).

Procedure

A one-to-one interview (i.e., an interviewer interviewed a candidate) was applied to 379 candidates in 45 civil service examinations held after 2006. The results of interviews acted in the hands of IVAP, but to gather performance data, 12 public entities collaborated by engaging 45 supervisors –14 women and 31 men – to evaluate the performance of the 125 people hired after passing an interview. To counter the artifactual errors, validity coefficients were corrected first for criterion unreliability and then for direct range restriction.

Results

Cronbach's alpha coefficient was calculated in those examinations in which the number of candidates –hired and not hired– exceeded an N of 20. Alpha coefficients ranged between .75 and .93. The alpha coefficient of overall performance, task performance, and contextual performance in this study was .87, .75 and .79 respectively.

When comparing the results of men and women, [Table 4](#) shows that hired women have a mean score slightly higher ($M = 3.15$, $SD = 0.34$) than hired males ($M = 3.06$, $SD = 0.44$). This difference is not significant, $t(123) = -1.18$, $p = .24$, and the effect size ($d = -0.23$) is small according to [Cohen \(1988\)](#). The same is true when the differences in the sample of 379 candidates are considered, $t(377) = -0.90$, $p = .37$, with an effect size ($d = -0.09$) small in terms of [Cohen \(1988\)](#).

Observed correlations ([Table 5](#)) between the interview and performance measures are significant: .29 with task performance, .23 with contextual performance, and .27 with overall performance. Corrected correlations with the three performance measures are significant: ($\rho_{xy} = .47$) with TP, ($\rho_{xy} = .37$) with CP, and ($\rho_{xy} = .44$) with OP. According to [Cohen \(1988\)](#), correlation coefficients in the order of .30 can be considered “medium,” and those of .50, “large” in terms of magnitude of effect sizes

Discussion

According to previous studies, interviews are a valid predictor of the three performance criteria used, which supports the first hypothesis formulated. Observed correlations' values are very similar to those obtained in previous meta-analyses ([Huffcutt et al., 2004](#); [Taylor & Small, 2002](#)). Besides, these observed values are similar to those obtained by the best predictors of job performance: work sample tests and intelligence ([Ones, Dilchert, Viswesvaran, & Salgado, 2009](#); [Roth, Bobko, & McFarland, 2005](#); [Salgado et al., 2003](#)).

Regarding the corrected correlations, the values obtained are slightly lower than those obtained in the three studies that are most similar to this paper, while the values of the corrected correlation between the interview and overall and task performance are similar to those indicated by [McDaniel et al.'s \(1994\)](#), .44, or [Huffcutt et al.'s \(2004\)](#), .51, meta-analyses. It must be mentioned that unreliability of criteria has been corrected taking into account the meta-analytical estimate of interrater reliability, $r_{xx} = .52$ ([Viswesvaran, Ones, & Schmidt, 1996](#)), despite the criticism about its use made by [Lebreton, Scherer and James \(2014\)](#), as there is great consensus on the value of .52 as an accurate estimate of the interrater reliability of overall job performance for a single rater ([Sackett, 2014](#); [Salgado, Moscoso, & Anderson, 2016](#); [Shen, Cucina, Walmsley, & Seltzer, 2014](#); [Viswesvaran, Ones, & Schmidt, 2014](#)).

It could be argued that a greater degree of interview structure, such as adding a second interviewer or using the same questions without rephrasing them, could increase the validity of the interviews, something that future research should confirm. In any case,

the corrected validity coefficients, at least to predict overall and task performance, are similar to the meta-analytical values of other tests frequently used in civil service examinations, such as knowledge tests, .48, or work sample tests, .54, and superior to the values obtained by the evaluation of job experience, .18 (Gorriti, 2010; Salgado, 2010; Schmidt & Hunter, 1998).

Furthermore, interview scores have not been affected by the gender of candidates – as indicated previously by the meta-analyses by Alonso (2011) or Levashina et al. (2014), which empirically support the second hypothesis advanced.

However, to assess the contribution of this paper it is necessary to mention its limitations. The most important limitation is the lack of data about the internal consistency of all the interviews analyzed or about interrater reliability, as all the interviews were conducted by a single interviewer. As indicated by Campion et al. (1997), adding interviewers can cancel random errors, memory can improve, and the range of information and judgments from different perspectives can increase accuracy. However, the results are similar to those obtained in previous research on structured interviews taken as a whole, which is an argument in favour of one-to-one interviews. Thus, it could be argued that standardization of the interview helps to improve the effectiveness of individual interviews, in addition to improving its internal validity.

These limitations suggest caution regarding the generalizability of the results. Nevertheless, the results confirm with ample confidence that one-to-one SBIs are a valid predictor of performance in different positions in the field of selection in the public sector and that they do not produce adverse gender impact, something that confirms the conclusions of previous similar papers. Overall, the results make a strong case for interviews to have a place in civil service examinations in the same way as traditionally used selection procedures (knowledge tests, work sample tests, assessment of job experience).

Conflicto de intereses

El autor de este artículo declara que no tiene ningún conflicto de intereses.

Referencias

- Aguinis, H., Henle, C. A. y Ostroff, C. (2001). Measurement in work and organizational psychology. En N. Anderson, D. S. Ones, H. K. Sinangil y C. Viswesvaran (Eds.), *Handbook of Industrial, Work and Organizational Psychology* (1) (pp. 27–50). London, UK: SAGE. <http://dx.doi.org/10.4135/9781848608320.n3>
- Alfama, E. y Alonso, A. (2015). Las políticas de género en la administración pública. Una introducción. *Revista Vasca de Gestión de Personas y Organizaciones Públicas*, 8, 24–41.
- Alonso, P. (2011). ¿Producen resultado adverso de género las entrevistas estructuradas de selección de personal? *Revista de Psicología del Trabajo y de las Organizaciones*, 27, 43–53. <http://dx.doi.org/10.5093/tr2011v27n1a5>
- Alonso, P., Moscoso, S. y Cuadrado, D. (2015). Procedimientos de selección de personal en pequeñas y medianas empresas españolas. *Revista de Psicología del Trabajo y de las Organizaciones*, 31, 79–89. <http://dx.doi.org/10.1016/j.rpto.2015.04.002>
- American Educational Research Association, American Psychological Association and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Bartram, D. (2007). Increasing validity with forced-choice criterion measurement formats. *International Journal of Selection and Assessment*, 15, 263–272. <http://dx.doi.org/10.1111/j.1468-2389.2007.00386.x>
- Beck, J. W., Beatty, A. S. y Sackett, P. R. (2014). On the distribution of job performance: the role of measurement characteristics in observed departures from normality. *Personnel Psychology*, 67, 531–566. <http://dx.doi.org/10.1111/peps.12060>
- Berges, A. (2011). Validación de procedimientos de selección de policías locales. *Anuario de Psicología Jurídica*, 21, 83–96. <http://dx.doi.org/10.5093/jr2011v21a8>
- 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 (Eds.), *Personnel selection in organizations* (pp. 71–98). San Francisco, CA: Jossey-Bass.
- Campbell, J. P. (1990). Modelling the performance prediction model in industrial and organizational psychology. En M. Dunnette y L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (1) (pp. 687–731). Palo Alto, CA: Consulting Psychologist Press.
- Campion, M. A., Palmer, D. K. y Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology*, 50, 655–702. <http://dx.doi.org/10.1111/j.1744-6570.1997.tb00709.x>
- Choragwicka, B. (2009). El efecto de la experiencia laboral y del género de los entrevistados sobre las valoraciones y fiabilidad de una entrevista conductual. *Revista de Psicología del Trabajo y de las Organizaciones*, 25, 39–45.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: LEA.
- Congreso de Los Diputados. (2007a). Estatuto Básico del Empleado Público. *Boletín Oficial de las Oficinas del Estado*, 89, 16270–16299, 13 de abril.
- Congreso de Los Diputados. (2007b). Ley orgánica 3/2007, de 22 de marzo, para la igualdad efectiva de mujeres y hombres. *Boletín Oficial del Estado*, 71, 12611–12645, 23 de marzo.
- Conway, J. M., Jako, R. A. y Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology*, 80, 565–579. <http://dx.doi.org/10.1037/0021-9010.80.5.565>
- Chaves, J. R. (2009). *Control de concursos y oposiciones en la jurisprudencia*. Madrid: Editorial Reus.
- Dipboye, R. L. (1992). *Selection interviews: Process perspectives*. Cincinnati, OH: South-Western.
- Dipboye, R. L., Macan, T. y Shahani-Denning, C. (2012). The selection interview from the interviewer and applicant perspectives: Can't have one without the other. En N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 323–352). NY: Oxford University Press.
- Gorriti, M. (2010). *Modelo de proceso selectivo por grupos de clasificación en la Administración pública*. Barcelona: CEMICAL.
- Gorriti, M. y López, J. (2009). *Identificación de criterios de desempeño a partir del análisis de puestos*. Oñati: IVAP.
- Gorriti, M. y López, J. (2010). *Análisis de puestos en la Administración Pública*. Oñati: IVAP.
- Gorriti, M. y López, J. (2014). *Guía para la validación de sistemas de acceso a la Función Pública*. Oñati: IVAP.
- Huffcutt, A. I. y Arthur, W., Jr. (1994). Hunter & Hunter revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, 79, 184–190. <http://dx.doi.org/10.1037/0021-9010.79.2.184>
- Huffcutt, A. I., Conway, J. M., Roth, P. L. y Klehe, U. (2004). The impact of job complexity and study design on situational and behavior description interview validity. *International Journal of Selection and Assessment*, 12, 262–273. <http://dx.doi.org/10.1111/j.0965-075x.2004.280.1.x>
- Hunter, J. E. y Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.
- Janz, T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology*, 67, 577–580. <http://dx.doi.org/10.1037/0021-9010.67.5.577>
- Janz, T. (1989). The patterned behavior description interview: The best prophet of the future is the past. En R. W. Eder y G. R. Ferris (Eds.), *The Employment Interview: Theory, research and practice* (pp. 158–168). Newbury Park, CA: Sage.
- Latham, G. P., Saari, L. M., Pursell, E. D. y Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology*, 65, 422–427. <http://dx.doi.org/10.1037/0021-9010.65.4.422>
- Lebreton, J. M., Scherer, K. T. y James, L. R. (2014). Corrections for criterion reliability in validity generalization: A false prophet in a land of suspended judgment. *Industrial and Organizational Psychology*, 7, 478–500. <http://dx.doi.org/10.1111/iops.12184>
- Levashina, J., Hartwell, C. J., Morgeson, F. P. y Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, 67, 241–293. <http://dx.doi.org/10.1111/peps.12052>
- Marín, N. (2009). El acceso a la función pública. El Estatuto Básico del Empleado Público. Especial referencia a la administración local. *Revista CEMCI*, 5, 1–25.
- Mauri, J. (2009). *La selecció dels funcionaris al servei de les entitats locals de Catalunya*. Barcelona: Generalitat de Catalunya.
- McCarthy, J. M., Van Iddekinge, C. H. y Campion, M. A. (2010). Are highly structured job interviews resistant to demographic similarity effects? *Personnel Psychology*, 63, 325–359. <http://dx.doi.org/10.1111/j.1744-6570.2010.01172.x>
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L. y Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, 79, 599–616. <http://dx.doi.org/10.1037/0021-9010.79.4.599>
- Motowidlo, S. J., Carter, G. W., Dunnette, M. D., Tippins, N., Werner, S., Burnett, J. R. y Vaughan, J. (1992). Studies of structured behavioral interview. *Journal of Applied Psychology*, 77, 571–588. <http://dx.doi.org/10.1037/0021-9010.77.5.571>
- Motowidlo, S. J. y Kell, H. J. (2013). Job performance. En N. W. Schmitt, S. Highhouse y I. Weiner (Eds.), *Handbook of psychology, volume 12: Industrial and organizational psychology* (pp. 82–103). Hoboken, NJ: Wiley. <http://dx.doi.org/10.1002/9781118133880.hop212005>
- Ones, D. S., Dilchert, S., Viswesvaran, C. y Salgado, J. F. (2009). Cognitive ability. En J. L. Farr y N. T. Tippins (Eds.), *Handbook of employee selection*. Mahwah, NJ: Erlbaum.
- Oasca, A. y López-Araujo, B. (2009). ¿La justicia en selección predice las intenciones de los candidatos? *Revista de Psicología del Trabajo y de las Organizaciones*, 25, 219–229. <http://dx.doi.org/10.4321/s1576-59622009000300003>
- Ramírez, R. G. (2013). *Diseño del trabajo y desempeño laboral individual* (tesis doctoral). Universidad Autónoma de Madrid.
- Roth, P. L., Bobko, P. y McFarland, L. A. (2005). A meta-analysis of work sample test validity: updating and integrating some classic literature. *Personnel Psychology*, 58, 1009–1037. <http://dx.doi.org/10.1111/j.1744-6570.2005.00714.x>
- Rotundo, M. y Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance:

- A policy capturing approach. *Journal of Applied Psychology*, 87, 66–80. <http://dx.doi.org/10.1037//0021-9010.87.1.66>
- Sacco, J. M., Scheu, C. R., Ryan, A. M. y Schmitt, N. (2003). An investigation of race and sex similarity effects in interviews: A multilevel approach to relational demography. *Journal of Applied Psychology*, 88, 852–865. <http://dx.doi.org/10.1037/0021-9010.88.5.852>
- Sackett, P. R. (2014). When and why correcting validity coefficients for interrater reliability makes sense. *Industrial and Organizational Psychology*, 7, 501–506. <http://dx.doi.org/10.1111/iops.12185>
- Sackett, P. R. y DeVore, C. J. (2001). Counterproductive behaviors at work. En N. Anderson, D. S. Ones, H. K. Sinangil y C. Viswesvaran (Eds.), *Handbook of Industrial and Organizational Psychology* (1) (pp. 145–164). London: SAGE.
- 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. (2010). *Selección de recursos humanos por competencias en las Administraciones públicas: evidencias basadas en la investigación*. Barcelona: 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. <http://dx.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. <http://dx.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. y Moscoso, S. (1995). Validez de las entrevistas conductuales estructuradas. *Revista de Psicología del Trabajo y las Organizaciones*, 11, 9–24.
- Salgado, J. F. y Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work and Organizational Psychology*, 11, 299–324. <http://dx.doi.org/10.1080/13594320244000184>
- Salgado, J. F. y Moscoso, S. (2005). *La entrevista conductual estructurada de selección de personal. Teoría, práctica y rentabilidad*. Madrid: Pirámide.
- 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. <http://dx.doi.org/10.1016/j.rpto.2015.12.002>
- Salgado, J. F., Moscoso, S. y Gorriti, M. (2004). Investigaciones sobre la entrevista conductual estructurada (ECE) en la selección de personal en la administración general del País Vasco: metaanálisis de fiabilidad. *Revista de Psicología del Trabajo y las Organizaciones*, 20, 107–140.
- Salgado, J. F., Viswesvaran, C. y Ones, D. S. (2001). Predictors used for personnel selection: An overview of construct, methods and techniques. En N. Anderson, D. S. Ones, H. K. Sinangil y C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational Psychology* (1) (pp. 165–199). Londres: SAGE. <http://dx.doi.org/10.4135/9781848608320.n10>
- 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. <http://dx.doi.org/10.1080/1359432x.2012.716198>
- 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, 262–274. <http://dx.doi.org/10.1037/0033-2909.124.2.262>
- 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. <http://dx.doi.org/10.1037/1082-989x.8.2.206>
- Schuler, H. y Funke, U. (1989). The interview as a multimodal procedure. En R. W. Eder y G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 183–192). Thousand Oaks, CA: Sage Publications, Inc.
- 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. <http://dx.doi.org/10.1111/iops.12187>
- Smith, M. y Smith, P. (2005). *Testing people at work*. Malden, MA: Wiley-Blackwell.
- Smith, P. y Kendall, L. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors to rating scales. *Journal of Applied Psychology*, 47, 149–155. <http://dx.doi.org/10.1037/h0047060>
- Taylor, P. J. y Small, B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational and past behavior employment interview questions. *Journal of Occupational and Organizational Psychology*, 75, 277–294. <http://dx.doi.org/10.1348/096317902320369712>
- Thorndike, R. L. (1949). *Personnel selection*. New York: Wiley.
- 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 (Eds.), *Handbook of Industrial and Organizational Psychology* (1) (pp. 110–126). London: SAGE. <http://dx.doi.org/10.4135/9781848608320.n7>
- Viswesvaran, C. y Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment*, 8, 216–226. <http://dx.doi.org/10.1111/1468-2389.00151>
- 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. <http://dx.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. <http://dx.doi.org/10.1111/iops.12186>
- Wiesner, W. H. y Cronshaw, S. F. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology*, 61, 275–290. <http://dx.doi.org/10.1111/j.2044-8325.1988.tb00467.x>