

Valoraciones de Méritos (Training and Experience) en la Administración Pública y la Empresa: Fiabilidad, Validez y Discriminación de Género

Merit Ratings (Training and Experience) in the Public Administration and Industry: Reliability, Validity and Gender Differences

Pamela Alonso, Gabriel Táuriz y Beata Choragwicka
Universidad de Santiago de Compostela

Resumen. Este artículo presenta dos estudios en los que se ha examinado la fiabilidad (consistencia interna, equivalencia y estabilidad), validez de constructo y discriminación de género de las valoraciones de méritos como instrumento de selección de personal. En el primer estudio (N=72) se encontró que la valoración de méritos presentaba una fiabilidad test-retest y de acuerdo entre valoradores elevada ($r_{xx}=.93$) pero una consistencia interna baja ($\alpha=.53$). Igualmente, se observó evidencia de discriminación indirecta contra el grupo de mujeres. En el segundo estudio, dos muestras (N=42 y N=98) sirvieron para examinar la consistencia interna, la validez de constructo y la discriminación de género. Los resultados mostraron coeficientes alfa inferiores a los del primer estudio y mayor discriminación de género. Por último, se discuten las implicaciones de los resultados para la investigación y la aplicación de este instrumental en la selección de personal.

Palabras clave: valoraciones de méritos, experiencia, formación, fiabilidad, validez, discriminación.

Abstract. This article presents two studies examining reliability (internal consistency, equivalence, and stability), construct validity, and gender discrimination of merit ratings as a personnel selection procedure. The first study (N=72) found that merit rating showed a large test-retest reliability and rater's agreement ($r_{xx}=.93$) but low internal consistency ($\alpha=.53$). It was also observed evidence of gender discrimination against women. In the second study, two samples (N=42 and N=98) were used for estimating internal consistency, construct validity and gender discrimination. Results showed Alpha's coefficients smaller than study 1 and larger gender discrimination. Finally, the implications of this procedure for the research and practice of personnel selection are discussed.

Key words: merit ratings, experience, training, validity, reliability, discrimination.

La valoración de los méritos de los candidatos, basada en la puntuación de la experiencia, la educación y la formación, es uno de los métodos más utilizados por las Administraciones Públicas y, en menor medida, por las empresas en sus procesos de selección de personal y de promoción. Así, por ejemplo, en la Administración Pública española, la Ley 7/2007, de 12 de Abril, del Estatuto Básico del Empleado Público, en su artículo 61, establece que los sistemas selectivos para el acceso al empleo público como funcionario de carrera y personal laboral fijo consistirán en oposición y concurso-oposición en la que podrá haber una valoración de méritos y que está última será la única posi-

ble prueba en la fase de provisión de puestos. Una situación similar se presenta en Estados Unidos, donde el concurso de méritos es el sistema más empleado en el sector público (Cook, 1980). Por ejemplo, recientes encuestas llevadas a cabo por la Internacional Public Management Association for Human Resources (IMPA-HR, 2006) muestran que más del 70% de las administraciones públicas utilizan métodos de valoración de méritos para la selección de personal.

La valoración de méritos se incluye dentro de la categoría de métodos de selección denominados en la literatura anglosajona como "Training and Experience Evaluations" (T&E) (Gatewood y Field, 1998). Las T & E Evaluations en general y la valoración de méritos en particular se basan en el supuesto de que la experiencia previa en el puesto y la formación se relacionan positivamente con el futuro desempeño en el trabajo (Ash, Johnson, Levine y McDaniel, 1989), apoyándose en el principio de que el mejor predictor de la conducta futura es la conducta pasada (Principio de la consistencia conductual). Las T&E consisten en la evalua-

La investigación incluida en este artículo ha sido financiada con cargo a los proyectos de investigación del Plan Nacional de I+D PSI2008-03617 y EDU2008-03592. Todos los autores contribuyeron igualmente. La correspondencia sobre este artículo deberá enviarse al primer autor a la siguiente dirección: E.U. de Relaciones Laborales, Universidad de Santiago de Compostela, 15782 Santiago de Compostela. E-mail: pamela.alonso@usc.es o gabriel.tauriz@usc.es

ción de la información aportada por el solicitante para determinar sus capacidades en relación al puesto, en vez de en medidas directas de sus competencias. Dicha información suele ser aportada en forma escrita, ya sea a través de formularios cumplimentados, o a través de certificados o acreditaciones. Un evaluador valorará la información aportada a partir de un conjunto de criterios o indicadores previamente diseñados. Planteado de esta forma, este método de evaluación de los candidatos está basado en criterios racionales, no empíricos, que establecen que la acreditación de una determinada formación o experiencia son indicadores suficientes para predecir el desempeño laboral en un puesto concreto (Ash et al., 1989).

Dentro de las T&E Evaluations existen al menos cinco tipos de métodos diferentes: (1) Método basado en la Tarea (Task-Based Method; Ash et. al. 1989) en el que se le presenta al evaluador un listado con las tareas que engloba el nuevo puesto y éste debe puntuarlas en función de la experiencia previa aportada por el candidato en esa tarea en otros puestos y en qué grado llega a dominarla; (2) Método de Conocimiento, Destreza y Capacidad (Knowledge, Skill & Ability (KSA) Method; Levine, Ash y Levine, 2004), muy similar al anterior, pero en este caso, se valora el grado de dominio de los candidatos en unos determinados conocimientos, destrezas y habilidades, en vez de puntuar tareas; (3) Método de la Consistencia Conductual (Behavioral Consistency Method; Schmidt, Caplan, Bemis, Decuir, Dunn y Antone, 1979), consistente en que el candidato tiene que describir con detalle al menos dos logros pasados que demuestren su dominio en ciertas competencias consideradas clave para el puesto, debiendo incluir en su redacción la referencia de alguien que pueda confirmar lo dicho; (4) Método de Agrupamiento (Grouping Method; Porter, Levine y Flory, 1976), en el que se clasifica a los candidatos en varios grupos, por ejemplo, muy cualificados, cualificados y no cualificados; (5) por último, el Método de Puntos (Point Method) consistente en otorgar una puntuación al candidato en función de que posea o acredite un determinado mérito, como puede ser el número de meses o años de formación, educación o experiencia laboral (Ash et. al., 1989). La valoración de méritos, tal y como se lleva a cabo en nuestro país, encajaría dentro de los denominados "Método de Puntos" ya que se trata de la valoración de las acreditaciones de los candidatos a través de unos criterios y baremos preestablecidos, en los que se establece una puntuación máxima para cada indicador o apartado puntuable, en función de la importancia que se le otorgue al mismo.

La larga tradición en el uso de los sistemas de valoración de méritos parece deberse a tres suposiciones: (1) por un lado, a que se trata de un método económico y rápido y que además permite hacer un ranking con los candidatos y, al mismo tiempo, a que es considerado un método "justo" por candidatos y sindicatos, ya

que, aparentemente, es independiente del criterio de los jueces (valoradores), totalmente objetivo y no discriminatorio; (2) por otro lado, a que es un método válido, es decir a que las puntuaciones de la valoraciones se relacionan ampliamente con el desempeño en el puesto de trabajo, de tal modo que las personas que reciben mayores puntuaciones en la valoración de méritos serían también las que mostrarían más tarde un mejor desempeño en el trabajo; (3) por último, a que es un método totalmente fiable, en tanto que no existiría error de medida en las puntuaciones asignadas a los candidatos; en otras palabras, que si dos valoradores puntúan independientemente uno de otro los méritos de los mismos candidatos, la puntuación final otorgada a cada candidato por cada uno de los valoradores será la misma (acuerdo entre jueces o valoradores) o bien, que si un mismo valorador puntúa los mismos méritos del mismo grupo de candidatos en dos momentos temporales distintos, separados entre si unas semanas, la puntuación en el primer momento y en el segundo será exactamente la misma. Seguidamente, examinamos más detalladamente estas tres suposiciones y las evidencias empíricas disponibles.

La "justicia percibida" del método se debe al supuesto de que existe total objetividad en su uso. Muchos usuarios del mismo, honestamente creen que al asignar, con anterioridad a su aplicación, una puntuación numérica a cada mérito se evita totalmente la subjetividad del evaluador. Sin embargo, la simplicidad y objetividad del método son ilusorias, dado que éste cuenta con una serie de inconvenientes importantes. En primer lugar, la utilización de los baremos no es tan simple como aparenta ser, ya que de no estar bien definidos, podrían dar lugar a diferentes interpretaciones por parte de los evaluadores. Otro aspecto a tener en cuenta es el modo en que los candidatos presentan su documentación, pues de no estar bien ordenada, puede suponer un problema para el evaluador, que podría pasar por alto alguna puntuación, o que le resultase complicado establecer los cálculos de determinadas experiencias. Lo mismo ocurre con las fechas en los documentos que acreditan experiencia, así, cuanto más concretos son estos documentos, más perjudicados salen los candidatos, puesto que son valorados únicamente por los días que especifican, mientras que otros candidatos que aportan información más genérica, ante el desconocimiento de fechas concretas, resultan valorados de una forma más positiva. Así, una persona que haya trabajado entre el 25 de noviembre y el 15 de diciembre, si en la documentación aportada especifica las fechas, será puntuada por 20 días de experiencia, mientras que si la documentación aportada indica que trabajó entre noviembre y diciembre, será puntuada por dos meses. En cuanto a la objetividad del método, queda patente si consideramos que existe un amplio margen a la discrecionalidad del evaluador, dado que en la mayoría de los casos los baremos de puntuación están fijados de un modo muy general, que se prestan a

múltiples interpretaciones y que difícilmente pueden adaptarse a cada candidato en concreto que debe ser evaluado. Otro aspecto a tener en cuenta es la fijación de los baremos, quién es el encargado de fijarlos y bajo qué criterios. Estos baremos solo podrían fijarse de un modo objetivo en base a un previo análisis del puesto que permitiese conocer cuales son las características necesarias para el ocupante del mismo. Sin embargo, en la mayoría de los casos, este análisis previo no existe (Villasante y Bretones, 1995). Dada la ausencia de este análisis de puestos previo, se puede comprender fácilmente que los baremos diseñados raramente se ajustan a las necesidades y requisitos del puesto. De algunos de estos problemas se hizo eco la Comisión redactora del Informe para el estudio y preparación del Estatuto Básico del Empleado Público, cuando señaló que la valoración de méritos “no siempre ofrece las necesarias garantías de igualdad”, dado que existe “una amplia discrecionalidad de cada administración para determinar los méritos del baremo correspondiente”. De tal manera, que es posible que el concurso de méritos acabe siendo utilizado “para la consolidación en el empleo de interinos o contratados temporales que han accedido a esa condición sin un verdadero procedimiento competitivo, sobrevalorando como mérito los servicios prestados”.

Además de lo anterior y relacionado con la “justicia percibida”, otro importante inconveniente, mencionado por la literatura americana pero que, sin embargo, no ha tenido el suficiente eco en nuestro país, es la posibilidad de que la valoración de méritos ejerza impacto adverso en minorías raciales, mujeres o jóvenes. En otras palabras, que produzca discriminación indirecta o, más técnicamente, impacto adverso. Así, por ejemplo, Ash y Levine (1981) encontraron evidencias de impacto adverso de la valoración de méritos en mujeres en dos de los tres trabajos evaluados. En aquellos trabajos en los que existe un índice bajo de representación femenina, o en puestos de responsabilidad y directivos, las mujeres han tenido menos oportunidades para acumular años de experiencia que puedan presentar como mérito. La misma situación se encontraría en minorías raciales. Esto implica que la valoración (concurso) de méritos puede contribuir a la perpetuación de situaciones de discriminación. Esta situación podría propiciar la aparición de demandas por discriminación ante la utilización de la valoración de méritos, al vulnerar los principios de igualdad entre mujeres y hombres establecidos por la Constitución Española y recientemente regulados por la Ley Orgánica 3/2007, de 22 de marzo, para la igualdad efectiva de mujeres y hombres. Estas demandas podrían prosperar si se demuestra la existencia de impacto adverso sin el respaldo de un estudio que demuestre la validez de estas pruebas. Así, en el caso de Estados Unidos, ya hay precedentes de demandas por este motivo (por ejemplo, Kennedy vs. Crittenden, Civil Action No. 77-200-MAC-WDO).

En cuanto a la validez de este tipo de valoraciones de méritos, en un primer meta-análisis, Hunter y Hunter (1984) encontraron una validez operativa de .13. Pocos años más tarde, en un meta-análisis más exhaustivo, McDaniel, Schmid y Hunter (1988) hallaron un coeficiente de validez, corregido por restricción de rango y falta de fiabilidad en el criterio, de .11 ($N=6741$). Además, los resultados indicaron que este coeficiente no puede ser generalizado a través de las situaciones dado que el intervalo de credibilidad para el 90% incluye el valor 0, es decir, hay ocasiones en las que el método proporciona una validez de 0 o negativa. La conclusión a la que se llega a partir de estos datos es que este método tiene una validez muy baja, que además no es generalizable. Son varias las razones que motivan esta baja validez. El hecho de que dos personas cuenten con la misma formación o con la misma experiencia laboral no indica que hayan adquirido los mismos conocimientos, habilidades o destrezas. Así, dos personas que han recibido la misma formación asimilan conocimientos diferentes y resulta fácil distinguir en un mismo aula importantes diferencias en cuanto a la capacitación de los distintos alumnos. Lo mismo es aplicable en cuanto a la experiencia laboral. Dos trabajadores en un mismo puesto desarrollan capacidades diferentes aunque hayan empezado a trabajar el mismo día. También se ha observado que este método es más válido cuando el conjunto de aspirantes tiene bajos niveles de experiencia laboral y menos válido con solicitantes con amplia experiencia (McDaniel et al., 1988).

Por lo que se refiere a la fiabilidad del método, se suele asumir que éste es totalmente fiable. Sin embargo, antes de examinar la evidencia disponible, es preciso tener en cuenta a qué tipo de error nos estamos refiriendo cuando mencionamos la fiabilidad de las valoraciones de métodos. Schmidt y Hunter (1996) señalan que los errores de medida pueden ser intrasujeto o entre sujetos. Por lo que se refiere al primer tipo, los coeficientes de fiabilidad podrían estimar la consistencia interna de la medida, la equivalencia de las puntuaciones o los errores debidos al paso del tiempo. En el segundo tipo, la correlación entre jueces o valoradores sería el modo de estimar la fiabilidad entre sujetos. En relación con estos tipos de coeficientes de fiabilidad, pocas investigaciones han sido publicadas sobre la fiabilidad de las evaluaciones de experiencia y formación en general y sobre las valoraciones de méritos en particular, por lo que se necesita más investigación al respecto, sobre todo teniendo en cuenta que la fiabilidad de la medida es una condición necesaria para su validez y, además, de su escasez, la evidencia disponible es contraria al supuesto de la extrema fiabilidad del método. Así, por ejemplo, Ash y Levine (1985) encontraron en su estudio una fiabilidad entre jueces media de .83. Por su parte, Frank Schmidt y sus asociados (1979), en un estudio inédito, encontraron que la fiabilidad promedio de los métodos T&E se sitúa en torno

a .80. Debemos tener en cuenta, como señalan los propios autores, que esta fiabilidad es el resultado de evaluar dos veces la misma información aportada por los candidatos. Si se les permitiese aportar de nuevo la información, la fiabilidad podría resultar más baja al tener en cuenta la variabilidad intra-solicitante. Así pues, a partir de los datos anteriores se puede decir que la evidencia disponible se ha concentrado en el examen de la fiabilidad entre valoradores y que el tamaño promedio de los coeficientes es muy aceptable. Sin embargo, la evidencia disponible no nos permite indicar cuál puede ser el rango de los coeficientes de consistencia interna (por ej. alfa de Cronbach) de las valoraciones de méritos, ni tampoco cual es la estabilidad temporal de las puntuaciones. En otras palabras, si existe homogeneidad en la medida (experiencia y formación medirían los mismos constructos subyacentes) y si las puntuaciones dadas a los méritos varían con el paso del tiempo, sin haberse modificado los datos originales. Es preciso también tener en cuenta que los coeficientes mencionados han sido obtenidos con muestras y procedimientos de Estados Unidos y que en la literatura de investigación española no se encuentran investigaciones sobre la fiabilidad de la valoración de méritos tal y como se emplea en nuestro país.

Teniendo en cuenta lo indicado hasta el momento, el objetivo de este artículo es triple: (1) Por un lado, examinar la fiabilidad de las valoraciones de méritos en la Administración Pública y en la empresa, estimando tanto coeficientes intravalorador (consistencia interna y estabilidad temporal) como entre valoradores (acuerdo entre valoradores); (2) examinar la validez de constructo de las valoraciones de méritos; y (3) examinar si las valoraciones de méritos son imparciales y justas y

no producen discriminación indirecta para ninguno de los dos géneros. A este fin, presentamos los resultados de dos estudios independientes, con tres muestras independientes, reclutada cada una de ellas para un puesto de trabajo diferente.

ESTUDIO 1

Método

Muestra

Para la realización de este estudio se valoraron 72 candidaturas para un puesto de docente de idiomas en una organización asociada a una universidad pública española. La muestra estaba formada por 14 hombres (19.4%) y 58 mujeres (80.6%). La mayor parte poseían experiencia docente (83.3%), y el 33.3% había trabajado anteriormente en esa misma organización, en ese mismo puesto. Un 52% de los candidatos evaluados poseían la licenciatura en filología hispánica, mientras que el 43% tenían otra licenciatura afín.

Instrumento

El método de selección empleado fue la valoración de méritos, basados en experiencia laboral, educación y formación. Los méritos que se tuvieron en cuenta en la baremación fueron, atendiendo a la formación, el título académico (licenciatura en filología hispánica, 2 puntos, u otras licenciaturas afines, 1 punto), la formación de postgrado (máximo 2 puntos), formación espe-

Tabla 1. Sistema de valoración de méritos

Méritos	Puntuación máxima
1. Formación Académica	8 puntos
1.1 Licenciatura específica	2 puntos
1.2 Licenciatura afín.	1 punto
1.3 Otras titulaciones (postgrado, cursos especializados) sobre didáctica de la lengua objeto del concurso o de lenguas en general, distintas a las evaluadas en los puntos anteriores.	2 puntos
1.4 Cursos de formación EFE (Español con fines específicos)	2 puntos
1.5 Tesis licenciatura/doctoramiento.	1 punto
2. Experiencia Docente	12 puntos
2.1. Experiencia en la enseñanza de la lengua objeto del concurso, en la empresa que convoca la plaza (0,1 por mes trabajado en la empresa)	7 puntos
2.2. Experiencia docente en otras áreas.	5 puntos
a) En centros públicos: 0,1 puntos por mes trabajado	
b) En otros centros: 0,05 por mes trabajado	
3. Otros méritos	5 puntos
3.1. Coordinación y gestión docente (cargo responsabilidad)	2 puntos
3.3. Publicaciones (artículos, libros, dossier y traducciones)	1 punto
3.3. Ponencias y comunicaciones en congresos, seminarios, cursos, monográficos, etc. relacionados con la enseñanza de lenguas	1 punto
3.4. Conocimientos de otras lenguas	1 punto

cífica de didáctica de la lengua (máximo 2 puntos) y el doctorado (máximo 1 punto) (ver Tabla 1). En cuanto a la experiencia laboral, se tuvo en cuenta la experiencia docente en centros públicos y privados (máximo 5 puntos entre las dos) y la experiencia previa en la entidad que convocaba el concurso (máximo 7 puntos). Otros méritos valorados fueron la experiencia previa en coordinación y gestión docente (máximo 2 puntos), las publicaciones (máximo 1 punto) y ponencias (máximo 1 punto) relacionadas con la didáctica de la lengua, y el conocimiento de idiomas (máximo 1 punto).

Procedimiento

Se realizó una oferta pública de empleo y las personas que, cumpliendo los requisitos señalados en la convocatoria, lo estimaron oportuno, se presentaron a la misma. Para ello cumplimentaron el formulario de solicitud y aportaron la documentación que consideraron necesaria para la evaluación conforme a los baremos publicados. La documentación fue, posteriormente, remitida a los evaluadores para su valoración. El propósito de la ponderación era la elaboración de una jerarquía de los candidatos, ordenados de mayor a menor según su adecuación a la convocatoria. Conscientes de las dificultades de la utilización de este método de evaluación, y tratando de reducir las posibilidades de error al mínimo posible, se repasaron los baremos entre los jueces, determinando las puntuaciones al detalle hasta alcanzar un acuerdo en las cuestiones más problemáticas, como por ejemplo, el cómputo de la experiencia laboral, en el caso de que las fechas aportadas no fuesen específicas. Así mismo, se realizó un ensayo piloto para aclarar las dudas y asegurar que todos los evaluadores empleasen los mismos criterios. Para ello, dos de las candidaturas fueron evaluadas en conjunto por todos los evaluadores. Las dificultades y ambigüedades encontradas permitieron clarificar los criterios de valoración. Con la finalidad de conocer la estabilidad temporal de las puntuaciones y el acuerdo entre jueces se realizó una segunda evaluación un mes después de la primera. En esta ocasión, los evaluadores puntuaron candidaturas diferentes a las que habían valorado en la ocasión anterior. Hay que señalar que, en este punto, los jueces no tenían acceso a las puntuaciones obtenidas por los candidatos en la primera baremación. Con el propósito de facilitar la interpretación de los datos se agruparon conceptualmente las puntuaciones bajo cuatro epígrafes: Formación, Experiencia, Titulación y Publicaciones.

Resultados

En la Tabla 2 figuran los coeficientes de fiabilidad calculados para la puntuación total de la valoración de

méritos utilizada en el presente estudio. Hemos calculado la consistencia interna de los elementos mediante el coeficiente alfa de Cronbach, la equivalencia entre las partes mediante la fórmula de Spearman-Brown y la estabilidad temporal a través de la correlación entre dos puntuaciones separadas entre sí 4 semanas y realizadas por valoradores diferentes. Por tanto, hemos examinado tanto errores intrasujeto como entre sujetos (valoradores). Los resultados indican que la puntuación final de los méritos presenta, en el presente caso, una fiabilidad relativamente baja (.55) en comparación con los estándares habitualmente utilizados para el coeficiente alfa de Cronbach. El indicador de equivalencia entre las partes, obtenido a partir de la fórmula de Spearman-Brown se sitúa en los límites de lo que sería aceptable (.73) para una investigación, no tanto para una aplicación profesional y, por último, la estabilidad temporal de la puntuación total es muy elevada (.93), lo que indica que los valoradores no tienden a cambiar sus puntuaciones de un modo importante.

Tabla 2. Fiabilidad de la valoración de méritos para el puesto de docente

Alfa de Cronbach	.55
Spearman-Brown	.73
Test-Retest	.93

Nota. N = 72

Analizando más específicamente la estabilidad temporal de cada una de las partes que componen la valoración de méritos y que contribuyen a la puntuación total, se puede observar (ver Tabla 3) que en general los coeficientes son muy aceptables, con un rango de valores que va desde .85 para las publicaciones a .91 para la educación (titulación) y la experiencia. La formación se sitúa en un nivel intermedio (.89). Tomados en conjunto, estos resultados vuelven a indicar la robustez de la puntuación de las valoraciones de méritos frente al paso del tiempo, si se utiliza un procedimiento como el descrito en la sección de método, es decir, si se especifican detalladamente los criterios de valoración y se realiza un entrenamiento previo de los evaluadores.

Por lo que se refiere a la relación entre las distintas partes componentes del instrumento y que contribuyen a la puntuación final, las correlaciones entre las mismas aparecen en la Tabla 4. Estas correlaciones sirven para obtener una indicación aproximada de la validez de constructo del instrumento. Como puede apreciarse, las correlaciones son en todos los casos bajas o muy bajas (inferiores a .30), a excepción de la correlación entre la puntuación en formación y la puntuación en experiencia que muestra una correlación de tamaño moderado (.56). No obstante, esta correlación debe interpretarse con prudencia ya que puede estar afectada por la edad de los sujetos. Al no disponer de este

Tabla 3. Correlaciones entre las dos valoraciones de la muestra de docentes

	$\bar{X}1$	$\bar{X}2$	SD1	SD2	r
Total	7.60	7.71	4.27	3.85	.93
Educación	1.68	1.59	0.84	0.85	.91
Formación	2.15	2.07	1.15	1.24	.89
Experiencia	3.13	3.17	2.44	2.88	.91
Publicaciones	0.75	0.77	0.86	0.89	.85

Nota. N = 71. $\bar{X}1$ = media primera valoración; $\bar{X}2$ = media segunda valoración; SD1= desviación típica primera valoración; SD2= desviación típica segunda valoración.

dato, no hemos podido calcular la correlación parcial entre experiencia y formación, después de controlar los efectos de la edad, pero no es muy aventurado sospechar que el tamaño de la correlación hubiese sido menor, ya que más edad permite acceder a más cursos de formación y tener más oportunidades de empleo que proporcionen experiencia. En síntesis, los datos mostrados en la Tabla 4 sugieren que la puntuación asignada mediante un instrumento de valoración de méritos es de carácter multidimensional y compensatoria, ya que puntuaciones bajas en una variable pueden ser compensadas con puntuaciones altas en otra. Por ejemplo, se obtendría la misma calificación final, mediante una puntuación de 7 (notable) en formación y de 3 (insuficiente) en experiencia que a la inversa, 3 en formación y 7 en experiencia, pero las cualificaciones y competencia personal para realizar el trabajo no serían idénticas en ambos casos.

Tabla 4. Correlaciones entre los conceptos medidos en la muestra de docentes

	Educ.	Form.	Exp.	Publ.
Educación	1.00			
Formación	.13	1.00		
Experiencia	.19	.053	1.00	
Publicaciones	.30	.30	.26	1.00

Nota. N = 72

Por último, se ha analizado la posible discriminación indirecta en función del género de los sujetos valorados. En el presente caso, y como es usual en los estudios de discriminación, se ha utilizado la *d* de

Cohen como indicador de la existencia de ventaja en las puntuaciones de un grupo sobre otro. En la Tabla 5 aparecen los resultados de este análisis. Como puede apreciarse, en el presente caso y para la puntuación acumulada de experiencia y formación, las mujeres puntúan .29 desviaciones típicas menos que los hombres. Lo que siguiendo a Cohen nos situaría en un tamaño de efecto pequeño. Por lo que se refiere a la experiencia, el tamaño del efecto es también pequeño, puntuando las mujeres por debajo de los hombres. Finalmente, en el caso de la formación, las mujeres puntúan .43 desviaciones típicas por debajo de los hombres, lo que indica un tamaño de efecto moderado. Así pues, los resultados indican que aunque la cuantía de la discriminación es pequeña o moderada, la valoración de méritos no es imparcial para las mujeres ya que se observa una cierta discriminación indirecta como consecuencia del instrumento utilizado. Es importante señalar en este contexto que la discriminación no es obra de los valoradores en absoluto, sino de que los elementos que contribuyen a la valoración (es decir, formación, experiencia, etc.) suponen una relativa ventaja para los hombres en comparación con las mujeres, ventaja que es independiente de quien tiene que aplicar el instrumento valorativo.

Discusión

Los resultados de este estudio indican que la puntuación total del instrumento de valoración de méritos utilizado no se ve muy afectada por cambios de valoradores o por el paso del tiempo (lo que podría hacer cambiar de criterios a los valoradores), lo que lo convierte en un instrumento robusto frente a este error de

Tabla 5. Tamaño del efecto de la discriminación por género (estimado como delta de Cohen) en la muestra de los docentes

	\bar{X}_m	\bar{X}_h	SDm	SDh	d
Experiencia (E)	3.04	3.51	2.50	2.21	-.19
Formación (F)	2.05	2.54	1.20	0.84	-.43
E + F	5.10	6.04	3.30	2.73	-.29
	n= 58	n=14			

Nota: \bar{X}_m = media mujeres; \bar{X}_h = media hombres; SDm= desviación típica mujeres; SDh= desviación típica hombres; d= delta de Cohen; E+F= experiencia y formación.

medida. Esto iría en la línea de los resultados hallados por Ash y Levine (1985) y Schmidt y colaboradores (1979). Sin embargo, por lo que se refiere a la consistencia interna del mismo y a la relación entre los elementos constituyentes, la valoración de méritos muestra una importante debilidad, ni es consistente ni hay una relación sólida entre los elementos, lo que lo hace susceptible de problemas de decisión de selección al estar basado en un formato compensatorio. Sería necesaria la realización de más estudios sobre esta temática para comprobar si los resultados obtenidos podrían ser generalizables.

Por último, los datos indican que, frente a la creencia generalizada que la valoración de méritos es objetiva e imparcial, la puntuación discrimina indirectamente a las mujeres con una magnitud entre pequeña y moderada. Este aspecto es especialmente importante si se tiene en cuenta que de ser así, sería contrario a la Ley Orgánica 3/2007, de 22 de marzo, para la igualdad efectiva de mujeres y hombres, que vela por la igualdad de trato en lo que se refiere al acceso al empleo, a la formación y a la promoción profesional. Esto podría ocasionar la aparición de demandas por discriminación indirecta si no se contraponen evidencias de validez.

ESTUDIO 2

Método

Muestra

Para la realización de este estudio se emplearon dos muestras de solicitantes de dos puestos diferentes. La primera muestra se constituye de 48 solicitantes para un puesto de director técnico. La muestra se divide en 24 hombres (57.14%) y 18 mujeres (42.86%). Un 55.6% de los cuales no cuentan con experiencia previa alguna, y un 15.6% obtienen la máxima puntuación en esta dimensión. En cuanto a la formación, un 13.3% no puntúan en la misma, mientras que más de la mitad de los solicitantes (53.3%) alcanza la puntuación máxima en este caso. En lo que se refiere a informática, algo más de la mitad de los solicitantes (51.1%) consigue la máxima puntuación y un 28.9% no puntúa en este aspecto.

La segunda muestra está formada por 98 solicitantes para un puesto de técnico. En este caso la muestra se divide a partes iguales entre hombres y mujeres (49 candidatos de cada sexo). En cuanto a la puntuación en experiencia, un 10.2% de los solicitantes no obtiene puntuación, frente a un 32.7% que alcanza el máximo. Más del 65% puntúa lo máximo en formación, y un 6.1% no puntúa en este criterio. En lo que se refiere a informática, la mayoría de los solicitantes (71.4%) obtiene la puntuación máxima y un 14.3% no puntúa.

Instrumento

El método empleado para este proceso de selección fue la baremación de méritos. Se valoró la experiencia profesional en puestos similares al convocado, en entidades o empresas del sector público (máximo 3 puntos); la formación relacionada con el puesto, en forma de cursos impartidos por las administraciones públicas u homologados por ellas (máximo 2 puntos) y, por último, los conocimientos de informática del solicitante (máximo 1 punto).

Procedimiento

Una organización vinculada a la administración pública autonómica convocó, mediante concurso público, varias plazas para los puestos de técnico provincial y director técnico. Aquellas personas interesadas debían remitir la documentación acreditativa de sus méritos a la organización convocante, junto con la solicitud debidamente cumplimentada y sus documentos identificativos, especificando el puesto o puestos a los que se optaba. Una vez que los evaluadores recibieron la documentación acreditativa procedieron a la evaluación de las candidaturas para cada uno de los puestos.

Resultados

En este estudio hemos calculado la consistencia interna de los elementos mediante el coeficiente alfa de Cronbach y la equivalencia entre las partes mediante la fórmula de Spearman-Brown para los dos puestos convocados. En la Tabla 6 se muestran los coeficientes de fiabilidad calculados para la puntuación total de las dos muestras, es decir, tanto para el puesto de director técnico, como para el de técnico provincial. En el caso del puesto de director técnico, encontramos valores muy bajos y, en consecuencia, inaceptables desde el punto de vista psicométrico. Así, los resultados hallados muestran que la puntuación final de los méritos presenta un coeficiente Alfa igual a .41, notablemente más bajo que el encontrado en el estudio 1, ya de por sí bajo. En cuanto al indicador de equivalencia entre las partes (fórmula de Spearman-Brown), el resultado también es bajo y considerablemente menor que el del

Tabla 6. Fiabilidad de la valoración de méritos en las muestras para los puestos de director técnico y técnico.

<i>Director Técnico (n= 42)</i>	
Alfa de Cronbach	.41
Spearman-Brown	.44
<i>Técnico (n= 98)</i>	
Alfa de Cronbach	Negativo
Spearman-Brown	Negativo

Tabla 7. Estadísticos descriptivos y correlaciones entre las valoraciones de méritos para los puestos de director técnico y técnico

	\bar{X}	SD	Formación	Experiencia	Informática
<i>Director Técnico (n= 42)</i>					
Formación	1.43	0.79	1.00		
Experiencia	0.79	1.43	.28	1.00	
Informática	0.65	0.46	-.07	.06	1.00
<i>Técnico (n=98)</i>					
Formación	1.60	0.66	1.00		
Experiencia	1.75	1.10	-.49	1.00	
Informática	0.82	0.36	-.28	.06	1.00

estudio anterior, al obtenerse un valor de .44. Para el puesto de técnico provincial nos encontramos con resultados aberrantes, tanto para la consistencia interna, como para la equivalencia entre las partes, ya que resulta una fiabilidad negativa e inferior a -1. Esto significa que la baremación de méritos empleada en este caso muestra una fiabilidad nula.

En la Tabla 7 se muestra la correlación entre las diferentes partes que componen el baremo de méritos, lo que, al igual que en el estudio anterior, nos permite conocer de un modo aproximado la validez de constructo de este método. En el caso del puesto de director técnico, encontramos una correlación entre formación y experiencia igual a .28, lo que indica una relación baja entre estos dos conceptos e inferior a la encontrada en el estudio anterior. En cuanto a la correlación de los conocimientos de informática con formación y experiencia, nos encontramos con índices muy bajos (.07 y .06), llegando a ser negativa en el caso de la formación. Por lo que se refiere al puesto de técnico provincial, la correlación entre experiencia y formación es de -.49, es decir, se trata de una relación de tamaño moderado, pero con signo negativo. Lo mismo ocurre entre formación e informática, donde la correlación es menor, pero también con signo negativo (-.28). En el caso de la correlación entre los conocimientos de informática y la experiencia, el resultado coincide con el puesto de director técnico (.06).

Al igual que en el estudio 1, se debe ser prudente al interpretar estas correlaciones, pues en este caso, como en el anterior, tampoco se conoce la edad de los candidatos para poder controlar sus efectos sobre las correlaciones. El hecho de que la relación sea negativa indica que las puntuaciones se compensan entre sí. Así, la tendencia es que aquellas personas que puntúan alto en una dimensión, puntúan bajo en la otra y viceversa.

Del mismo modo que en el estudio anterior, en este caso también se ha analizado la posibilidad de la existencia de discriminación indirecta en función del género de los candidatos. En la Tabla 8 pueden verse los resultados, donde, una vez más, se ha empleado la *d* de Cohen como indicador. En el caso del puesto de director técnico, nos encontramos con un tamaño del efecto de -.47 desviaciones típicas para la suma de las puntuaciones de experiencia y formación. Este efecto se ve incrementado hasta -.72 en el caso de la experiencia, algo significativo dado que se trata de un puesto de dirección. En cuanto al puesto de técnico provincial, la diferencia entre las puntuaciones es de -.59 en el caso de la suma de experiencia y formación, y de -.60 desviaciones típicas, en el caso de la experiencia. En lo que se refiere a formación, el tamaño del efecto es más bajo en los dos casos (.07 en el caso de directores técnicos, y .12 en el de técnicos provinciales), en este caso la diferencia es a favor del género femenino.

Tabla 8. Tamaño del efecto de la discriminación por género (estimado como delta de Cohen) en las muestras de directores y técnicos

	\bar{X}_m	\bar{X}_h	SD _m	SD _h	<i>d</i>
<i>Directores</i>					
Experiencia	0.33	1.14	0.83	1.30	-.72
Formación	1.47	1.41	0.82	0.78	.07
E + F	1.80	2.54	1.32	1.73	-.47
	n= 18	n= 24			
<i>Técnicos</i>					
Experiencia	1.43	2.07	1.05	1.08	-.60
Formación	1.64	1.56	0.66	0.67	.12
E + F	3.07	3.62	0.86	1.00	-.59
	n= 49	n= 49			

Nota. \bar{X}_m = media mujeres; \bar{X}_h = media hombres; SD_m= desviación típica mujeres; SD_h= desviación típica hombres; *d*= delta de Cohen; E+F= Experiencia y formación.

Discusión

A la vista de los resultados obtenidos, podemos decir que, una vez más, se encuentran debilidades en cuanto a la fiabilidad del instrumento, ya que ésta obtiene valores bajos o incluso nulos. Esto invita a cuestionar la viabilidad de este método, dados los problemas que puede ocasionar la utilización de un instrumento poco fiable. No se encuentran otros estudios previos que analicen estas cuestiones, por lo que es imposible realizar cualquier tipo de comparación. Sería muy interesante disponer de otros estudios con muestras más amplias y otros tipos de puestos de trabajo para poder estudiar mejor este aspecto. Por otro lado, en cuanto a la validez de constructo, los resultados muestran que se trata de un método en el que no se encuentra una relación sólida entre los elementos, y ésta puede, incluso, llegar a ser negativa.

Por último, en el caso de este estudio, nos encontramos con un sistema de baremación de méritos más discriminatorio, puesto que las diferencias entre las puntuaciones obtenidas en función del género son mayores, sobre todo, en lo que concierne a la experiencia y en el puesto directivo, lo que concuerda con la teoría del techo de cristal (Bastida, 2008; Hoobler, Wayne y Lemmon, 2009; Ng y Wyesner, 2007), que hace referencia a las dificultades que encuentran las mujeres para acceder a puestos directivos. Sería conveniente continuar esta línea de investigación, para así poder llegar a conclusiones más firmes.

Discusión General

La valoración de méritos es uno de los instrumentos más utilizados en las administraciones públicas en particular, pero también en las empresas para la toma de decisiones en selección de personal. Esto ocurre tanto en los EEUU como en los países europeos. En nuestro país su uso cuenta con una larga tradición y con un amplio respaldo de la ley. Durante muchos años fue uno de los dos únicos métodos de selección que la legislación contemplaba para la provisión de funcionarios públicos. El reciente Estatuto Básico del Empleado Público ha limitado su empleo en la selección de funcionarios al de una prueba complementaria, aunque sigue estando plenamente vigente para la contratación de personal laboral fijo. La relevancia social de este método y las implicaciones que conlleva para el buen funcionamiento de las administraciones públicas hacen necesario un análisis pormenorizado de sus características psicométricas y de otros aspectos derivados de su uso.

A tenor de lo observado en los dos estudios presentados, la valoración de méritos presenta una buena fiabilidad examinada ésta en términos de acuerdo entre valoradores y de estabilidad temporal, con coeficientes alrededor de .90. No obstante, debe tenerse en cuenta

también que esta elevada magnitud puede deberse en alguna medida a las características propias de nuestro estudio, en el cual se incidió especialmente en la preparación previa de los evaluadores y los baremos. Este acuerdo entre valoradores y esta estabilidad de las puntuaciones contrasta, sin embargo, con la relativamente escasa consistencia interna hallada, que puede llegar, incluso, a resultar nula en algunos casos. Por todo ello, los usuarios deberían tener presente que la valoración de méritos, cuando los requisitos de la plaza y los criterios de valoración no están bien establecidos, son muy genéricos o ambiguos, puede conducir a procesos selectivos nada fiables. A este respecto, es preciso recordar que un instrumento con una fiabilidad baja o nula implica una arbitrariedad o falta de objetividad en la medida que invalidaría el proceso selectivo.

Por otra parte, los análisis realizados para examinar la validez de constructo sugieren que la relación entre experiencia y formación puede ser pequeña o incluso negativa y, como conjetura, podría ser incluso menor (o más negativa) si se controlase la influencia de la edad de los sujetos en la relación. Esto tiene importantes implicaciones desde el punto de vista práctico, ya que la puntuación final de dos personas podría ser semejante a pesar de originarse con una cualificación y competencia profesional diferente, al producirse un proceso compensatorio entre experiencia y formación. En este sentido, para paliar en cierta medida los efectos compensatorios a los que estamos haciendo referencia, podrían introducirse puntos de corte mínimos para cada uno de los elementos del instrumento de valoración con antelación a la obtención de la puntuación final, de tal modo que aquellas personas que no hubiesen superado dichos puntos de corte no superarían el proceso selectivo.

Otro aspecto relevante que ha sido examinado en la presente investigación es el relativo a la posibilidad de discriminación indirecta derivada del uso de la valoración de méritos. Uno de los puntales principales en los que se apoya el uso del concurso de méritos como método de selección es la supuesta justicia y objetividad de sus puntuaciones. Sin embargo, la aparente objetividad del método, al menos en los datos de que se dispone en este artículo, queda cuestionada si se tiene en cuenta la discriminación indirecta que se ha observado en el caso de las mujeres. Los resultados obtenidos en los dos estudios presentados indican que el colectivo femenino obtiene puntuaciones inferiores a los hombres en los tres puestos evaluados. Esta diferencia varía entre .29 y .59 desviaciones típicas sobre el resultado final, llegando a una diferencia de .72 para la valoración de la experiencia laboral en el caso del puesto directivo. La explicación de esta diferencia en las puntuaciones podría deberse a las dificultades para el acceso a determinados puestos que se ha encontrado la mujer por razones históricas, culturales, políticas y sociales; y a la imposibilidad, por lo tanto, de acumular experiencia en esos puestos. Un ejemplo de esto sería el

fenómeno conocido como “techo de cristal”, que hace referencia a la dificultad de las mujeres para acceder a puestos directivos (Bastida, García-Izquierdo, Moscoso y Ramos-Villagrasa, 2009; Kyrchmeyer, 2002; Stroh, Langlands y Simpson, 2004). Estaríamos, por lo tanto, ante un instrumento que, de generalizarse los resultados de nuestro estudio, contribuye a la perpetuación de la discriminación por razones de género. En relación con esta cuestión, es importante que los profesionales y usuarios de instrumentos de valoración de méritos que quieran usar la recomendación sugerida más arriba relativa a incorporar puntos de corte mínimos, sean conscientes que tal recomendación se aplicaría sólo en contexto de no discriminación de género, porque si existe evidencia de su existencia, la incorporación de puntos de corte mínimos podría incrementar la discriminación indirecta de las mujeres.

Teniendo en cuenta la escasa validez predictiva de este método y los resultados anteriormente expuestos, sería recomendable una cierta de cautela en su uso por parte de las todas las organizaciones, pero especialmente por administraciones públicas. En este sentido los resultados apoyan un uso muy restringido de la baremación de méritos, supeditado a la resolución de los problemas de fiabilidad, validez y discriminación encontrados.

Implicaciones para la práctica

Habida cuenta lo expuesto anteriormente y de cara a señalar algunas implicaciones para la práctica profesional, creemos que sería recomendable seguir una serie de pautas para su uso: (a) no es conveniente usar una valoración de méritos sin haber realizado previamente un análisis del puesto de trabajo, que permita relacionar los criterios a evaluar con el contenido de las tareas y funciones del puesto; (b) es necesario asumir que la valoración de méritos es un instrumento de poca utilidad para la predicción del desempeño, por lo que nunca debería ser utilizado en solitario, sino en conjunto con otras pruebas e instrumentos que garanticen una validez mayor; como la Entrevista Conductual Estructurada (Choragwicka y Moscoso, 2008; Salgado y Moscoso, 2002; Sáez, 2007), los tests cognitivos (Salgado y Anderson, 2003a; Salgado, Anderson, Moscoso, Bertua, De Fruyt y Rolland, 2003b) y otros (véase Tabla 9); (c) Los profesionales y responsables de la aplicación de una valoración de méritos deberían examinar los resultados de dicha aplicación en relación con su fiabilidad y posible discriminación, antes de comunicarlos públicamente y utilizarlos en la toma decisiones, ya que como se ha visto son susceptibles de provocar discriminación y muestran una ausencia total de fiabilidad; (d) por último, es recomendable sustituir o complementar los actuales instrumentos de valoración de méritos por otros más efectivos como pudieran ser los métodos de consistencia conductual (Schmidt et. al. 1979).

Limitaciones del estudio

Nuestra investigación presenta algunas limitaciones que deben ser tenidas en cuenta para interpretar adecuadamente los resultados. Una primera limitación que presenta nuestro estudio tiene que ver con el reducido número de ocupaciones que han sido evaluadas. Las conclusiones mostradas en este artículo son resultado del análisis de sólo tres puestos de trabajo (director, técnico y profesor), por lo que cualquier generalización a otras ocupaciones debe ser efectuada con cautela. Por otro lado, el tamaño de las muestras también es reducido (72, 42 y 98 respectivamente para cada puesto analizado), con lo que, una vez más, las generalizaciones resultarían demasiado arriesgadas. Por ello, y dada la importancia que tiene el objeto de estudio de este artículo y las implicaciones que tiene para las administraciones públicas y para la sociedad, sería recomendable la realización de más estudios que amplíen y completen los resultados aquí mostrados. Entre las múltiples posibles investigaciones a realizar, sugerimos el estudio de otros sistemas de valoración de méritos, que puedan superar las limitaciones que posee el actual sistema. Por otro lado, sería interesante investigar si la discriminación por razón de género detectada en este estudio es generalizable a otros puestos de trabajo.

En resumen, la valoración de méritos es un instrumento ampliamente utilizado para la selección de personal, especialmente en las administraciones públicas de todo el mundo. Tanto los usuarios como las organizaciones sindicales tienden a percibir este instrumento como objetivo, fiable, válido y no discriminatorio. Sin embargo, los resultados de la presente investigación indican que este instrumento puede no ser fiable y válido, además de ser relativamente discriminatorio contra las mujeres. Por ello se sugiere un uso prudente del mismo y más investigaciones que permitan un posicionamiento más firme sobre su uso.

Tabla 9. Fiabilidad y validez de los procedimientos de selección de personal

Método	Fiabilidad	Validez Operativa
Entrevista Conductual Estructurada	.83	.63
Exámenes (Tests) de Conocimientos	.80	.45
Assessment Center – Simulaciones	.70	.37
Entrevistas Convencionales Estructuradas	.65	.33
Referencias Personales	.60	.26
Valoración de Méritos y Formación	.80	.18
Entrevistas No Estructuradas	.50	.14
Capacidad Cognitiva General	.83	.71
Razonamiento Matemático	.85	.52
Razonamiento Espacial-Mecánico	.77	.51
Capacidad Perceptiva	.67	.52
Memoria	.77	.56
Personalidad – Conciencia	.80	.30

Nota. Validez Operativa= validez observada corregida por fiabilidad de criterio y restricción en el rango. (Basado en Salgado y Moscoso 2008).

Extended Summary

The rating of the merits of candidates, based on the score of experience, education and training, is one of the methods used by public administrations, and to a lesser extent by firms, in their recruitment and promotion processes. The merit rating is included within the category of selection methods known in Anglo-Saxon literature as "Training and Experience Evaluations" (T & E) (Gatewood & Field, 1998). The T & E consists of evaluating the information provided by the applicant to determine their capabilities for the job. Such information is often provided in written form, either through forms, certificates or accreditations. An assessor shall assess the information provided from a previously designed set of criteria. The long tradition in the use of merit rating systems seems to be based on three assumptions: (1) the method is cheaper, faster and considered "fair" by candidates and unions, (2) it is a valid method, i.e. the rating scores relate largely to performance in the workplace, (3) it is a completely reliable method, as there would be no measurement error in the scores assigned to candidates.

However, this method has a number of important drawbacks. First, the use of ratings is not as simple as it appears to be, if it is not well defined, it could lead to different interpretations by the evaluators. Another aspect to consider is how the candidates present their documentation as it may pose a problem for the evaluator. As to the objectivity of method, there is ample room for discretion of the evaluator, since in most cases the rating scales are set out in a very general way, which can lend itself to multiple interpretations and are difficult to adapt to each single candidate to be evaluated. Another aspect to consider is the setting of the ratings, they could only be determined in an objective manner based on a prior job analysis to know what the necessary characteristics are for the job incumbent. However, in most cases, this preliminary analysis does not exist (Villasante & Bretones, 1995). Another important drawback, mentioned by American literature, is the possibility that merit ratings produce adverse impact on racial minorities, women or young people. This implies that the ratings (competition) of merit can contribute to the perpetuation of discriminatory situations.

With regard to the criterion validity of merit ratings, McDaniel, Schmidt and Hunter (1988), in their comprehensive meta-analysis of training and experience, found a validity coefficient corrected for range restriction and unreliability in the criterion of .11 ($N = 6741$). The conclusion reached from these data is that this method has low validity, which also is not generalizable. In connection with the reliability of the method, it is often assumed that it is entirely reliable, and few studies have been published on the reliability of assessments of experience and training in general and on merit ratings in particular, so that needs more investi-

gation. The available evidence is contrary to the case of extreme reliability of the method. For example, Ash and Levine (1985) found an average inter-rater reliability of .83. However, the available evidence does not allow us to indicate what the range of internal consistency coefficients (e.g. Cronbach's alpha) of the merit ratings may be, nor what the temporal stability of scores is. Taking all of this into account, the aim of this paper is threefold: (1) to examine the reliability of the merit ratings with regard to the internal consistency, temporal stability and the agreement between raters, (2) to examine the construct validity of the merit ratings, and (3) to examine whether the merit ratings are impartial and fair and indirect discrimination does not occur for either gender.

STUDY 1

Method

For the realization of this study 72 applicants for the position of language teacher were evaluated in a partner organization to a Spanish public university, of which 14 were males and 58 females. The selection method used was merit rating, based on experience, education and training related to the teaching of language. Before making the ratings, the evaluators conducted a pilot study to clarify the evaluation criteria. In order to know the temporal stability of scores and the degree of agreement between judges, a second evaluation was conducted one month after the first. The second time, each rater assessed different nominations than the previous occasion.

Results

The results of the reliability analysis indicate a relatively low internal consistency ($\alpha = .55$), an equivalence between the parts (Spearman-Brown) in the limits of what would be acceptable (.73) and high temporal stability (.93). Furthermore, the low correlations found between different parts of the instrument suggest that merit ratings are a multidimensional variable. Regarding the adverse impact, it is observed that, on average, women scored .29 standard deviations less than men. The assessment of merit is not, therefore, totally fair to women because there is some indirect discrimination as a result of the instrument used.

Discussion

The results of this study indicate that the total score of the instrument is not severely affected by changes of

rater or the passage of time. However, the merit rating shows a significant weakness: neither is it consistent nor is there a strong relationship between the elements, making it susceptible to problems of selection decision. Finally, the data indicate that, contrary to the widespread belief that the assessment of merit is objective and fair, the score discriminates indirectly against women, which could give rise to discrimination claims for breaking the standards of the Organic Law 3 / 2007 of 22 March for the effective equality between women and men.

STUDY 2

Method

Two samples of applicants for two different jobs were used in this study. The first sample consisted of 48 applicants for the job of technical director, 24 men and 18 women. The second sample consisted of 98 applicants for the position of technician, 49 candidates of each sex. The method used for this selection process was the rating of merits. It valued experience, training and computer literacy.

Result

In the case of technical director, we find low values for the coefficients of reliability, specifically Cronbach's alpha of .41 and an indicator of equivalence (Spearman-Brown) of .44. For the technician, we find aberrant results, for both the coefficients of internal consistency and equivalence, resulting in a negative reliability coefficient and a greater than 1 coefficient. This means that the ratings of merit used in this case show no reliability. Furthermore, the low correlations found between different parts of the instrument indicate a lack of construct validity. We must highlight the presence of negative correlations, indicating that scores on different sections are offset by one another, complicating the hiring decision. Another aspect analyzed is the indirect discrimination based on gender. The effect size for the job of technical director is $-.47$ standard deviations. This figure indicates a moderate discrimination against women. In the case of experience rating, the figure increases to $-.72$. For the job of technician, the effect size is $-.59$ standard deviations.

Discussion

In view of these results, we can conclude that there are psychometric weaknesses in the reliability of the instrument and the construct validity, which calls into question the feasibility of this method as a screening instrument. The analysis of effect size of gender scores

indicates that merit ratings were discriminatory to women, especially in management positions, which is consistent with the theory of the glass ceiling.

General discussion

The social relevance of this method and the implications for the proper functioning of the Public Administration require a detailed analysis of the psychometric characteristics and other issues arising from their use. According to the findings of the two studies carried out, the merit ratings show good inter-rater reliability and temporal stability, with coefficients around .90. It should be noticed that this high magnitude of the coefficient may be due, in part, to the characteristics of our study, in which was particularly affected by the priming of assessors and scales. This contrasts, however, with the low internal consistency and low construct validity found. Therefore, users should keep in mind that when endpoints are not well established the assessment of merit in selective processes may lead to unreliable coefficients.

One of the main pillars which support the use of merits is its supposed objectivity. However, the results obtained in the two studies presented indicate that scores obtained by women are smaller than those of men in all three positions tested, reaching a difference of .72 standard deviations for the assessment of work experience for a management job. The explanation for this difference in scores could be due to difficulties in access to certain jobs that women have had for historical, cultural, political and social reasons, and the impossibility, therefore, of gaining experience in those positions. We would have, therefore, an instrument that, to generalize the results of our study, contributes to the perpetuation of gender discrimination. Given the low predictive validity of this method and the results outlined above, we would recommend a degree of caution in its use by all organizations, but especially by the Public Administration. In this sense, the results support a very restricted use of merit ratings, subject to resolution of problems of reliability, validity and unfairness found.

Referencias

- Alcaide, A. M. (2008). El acceso a la dirección en el sistema educativo español. Dificultades para la definición de un modelo. *Revista de Educación*, 347, 275-298.
- Ash, R. A., Johnson, J. C., Levine, E. L. y McDaniel, M. A. (1989). Job applicant training and work experience evaluation in personnel selection. *Research in personnel and human resources management*, 7, 183-226.
- Ash, R. A. y Levine, E. L. (1981). *An investigation of*

- career service unassembled examinations* (Informe final de la STAR grant No. 80-011). Tampa, FL: Center for evaluation Research, University of South Florida.
- Ash, R. A. y Levine, E. L. (1985). Job applicant training and work experience evaluation: An empirical comparison of four methods. *Journal of Applied Psychology, 70*, 572-576.
- Ash, R. A. (1983). The Behavioral Consistency Method of Training and Experience Evaluation: Content Validity Issues and Completion Rate Problems. *Public Personnel Management, 12*, 115.
- Ash, R. A. (1986). The Activity/Achievement Indicator: A Possible Alternative to the Behavioral Consistency Method of Training and Experience Evaluation. *Public Personnel Management, 15*, 325.
- Bastida, M. (2008). *La mujer en puestos directivos*. Ponencia presentada en las I Jornadas Muller e Igualdade: A perspectiva de xénero na xestión dos recursos humanos en organizacións públicas e privadas. Santiago de Compostela 19, 20 y 21 de Marzo.
- Bastida, M., García-Izquierdo, A., Moscoso, S. y Ramos-Villagrasa, P. (2009). *Attitudes toward affirmative action programs for women in Spanish organizations*. Comunicación presentada en el 14º European Congress of Work and Organizational Psychology. Santiago de Compostela, 13-16 de Mayo.
- Baughner, D. y Varanell Jr, A. (1994). Ten years of experience with a performance-based promotional selection and career development system within state government. *Public Personnel Management, 23*, 551.
- Buela-Casal, G. (2007). Consideraciones metodológicas sobre el procedimiento de acreditación y del concurso de acceso a cuerpos de funcionarios docentes universitarios. *Revista Electrónica de Metodología Aplicada, 12*, 1-14.
- Choragwicka, B. y Moscoso, S. (2007). Validez de contenido de una Entrevista Conductual Estructurada. *Revista de Psicología del Trabajo y las Organizaciones, 23*, 75-92.
- Comisión para el Estudio y Preparación del Estatuto Básico del Empleado Público (2005). *Estatuto Básico del Empleado Público. Informe de la comisión*. Madrid: INAP.
- 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.
- Gatewood, R.D. y Feild, H.S. (1998). *Human resource selection*. 4ª ed. Forth Worth: Dryden Press.
- Hoobler, J. M., Wayne, S. J. y Lemmon, G. (2009). Bosses' perceptions of family-work conflict and women's promotability: glass ceiling effects. *Academy of Management Journal, 52*, 939-957.
- International Public Management Association for Human Resources (IMPA-HR) (2006). *Recruitment and Selection Benchmarking*. Alexandria, VA: IMPA-HR. [accessible electrónicamente en la dirección de Internet <http://www.ipma-hr.org/pdf/final5.pdf>]
- Kirchmeyer (2002). Gender differences in managerial careers: Yesterday, Today, and Tomorrow. *Journal of Business Ethics, 37*, 5-24.
- Levine, E. L., Ash, R. A. y Levine, J. D. (2004). Judgmental assessment of job-related experience, training, and education for use in human resource staffing. En Thomas, J. C. y Hersen, M. (Eds.). *Comprehensive handbook of psychological assessment, Vol. 4: Industrial and organizational assessment*. (pp. 269-296). Hoboken, NJ, US: John Wiley & Sons Inc.
- Ley Orgánica 3/2007, de 22 de marzo, para *La Igualdad Efectiva de Mujeres y Hombres*.
- Ley 7/2007, de 12 de abril, del *Estatuto Básico del Empleado Público*.
- Lowry, P. E. (1994). Selection methods; Comparison of assessment centers with personnel records evaluations. *Public Personnel Management, 23*, 383.
- McDaniel, M. A., Schmidt, F. L. y Hunter, J. E. (1988a). A meta-analysis of the validity of methods for rating training and experience in personnel selection. *Personnel Psychology, 41*, 283-309.
- McGonigle, T. P. y Curnow, C. C. (2002). Development of a modified improved point method experience questionnaire. *Applied HRM Research, 7*, 12-21.
- McGonigle, T. P. y Curnow, C. K. (2007). Measures of Training and Experience. En Wheaton, G. R. (2007). *Applied measurement: Industrial Psychology in Human Resources Management*. Routledge.
- Ng, E. S. y Wiesner, W. H. (2007). Are men always picked over women? The effects of employment equity directives on selection decisions. *Journal of Business Ethics, 76*, 177-187.
- Porter, W. R., Levine, E. L. y Flory, A. III. (1976). *Training and experience evaluation. A practical handbook for evaluating job applications, resumes, and other applicant data*. Tempe, AZ: Personnel Services Organization.
- 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 las Organizaciones, 23*, 299-324.
- Salgado, J. F. y Anderson, N. (2003a). Validity generalization of GMA test across the European Community Countries. *European Journal of Work and Organizational Psychology, 12*, 1-17.
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., De Fruyt, F. y Rolland, J.P. (2003b). A meta-analytic study of GMA validity for different occupations in the European Community. *Journal of Applied Psychology, 88*, 1068-1081.

- 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.
- Salgado, J. F. y Moscoso, S. (2008). Selección de personal en la empresa y las administraciones públicas: de la visión tradicional a la visión estratégica. *Papeles del Psicólogo, 29*, 16-24.
- Schmidt, F. L., Caplan, J. R., Bemis, S. E., Decuir, R., Dunn, L. y Antone, L. (1979). *The behavioral consistency method of unassembled examining*. Washington, DC: US Office of Personnel Management.
- Steiner, D. D. y Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology, 55*, 404-406.
- Stroh, L.K., Langlands, C. L. y Simpson, P. A. (2004). Shattering the glass ceiling in the new millenium. En Stockdale, M. S. y Crosby, F. J. (Eds.). *The psychology and management of workplace diversity* (pp. 147-167). MA, US: Blackwell Publishing.
- Villasante, C. S. y Bretones, F. D. (1995). Selección, formación y desarrollo de carreras en la administración. En Rodríguez Fernández, A. (dir.) *Los Recursos Humanos en las Administraciones Públicas* (pp. 249-288). Madrid: Ediciones Tecnos.

Manuscrito Recibido: 07/09/2009

Revisión Recibida: 25/10/2009

Aceptado: 05/11/2009