



Artículo

Validez de contenido de una Entrevista Conductual Estructurada*

Content validity of a Structured Behavioral Interview

BEATA CHORAGWICKA¹
SILVIA MOSCOSO¹

Fecha de Recepción: 10-10-2006

Fecha de Aceptación: 19-02-2007

RESUMEN

Este trabajo presenta un examen empírico de la validez de contenido de una Entrevista Conductual Estructurada (ECE), realizado a partir de las valoraciones de un panel de 11 expertos, utilizando el método sugerido por Lawshe (1975). Se encontró la Razón de Validez de Contenido (RVC) significativa para 4 de las 14 preguntas y 4 de las 10 dimensiones. El Índice de Validez de Contenido (IVC) hallado para las dimensiones significativas era de .86 y para las preguntas significativas .68. También se estimó la fiabilidad del instrumento mediante un coeficiente de acuerdo entre 6 ó 12 evaluadores, obteniendo un valor promedio de .78, semejante al resultado .75 proveniente de los meta-análisis de Conway, Jako y Goodman (1995) y de Salgado y Moscoso (1995).

ABSTRACT

This paper reports on an empirical study of the content validity of a Structured Behavioural Interview (SBI). The study was based on 11 experts' ratings, using a method proposed by Lawshe (1975). A significant Content Validity Ratio (CVR) was found for 4 out of 14 questions and for 4 out of 10 general dimensions. A Content Validity Index (CVI) of .86 was obtained for the significant dimensions as a whole and of .68 for the questions.

* La investigación que se incluye este artículo ha sido parcialmente subvencionada por el Ministerio de Educación y Ciencia con cargo al proyecto SEJ2005-03842/PSIC en el que figura como investigadora Silvia Moscoso. Toda la correspondencia referida al contenido del presente trabajo debe dirigirse a Beata Choragwicka (beciach@usc.es) o bien a Silvia Moscoso (pssilvia@usc.es), Departamento de Psicología, Universidad Santiago de Compostela, Campus Sur, 15872 Santiago, A Coruña, España.

¹ Universidad de Santiago de Compostela.

Validez de contenido de una Entrevista Conductual Estructurada

Additionally, reliability was estimated by using the coefficient of agreement between raters (6 or 12) and the result was .78, similar to the .75 value obtained in meta-analyses by Conway, Jako and Goodman (1995) and by Salgado and Moscoso (1995).

PALABRAS CLAVE

Entrevista conductual estructurada, Validez de contenido, Propiedades psicométricas.

KEY WORDS

Structured behavioral interview, Content validity, Psychometric properties.

Varios estudios han confirmado que, independientemente del país del que se trate, las entrevistas son un núcleo central de cualquier proceso selectivo, siendo muchas veces la única base para tomar las decisiones de contratación. Su flexibilidad como método y su relativamente bajo coste de ejecución, comparando con otras técnicas de selección, permite su amplio uso: se estima que, dependiendo de los autores del estudio, el 90% de empresas europeas y entre el 80 y 99% de empresas americanas utilizan este instrumento en los procesos de selección [Harris (1989); Salgado y Moscoso (2001); Moscoso y Salgado (2001); McDaniel et. al (1994)]. Desde este punto de vista resulta sorprendente que los datos acumulados a través de numerosas investigaciones tanto cuantitativas (meta-análisis) como cualitativas sobre las propiedades psicométricas de las entrevistas de selección convencionales muestran que éstas poseen una fiabilidad de moderada a baja y una validez de criterio muy pequeña (Wagner, 1949; Mayfield, 1964; Urlich y Trumbo, 1965; Wright, 1969; Schmitt, 1976; Arvey y Campion, 1982; Reilly y Chao, 1982; y Hunter y Hunter, 1984).

Aunque pudiera tener un base real lo anteriormente dicho, hoy en día existe una gran bibliografía que indiscutiblemente demuestra que las propiedades psicométricas de la entrevista aumentan cuanto más estructurado es su formato (Wiesner y Cronshaw, 1988; Harris, 1989; Anderson 1992; Mc Daniel et al., 1994; Huffcutt y Arthur, 1994, Salgado y Moscoso, 1995; Campion, Palmer y Campion, 1997; Salgado, Ones y Viswesvaran, 2001). Tanto los estudios meta-analíticos como los estudios primarios ponen de manifiesto que las entrevistas pueden alcanzar unos niveles de validez y fiabilidad similares a los de los mejores instrumentos selectivos

(tests de habilidades cognitivas y tests de muestras de trabajo), siempre y cuando su formato convencional (es decir no estructurado) se sustituya por el formato de las entrevistas conductuales estructuradas (ECE) en las que se controlan e incluso eliminan varios de los sesgos existentes cuando un proceso de selección se basa en las entrevistas convencionales (Salgado, 1995).

La ECE es un modelo de entrevista relativamente reciente ya que se desarrolló a lo largo de los años 80 del siglo XX. Bajo la denominación de las ECE están incluidas la “entrevista de descripción de conducta” (Janz, 1977, 1982, 1989), la “entrevista desarrollada a partir de un análisis del puesto” (Campion, Pursell y Brown, 1988; Wright, Lichtenfels y Pursell, 1989), la “entrevista situacional” (Latham, Saari, Pursell, y Campion, 1980; Latham, 1989), la “entrevista multimodal” (Schüler y Funke, 1989), “entrevista conductual” (Green y Hogan, 1882), y la “entrevista conductual estructurada” (Motowidlo et al., 1992). Todas ellas unen una serie de condiciones, que permite que su denominación genérica sea igual (Salgado y Moscoso, 2001): (a) Las preguntas de la entrevista parten de un análisis de puestos; (b) a todos los candidatos se les hacen las mismas preguntas; (c) se utilizan las escalas de valoración basadas en la conducta; (d) Se realiza el mismo proceso con todos los entrevistados; (e) los entrevistadores son previamente formados sobre la ECE; (f) la decisión sobre la contratación se toma una vez terminadas todas las entrevistas.

A lo largo de los años, se han llevado a cabo múltiples estudios primarios y meta-analíticos que investigaban el tema de validez y de fiabilidad de las ECE, llegan-

do a conclusiones diferentes que en caso de los estudios dedicados a la entrevista convencional (Wagner, 1949; Mayfield, 1964; Ulrich y Trumbo, 1965; Wright, 1969; Schmitt, 1976; Arvey y Campion, 1982; Reilly y Chao, 1982; y Hunter y Hunter, 1984).

Con respecto a la validez, en 1988 Wiesner y Cronshaw en una investigación que incluía no solamente los estudios norteamericanos sino también de otros países (Australia, Alemania, Francia, Holanda, Inglaterra, Israel, Suecia), encontraron un valor de validez de criterio de la entrevista igual a .26 (no corregido) y .47 (corregido) (N=51.459). Descubrieron también que la validez de entrevista depende en gran medida del nivel de estructuración de la misma; la validez promedio corregida de las entrevistas individuales no estructuradas fue de .20, para las entrevistas de panel no estructuradas de .37 y para las entrevistas estructuradas individuales y de panel .63 y .60 respectivamente. A pesar de las limitaciones de este meta-análisis, cabe enfatizar que fue el primer estudio que de algún modo demostró que las entrevistas de selección podían ser un buen predictor del rendimiento. En el siguiente meta-análisis sobre la entrevista de empleo, McDaniel, Whetzel, Schmidt y Maurer (1994), al dividir las entrevistas en estructuradas y no estructuradas, hallaron que las primeras mostraban una validez corregida de .44 y las segundas de .33. El mismo año Huffcutt y Arthur llevaron a cabo un estudio meta-analítico (N=18.652) donde clasificaron las entrevistas según dos criterios: (a) la de estandarización de preguntas (4 niveles progresivos) y (b) la estandarización de las valoraciones de las respuestas (3 niveles progresivos). El coeficiente de la validez global de la entrevista corregido por restric-

ción del rango y error de medida en el criterio fue de .37, y para los 4 niveles de estandarización fue de .20, .35, .56, .57 (de 1 a 4 respectivamente). Salgado y Moscoso (1995) en su investigación meta-analítica sobre la ECE, analizaron los estudios realizados entre 1980 y 1994 (N=2.121) y encontraron una validez corregida de .68. Los autores concluyeron también que la ECE tiene una validez similar a la de los mejores métodos selectivos, siempre y cuando se respeten los pasos de un desarrollo correcto (en particular el análisis de puesto) y afirmaron que esta validez es generalizable a cualquier tipo de ocupación. Más recientemente, Schmidt y Rader (1999) realizaron un meta-análisis sobre la validez de la entrevista estructurada telefónica y alcanzaron una validez corregida de .40 para el criterio de rendimiento.

Todos los estudios mencionados indican que los valores de validez que pueda alcanzar la entrevista de selección en su formato estructurado son próximas a los valores de validez de los mejores métodos selectivos: los tests de habilidades cognitivas (.51) y los tests de muestras de trabajo (.54). No obstante cabe resaltar que la mayor parte de la investigación sobre la validez de la entrevista de selección, y de su versión estructurada, se ha centrado en la validez de criterio, pues desde el punto de vista de la selección de personal, es la evidencia de validez que suele interesar más. Sin embargo, los otros tipos de validez también son de gran relevancia, principalmente teniendo en cuenta que la validez es un concepto unitario, pero con diferentes fuentes para aportar su evidencia (SIOP, 2003). Cada una de estas fuentes, tanto la validez de contenido como la de criterio y la de constructo nos aportan una información útil que, dependiendo de los objetivos de estudio, puede ser más o

menos eficaz para confirmar la relación del instrumento de selección con el ámbito laboral en cuestión. En caso de la entrevista, un método de selección basado en las valoraciones subjetivas, también la validación del contenido parece ser de gran relevancia.

La evidencia de la validez de contenido consiste en la mayoría de los casos en demostrar un fuerte vínculo entre el contenido del procedimiento de selección y los requisitos del puesto o las habilidades y destrezas necesarias para un buen desempeño en el mismo. Este procedimiento aporta una base lógica para inferir los resultados en los comportamientos organizacionales de interés. El éxito del estudio basado en la validez de contenido está ligado a las valoraciones aportadas por los expertos en la materia (SIOP, 2003). Lawshe en su trabajo "Quantitative approach to content validity" (1975) desarrolló un modelo de cuantificación para relacionar el contenido de un instrumento de selección y el desempeño. El autor denomina a su método el Panel de Evaluación del Contenido y consiste en la utilización de un grupo de expertos con vasto conocimiento en el puesto de interés, los cuales deben evaluar de manera separada todos los ítems que contiene el instrumento. A continuación se calcula la fórmula llamada Razón de Validez de Contenido (RVC) que sirve para determinar cuales de las preguntas (dimensiones) del instrumento analizado son significativas y deben de ser retenidas en la versión final del mismo. La mínima RVC aceptable varía según el número de los miembros del grupo de expertos. Al final del procedimiento se calcula el Índice de Validez de Contenido (IVC) para el total de la herramienta que es un promedio de la validez de contenido de todas las preguntas (ítems) significativas.

Cabe mencionar que no es posible aplicar el método de Lawshe (1975) a las entrevistas convencionales, ya que éstas a menudo no poseen un único contenido, sus preguntas varían de candidato a candidato y, frecuentemente, están desvinculadas por completo del ámbito laboral (p. ej. preguntas sobre aficiones, opiniones o biográficas). Por esa razón no se puede medir la relación que tienen estas preguntas con el contenido del puesto en cuestión, aunque sí que parece posible utilizar la RVC en lo que concierne a las preguntas de las ECE que, partiendo de un análisis de puesto, se refieren a las conductas laborales específicas. Pero incluso así, no hay muchos trabajos previos sobre la validez de contenido de las entrevistas estructuradas. En nuestro conocimiento, en la actualidad existen solamente tres estudios publicados en los que se utilizó esta evidencia de validez aplicándola a la entrevista de selección. En el primer estudio, Schmitt y Ostroff (1986) desarrollaron un procedimiento, del cual formaba parte una entrevista situacional, para seleccionar operadores de un teléfono de urgencias. La entrevista desarrollada contenía 17 preguntas y evaluaba cuatro dimensiones principales: habilidades comunicativas, control emocional, juicio y compañerismo. Ocho expertos valoraron cada una de las preguntas y las RVC medias calculadas para cada juez varían desde .53 a .82. Aunque la entrevista se basaba en el análisis de puesto, solamente dos expertos evaluaron más de la mitad de las preguntas como esenciales o necesarias y las RVC provenientes de los restantes miembros de panel fueron negativos. En otro de los estudios mencionados, Carrier, Dalessio y Brown (1990) investigaron la relación entre las inferencias basadas en la validez de contenido y las basadas en la validez de criterio en un proceso selectivo para un

puesto de representante de compañía de seguros. Las aptitudes de los candidatos se verificaban mediante tres versiones de entrevista situacional. Como criterio se utilizaron la cantidad de comisiones obtenidas y la permanencia en el puesto. Las RVC para cada pregunta fueron estimadas partiendo de las valoraciones de 300 superiores. La RVC mínima aceptable para este número de las valoraciones es de .15, aunque los valores hallados para tres tipos de las entrevistas utilizadas fueron .03, -.21, -.22, respectivamente. El tercer estudio conocido fue llevado a cabo por Moscoso y Salgado (2001). Los autores calcularon la validez de contenido de una entrevista de descripción de conducta, desarrollada para un puesto de vigilante de seguridad. El instrumento contenía 7 preguntas y cada una de ellas fue valorada por un panel compuesto por 11 expertos. La RVC hallada para la mayoría de preguntas fue significativa y obtuvieron un IVC de .89.

En cuanto a la fiabilidad de las ECE existen varios estudios primarios cuyo propósito era analizar esta propiedad investigando algunos tipos de entrevistas estructuradas. No obstante, hasta hace poco la bibliografía más numerosa concernía las entrevistas situacionales. Latham, Saari, Pursell y Champion (1980) encontraron un coeficiente de fiabilidad interevaluador de .76 (N=49) y .79 (N=63) utilizando tres entrevistadores. Champion, Pursell y Brown (1988) obtuvieron un resultado de .88 (N=149), utilizando igualmente tres entrevistadores. Por otro lado, la primera investigación importante que confirmaría alta fiabilidad de la entrevista de la descripción de conducta fue el metaanálisis de Conway, Jako y Goodman (1995) donde el valor promedio de la fiabilidad interevaluador encontrado fue de .70 y cuando se utilizaban las preguntas con alto

nivel de estandarización (a todos los candidatos se les hace las mismas preguntas, no obstante el entrevistador tiene cierta libertad para adaptar las preguntas y sondear al candidato cuando lo considere oportuno) el valor de la fiabilidad fue de .75 (N=12.549). Dichos autores aportaron también los coeficientes alfa de las valoraciones de la entrevista, es decir de la fiabilidad por consistencia interna que fueron de .67 para las entrevistas altamente estructuradas, .56 para las entrevistas semiestructuradas y .34 para las entrevistas no estructuradas. Al mismo tiempo, Salgado y Moscoso (1995) analizando conjuntamente los datos disponibles sobre la fiabilidad interjuez de las ECE encontraron una fiabilidad promedio de .75, es decir, igual que en el meta-análisis de Conway, Jako y Goodman (1995). Más recientemente, Taylor y Small (2002) en una investigación meta-analítica encontraron un coeficiente de fiabilidad .79 para las entrevistas situacionales y .76 para las de descripción de conducta. Otro meta-análisis sobre la fiabilidad interjuez de la ECE es el de Salgado, Moscoso y Gorriti (2004) llevado a cabo en España, en la Administración General del País Vasco (AGPV). Dichos autores analizaron la fiabilidad de las entrevistas conductuales estructuradas utilizadas en el proceso de selección para los puestos de jefatura de la AGPV. Se investigaron las fiabilidades para un conjunto de 20 entrevistas desarrolladas para otros tantos puestos específicos del personal funcionario del Gobierno Vasco. Los paneles de expertos se componían de 3 ó 4 personas, entrevistadores con experiencia de los que uno siempre era un entrevistador homologado. Los coeficientes de fiabilidad de cada uno de los miembros de los paneles varían desde .41 hasta .96 y un valor promedio fue de .83. En cuanto a la fiabilidad de

distintos paneles, los coeficientes varían desde .73 a .93, con un promedio de .93.

Resumiendo, las ECE pueden aportar valores de validez y de fiabilidad muy elevados, siempre y cuando su desarrollo haya sido correcto. En cuanto a validez de las mismas cabe enfatizar que no siempre es posible aportar las evidencias para su validación de criterio, por lo que la validación de contenido puede ser una alternativa comprometedoras. No obstante la investigación al respecto es insuficiente y no aporta resultados contundentes, por lo que se necesitan más estudios dedicados a este tema.

El presente artículo contiene los resultados de dos estudios independientes pero relacionados, cuyo fin es determinar la validez de contenido y fiabilidad interjuez de la misma ECE desarrollada para seleccionar camareros. Los objetivos de este trabajo son averiguar: (a) en qué medida los valores de la validez de contenido de la ECE utilizada se aproximan al modelo de excelencia propuesto por Lawshe (1975) y (b) en qué medida la fiabilidad interjuez de la misma es cercana a los valores provenientes de los estudios meta-analíticos sobre la fiabilidad de las ECE (Conway, Jako y Goodman, 1995; Salgado y Moscoso, 1995). Nuestra hipótesis inicial es que la mayoría de preguntas y dimensiones alcanzarán la RVC significativa y que la fiabilidad promedio será alrededor de .75.

ESTUDIO 1

MÉTODO

Muestra

Once expertos en materia de hostelería, que incluían a supervisores directos de los

camareros de los restaurantes y hoteles más apreciados de Santiago de Compostela (Galicia), dueños de los restaurantes en Santiago, profesores del Centro Superior de Hostelería de Galicia y profesionales de la profesión, valoraron por separado las preguntas y las dimensiones de una ECE desarrollada para seleccionar camareros.

PROCEDIMIENTO

En el estudio se utilizó una entrevista de descripción de conducta de Janz (1977, 1982, 1989). El objetivo de esta modalidad de la ECE es conocer el comportamiento del candidato en situaciones pasadas, que se parezcan lo máximo posible a las situaciones a las cuales el candidato tendrá que enfrentarse en el puesto al que aspira. La entrevista utilizada fue desarrollada a partir de un número elevado (137) de incidentes críticos (véase Flanagan, 1954), respetando todos los pasos indicados por Janz (1977, 1982, 1989), con las modificaciones propuestas por Salgado y Moscoso (2001) y Moscoso y Salgado (2001). La versión original de la entrevista, aplicada al estudio de validación de contenido, la formaban 14 preguntas correspondientes a 10 dimensiones generales. A seis dimensiones les correspondía una pregunta (amabilidad, resistencia al estrés, solucionar problemas, "saber ganar clientes", paciencia, compañerismo) y a cuatro les correspondían dos preguntas (responsabilidad, atención, discreción, conocimientos del puesto). A cada pregunta principal, la seguían las preguntas de sondeo destinadas a profundizar el conocimiento del entrevistador en cuanto a la situación descrita por el candidato.

El cálculo de la validez de contenido se realizó a partir del método del Panel de

Evaluación de Contenido desarrollado por Lawshe (1975). A cada uno de los miembros se le proporcionó un ejemplar de la entrevista, la descripción de las dimensiones que ésta evalúa y una hoja de evaluación estandarizada. Los expertos dieron su puntuación en cada una de las dimensiones y también en cada una de las preguntas de la entrevista, respondiendo a la cuestión: "¿La característica (conocimiento, habilidad) que mide esta dimensión/pregunta le parece a Usted: (a) fundamental, (b) útil pero no fundamental, o (c) irrelevante para el desempeño adecuado en el puesto del camarero?".

A continuación se calculó la Razón de Validez de Contenido (RVC) para cada dimensión y para cada pregunta, y el Índice de Validez de Contenido (IVC) para el total de dimensiones y el total de las preguntas significativas de la entrevista.

RESULTADOS

En la Tabla 1 se pueden ver los resultados del procedimiento de validación de contenido de las dimensiones de la entrevista. Tal y como puede verse, ninguno de los miembros del panel consideró que alguna de las dimensiones de la entrevista era irrelevante y también puede verse que siempre más de la mitad de los miembros estaba de acuerdo en que cada una de las dimensiones es fundamental para un buen desempeño en el puesto de camarero, lo que indica que el instrumento ha sido creado exitosamente. En la Tabla 2 aparecen los resultados de las valoraciones de los jueces en cuanto a las preguntas de la entrevista. En este caso, todas las preguntas menos una (la pregunta 9.1) fueron valoradas por la mayoría de los expertos como fundamentales

para un rendimiento adecuado en el puesto. Si alguna está considerada irrelevante, es por un máximo de 3 de los 11 miembros del panel.

No obstante, cabe destacar que una RVC significativa la consiguieron solamente 4 de las 10 dimensiones y 4 de las 14 preguntas. Según el método de Lawshe, la RVC mínima para el Panel de Evaluación del Contenido compuesto por 11 expertos es de .59. Esta razón de validez de contenido resultó ser muy exigente, ya que no la alcanzaron incluso las dimensiones y preguntas sobre las cuales 8 de los 11 panelistas coincidían en que eran fundamentales para un desempeño adecuado en el puesto de camarero. Solamente cuando 9 de los 11 panelistas indicaron dicha dimensión (o la pregunta) como fundamental, el ítem valorado alcanzó la RVC significativa. Éste era el caso de las dimensiones: amabilidad (1), resistencia al estrés (.82), responsabilidad (.82) y atención (.82). Las preguntas evaluadas como fundamentales por nueve o más panelistas no siempre coincidieron con las dimensiones anteriormente mencionadas y fueron las preguntas 3.1 (solucionar problemas) RVC = .82; 5.1 (atención) RVC = .64, 6.1 (conocimientos del puesto) RVC = .64 y 8.1 (compañerismo) RVC = .64 (véase Tabla 2). El IVC para las dimensiones significativas es de .86 y para las preguntas significativas es de .68.

DISCUSIÓN

El estudio de la validez de contenido no confirmó completamente nuestra hipótesis ya que sólo una parte de las preguntas y dimensiones mostró la RVC significativa. Los resultados obtenidos se pueden considerar desde diferentes puntos de

Tabla 1.
Evaluaciones del Panel de Expertos por dimensiones

Dimensiones	Expertos											RVC
	1	2	3	4	5	6	7	8	9	10	11	
1. Amabilidad	F	F	F	F	F	F	F	F	F	F	F	1 ^a
2. Resistencia al estrés	F	F	F	F	F	F	F	Ú	F	F	F	.82 ^a
3. Solucionar problemas	F	Ú	Ú	F	F	Ú	F	Ú	Ú	F	F	.09
4. Responsabilidad	F	F	F	F	F	F	F	F	F	F	Ú	.82 ^a
5. Atención	F	F	F	F	F	F	F	F	Ú	F	F	.82 ^a
6. Conocimientos de puesto	Ú	F	F	F	F	Ú	F	F	Ú	F	F	.45
7. “Saber ganar los clientes”	F	Ú	F	F	F	Ú	F	F	F	F	Ú	.45
8. Compañerismo	F	Ú	F	F	F	Ú	F	Ú	F	F	F	.45
9. Discreción	F	F	Ú	F	F	Ú	F	Ú	F	F	F	.45
10. Paciencia	F	F	F	F	F	Ú	F	Ú	Ú	F	F	.45

Notas: RVC – La Razón de Validez de Contenido; F – Esta característica me parece fundamental; Ú – Esta característica me parece útil, pero no fundamental; I – Esta característica me parece irrelevante. El índice de Validez de Contenido (IVC) de las dimensiones significativas es de .86.

^a Dimensiones que han alcanzado la RVC mínima tal como lo indicó Lawshe (1975).

vista. Tal y como se puede apreciar en la Tabla 1, ninguno de los miembros del Panel de Evaluación del Contenido consideró alguna de las dimensiones evaluadas por la entrevista como irrelevante. Cabe subrayar también que solamente una de

las preguntas obtuvo la RVC negativa (véase Tabla 2), es decir, menos de la mitad de los expertos la consideró fundamental para un buen desempeño en el puesto del camarero (la pregunta 9.1). El resto de las preguntas (13 en total) han

Tabla 2.
Evaluación del Panel de Expertos por preguntas

Pregunta	Expertos											RVC
	1	2	3	4	5	6	7	8	9	10	11	
1.1	F	F	F	F	F	Ú	F	Ú	Ú	F	F	.45
2.1	F	F	F	F	F	F	F	Ú	I	Ú	F	.45
3.1	F	F	F	F	F	F	F	F	F	F	Ú	.82 ^a
4.1	F	I	Ú	F	F	F	F	Ú	Ú	I	F	.09
4.2	F	Ú	F	F	Ú	F	F	Ú	F	Ú	F	.27
5.1	F	F	F	F	F	Ú	F	F	F	Ú	F	.64 ^a
5.2	F	I	F	F	Ú	I	F	F	F	I	F	.27
6.1	F	F	F	F	F	Ú	F	F	F	Ú	F	.64 ^a
6.2	Ú	Ú	F	F	Ú	F	F	Ú	F	I	F	.09
7.1	F	Ú	F	F	F	Ú	F	F	F	Ú	Ú	.27
8.1	F	F	F	F	F	F	F	Ú	F	Ú	F	.64 ^a
9.1	F	F	Ú	F	Ú	I	F	Ú	Ú	Ú	F	-.09
9.2	F	F	F	F	F	F	F	Ú	I	I	F	.45
10.1	F	Ú	F	F	F	F	F	Ú	Ú	I	F	.27

Notas: RVC – La razón de Validez de Contenido; F – Esta característica me parece fundamental; Ú – Esta característica me parece útil, pero no fundamental; I – Esta característica me parece irrelevante. El índice de Validez de Contenido (IVC) de las preguntas significativas es de .68.

^a Preguntas que han alcanzado la RVC mínima tal como lo indicó Lawshe (1975).

sido valoradas como fundamentales para un desempeño adecuado en el puesto de camarero por la mayoría de los panelistas (mínimo por 6). Desde este punto de vista, el instrumento de selección que creamos es exitoso. Por otro lado, solamente

4 de las 10 dimensiones y 4 de las 14 preguntas alcanzaron la razón de validez de contenido significativa según el método desarrollado por Lawshe (1975). ¿Por qué ocurre eso con un instrumento basado en un número de incidentes críticos alto? Un

examen más detenido del método de Lawshe muestra que éste asigna el valor 1 a todos los ítems del instrumento valorados por los expertos como *fundamentales* y que los ítems considerados tanto *útiles* como *irrelevantes* tienen el mismo peso, es decir 0, lo que hace este procedimiento desmesuradamente estricto y que requiere un número de jueces muy alto para atenuar algo el grado de exigencia del método. A la vez, el grado de exigencia que el método de Lawshe impone a los procedimientos cuanto el número de jueces no es alto, nos lleva a sugerir la necesidad de la creación de un procedimiento alternativo al método de Lawshe en el que se podría asignar algún peso (menor que 1) a la respuesta *Este ítem me parece útil pero no fundamental*. Finalmente, para elegir las preguntas que formarían parte de la versión final de la entrevista hay que elegir uno de varios posibles caminos. Una de las opciones sería seguir estrictamente el procedimiento de Lawshe (1975) y dejar en la versión final del instrumento solamente las preguntas y dimensiones con la RVC significativa. No obstante este método no siempre es recomendable ya que, tal y como lo demuestran algunos de los estudios anteriores (Schmitt y Ostroff, 1986; Carrier et al., 1990), no en todos los casos es posible conseguir que un número de ítems elevado lo consiga. Otra posibilidad sería dejar todas las preguntas/dimensiones que tienen la RVC positiva, es decir, las que han sido valoradas como fundamentales por de la mitad de los panelistas. La siguiente de las posibilidades sería que el investigador estableciera a priori el número mínimo de jueces que deben considerar dicho ítem como fundamental, que en nuestro caso podría ser siete u ocho, y a partir del cual se incluirían las preguntas en la versión final de la entrevista.

ESTUDIO 2

MÉTODO

Muestra

Entrevistados. Seis hombres y seis mujeres fueron entrevistados mediante la ECE, en cada grupo la mitad tuvo experiencia previa en el puesto de camarero. Su edad varía entre 21 y 43 años. Ocho eran españoles y cuatro eran extranjeros y provenían de México, Honduras, Colombia y Polonia (dos pertenecían a candidatos con experiencia laboral previa y dos a candidatos sin experiencia; dos eran mujeres y dos hombres). Los entrevistados no estaban sujetos a un proceso de selección verdadero, la muestra se creó para satisfacer las necesidades del presente estudio, estando todos los candidatos expertos empleados como camareros en el momento de la realización de la entrevista.

Jueces. Doce personas (tres hombres y nueve mujeres) que no tenían experiencia previa a la hora de valorar las entrevistas conductuales, pero fueron entrenados en la función que debían desempeñar. Todos eran de nacionalidad española.

PROCEDIMIENTO

El total de 12 entrevistas de las que se componía nuestro estudio fueron llevadas a cabo por un único entrevistador y grabadas en un soporte digital de tal modo que solamente se oía la voz del candidato, sin que se pudiera ver su apariencia física. Esto se hizo para controlar que sólo las respuestas a las preguntas de entrevista eran evaluadas por los jueces. Para llevar a cabo las entrevistas se utilizaron dos ver-

siones resumidas de la entrevista descrita anteriormente en el estudio de la validación de contenido (para candidatos con y sin previa experiencia en el puesto). De las 14 preguntas originales se han utilizado 10 preguntas para candidatos expertos y 9 para los candidatos inexpertos que reflejaban las mismas 7 dimensiones generales (amabilidad, resistencia al estrés, responsabilidad, atención en el trabajo, conocimientos del puesto, compañerismo, discreción).

Como ya se indicó, los jueces recibieron un curso de formación sobre la ECE. En dicho curso, los jueces recibían información sobre qué son las ECE, que las diferencia de otras modalidades de entrevista, cómo se recoge información, cómo se formulan las preguntas, qué relación tienen con el empleo a desarrollar y, fundamentalmente, como se evalúan las respuestas mediante escalas de valoración con anclajes conductuales. Antes de comenzar el estudio propiamente dicho, los jueces hicieron varias prácticas de evaluación de preguntas y se contestaron las preguntas que tenían. Además del entrenamiento previo, los evaluadores fueron equipados con un manual que contenía las instrucciones específicas, las versiones de la entrevista para candidatos con y sin experiencia con las preguntas de sondeo, las descripciones de las dimensiones y las escalas de valoración de conducta con anclajes conductuales. Los panelistas disponían también de una hoja de evaluación estandarizada para cada candidato.

Cuatro candidatos fueron valorados por todos los jueces (12) y los ocho restantes por seis jueces (cuatro por un grupo de seis jueces y los demás cuatro por otro grupo de seis). Nunca se evaluaron más de

dos candidatos en una reunión de los miembros del panel. Los miembros del panel trabajaron por separado, sin poder consultar sus opiniones con los demás miembros. Se les pidió a los jueces que tomaran notas mientras escuchaban la entrevista y valorasen al candidato al final de la entrevista. Los candidatos fueron valorados en la escala de 1 (desempeño extraordinariamente malo) a 5 (desempeño extraordinariamente bueno) de forma separada en cada pregunta.

Para conseguir el valor de la fiabilidad de la entrevista en cuestión se utilizó un coeficiente de acuerdo entre los evaluadores. Se calculó la fiabilidad de los paneles ($r_{xx \text{ panel}}$) para todas entrevistas realizadas. Con este propósito se aplicó la fórmula profecía de Spearman-Brown, entendiendo como "n" el número de los miembros de panel. Puesto que había paneles de 6 ó 12 miembros, la fórmula podía adoptar una de las siguientes formas:

$$r_{xx \text{ panel}} = (6 * r_{xx \text{ individual}}) / [1 + (5 * r_{xx \text{ individual}})]$$

o bien

$$r_{xx \text{ panel}} = (12 * r_{xx \text{ individual}}) / [1 + (11 * r_{xx \text{ individual}})]$$

RESULTADOS

En la Tabla 3, la primera columna muestra el número de la entrevista, la segunda, la correlación promedio entre las valoraciones de todos los jueces miembros del panel (fiabilidad individual promedio), la tercera, la fiabilidad de panel, y las tres últimas columnas, el número de jueces que valoraron cada entrevista, el sexo del entrevistado y si éste poseía o no experiencia laboral previa en el puesto. Se puede apreciar que el coeficiente de fiabi-

lidad interjuez mínimo encontrado era de .34 y el máximo de .91, con un promedio ponderado de .78.

DISCUSIÓN

En este estudio se confirmó nuestra hipótesis inicial, que el valor de la fiabilidad de la ECE desarrollada será alrededor de .75, tal y cómo se indicó anteriormente

en los estudios meta-analíticos sobre las ECE (Conway, Jako y Goodman, 1995; Salgado y Moscoso, 1995). En concreto, encontramos la fiabilidad promedio ponderada por el tamaño de la muestra de .78, es decir ligeramente superior a la esperada.

Otro resultado importante que también cabe destacar es que de las 12 entrevistas, las 4 que han sido valoradas por todos los

Tabla 3.
Resultados del análisis de fiabilidad interjuez de la entrevista desarrollada

Entrevista	$r_{xx \text{ individual}}$	$r_{xx \text{ panel}}$	miembros de panel	Sexo del entrevistado	Experiencia laboral
1.	.46	.91	12	M	No
2.	.38	.88	12	M	Si
3.	.29	.71	6	M	Si
4.	.53	.87	6	H	No
5.	.08	.34	6	H	Si
6.	.40	.80	6	M	No
7.	.17	.55	6	H	Si
8.	.16	.53	6	M	No
9.	.37	.78	6	M	Si
10	.30	.72	6	H	No
11	.37	.87	12	H	No
12.	.44	.90	12	H	Si

Notas. $r_{xx \text{ individual}}$ – el coeficiente de correlación promedio entre todos los jueces miembros del panel, $r_{xx \text{ panel}}$ – la fiabilidad de panel, M=mujer, H=hombre.

miembros del panel (12) son las que han alcanzado los niveles de fiabilidad más altos con un promedio de .89. No obstante, entre los restantes ocho entrevistas, tres estaban muy por debajo del valor .70, considerado una fiabilidad aceptable de las ECE. Dichas entrevistas fueron valoradas por distintos paneles, dos fueron realizadas con hombres y una con una mujer; en dos casos el candidato tuvo experiencia previa en el puesto y en un caso no. Por las razones mencionadas anteriormente, cabe descartar que el bajo valor de fiabilidad en los tres casos pudiese ser el resultado de las características del panel o bien del perfil del candidato. Para contrastar, en el estudio de Salgado, Moscoso y Gorriti (2004) sobre las entrevistas conductuales estructuradas utilizadas en la AGPV se encontró que la mayoría de las entrevistas alcanzó los niveles de fiabilidad de panel por encima de .90 con un valor mínimo de .73.

Sin embargo, entre los puntos fuertes de este estudio cabe resaltar que el número de los miembros de panel fue de un mínimo de 6 personas y un máximo de 12. En la literatura existente acerca de la fiabilidad interjuez de la entrevista no hemos encontrado ningún estudio con el número igual (o superior) de jueces que valoraban a la vez a un sólo candidato. Cuantas más personas valoren dicho candidato, más fiable es el resultado final, y ésta es la base para tomar las decisiones sobre la eventual contratación del mismo. Por esa razón, es de interés para las empresas conseguir, en la medida de lo posible, que la proporción de varianza verdadera que hay en la varianza empírica de las puntuaciones de los candidatos sea lo más alta posible. Como muestra el presente estudio, podemos conseguirlo aumentando el número de evaluadores.

DISCUSIÓN GENERAL

El presente trabajo consistió en un examen empírico de la validez de contenido de una entrevista conductual estructurada, basándose en las opiniones de un panel de expertos, según el método sugerido por Lawshe (1975). Por otro lado se llevó a cabo un examen empírico de la fiabilidad de la entrevista, aplicándose a los candidatos expertos e inexpertos y calculando el coeficiente de fiabilidad a partir de las valoraciones aportadas por los paneles de 6 ó 12 jueces. Para poder realizar los estudios mencionados se creó una entrevista de descripción de conducta para un puesto de camarero, partiendo de un análisis de puesto.

Con respecto a la validación del contenido, siguiendo el método propuesto por Lawshe (1975), encontramos una RVC positiva para todas las dimensiones y para todas las preguntas menos una, lo que implica que estas han sido valoradas como fundamentales por más de la mitad de los miembros de panel. No obstante, la RVC mínima aceptable para el Panel de la Evaluación del Contenido compuesto por 11 expertos es de .59 y esta RVC resultó ser una exigencia muy elevada para la entrevista desarrollada, ya que no la alcanzaron incluso las dimensiones y preguntas sobre los cuales 8 de los 11 panelistas estaban de acuerdo en que eran fundamentales para un desempeño adecuado en el puesto del camarero. En dos de los tres estudios anteriores en los que se aplicó esta misma técnica, tampoco se alcanzó una RVC significativa en la mayoría de preguntas. Este hecho está relacionado con el número de expertos utilizados y con que la fórmula RVC `penaliza' la presencia de un número pequeño de jueces. También se relaciona con la característica de este coeficiente de

no asignar ningún peso a las preguntas consideradas por los jueces como *útiles pero no fundamentales* para un buen desempeño en el puesto. Por estas razones, sugerimos que la fórmula debería ser revisada para incorporar de algún modo las preguntas que han sido valoradas como útiles por la mayoría de los jueces. Por ejemplo, se podría asignar algún peso menor que 1 también a los ítems considerados *útiles, aunque no fundamentales*, que hasta el momento han sido tratados igual que los ítems valorados como irrelevantes.

Por otro lado, el coeficiente de fiabilidad promedio obtenido en esta investigación es de .78, lo que puede considerarse un resultado más que satisfactorio, en concreto comparándolo con el resultado del clásico meta-análisis de Conway, Jako y Goodman (1995) cuyo resultado promedio fue de .70, y de .75 para las entrevistas con el cuarto nivel de estandarización de preguntas (igual al utilizado en la entrevista en cuestión), o comparándolo con el resultado de un trabajo meta-analítico de Salgado y Moscoso (1995), donde la fiabilidad promedio encontrada de las ECE fue de .75.

Aunque el resultado obtenido en el presente estudio (.78) es más bajo que la fiabilidad promedio de .93 encontrada por Salgado, Moscoso y Gorriti (2004) para las entrevistas conductuales estructuradas utilizadas en la selección por la AGPV, cabe resaltar que en nuestro estudio no participaron los jueces expertos, ni entrevistadores homologados, indudablemente una de las limitaciones de este estudio. Al mismo tiempo se plantea una conclusión: El uso de los entrevistadores expertos y homologados puede ser un elemento crucial para la aplicación de una ECE con

éxito, siempre y cuando ésta haya sido desarrollada de acuerdo con todos los pasos debidos (véase Salgado y Moscoso, 2001). Por esta razón deducimos que otra de las limitaciones de este estudio fue un tiempo de entrenamiento de los jueces relativamente corto (una sesión de dos horas) que podría haberse alargado. Considerando todas estas desventajas, cabe destacar otra vez que el resultado de fiabilidad interjuez obtenido (.78) es exitoso, ya que los instrumentos de selección más comunes, como los tests de habilidad mental general o los tests de la personalidad basados en cinco grandes factores tienen una fiabilidad promedio de .83 y .80 respectivamente, es decir, muy cercana a la encontrada para el instrumento analizado. Por otra parte, la diferencia entre la fiabilidad encontrada por Salgado et al. (2004) y la encontrada en este estudio puede ser atribuible a la experiencia de los miembros del panel al juzgar las conductas relativas al puesto de trabajo. Si esta conjetura es correcta, la diferencia entre ser experto y no serlo añadiría (o reduciría) un 15% el error de medida (.93-.78).

Finalmente cabe mencionar que quedaría pendiente un estudio en el que se examine la validez predictiva de la entrevista creada, es decir, que confirmare o no que las personas seleccionadas con esta entrevista presentan mejores resultados en términos del desempeño de tarea, desempeño contextual y, globalmente, en su productividad. Esto implicaría el uso de dicho instrumento en el proceso selectivo y la medición del desempeño de personas evaluadas tras transcurrir un periodo de tiempo lo suficientemente largo para que permitiese la observación de su desempeño en el puesto. En nuestro caso, no hemos podido realizar tal estudio aunque creemos

que, dada la fiabilidad obtenida ($r_{xx}=.78$), la validez de contenido alcanzada (IVC=.86) y que fue desarrollada con todos los requisitos de una entrevista conductual estructurada, su validez de criterio debiera estar en torno a .45, al igual que el promedio obtenido en las revisiones meta-analíticas de las entrevistas conductuales estructuradas (Salgado y Moscoso, 1995; Salgado, 1999; Salgado, Viswesvaran y Ones, 2001).

REFERENCIAS BIBLIOGRÁFICAS

- Anderson, N. R. (1992). Eight decades of employment interview research: a retrospective meta-review and prospective commentary. *European Work and Organizational Psychologist*, 2 (1), 1-32.
- Arvey, R. D., y Campion, J. E. (1982). The employment interview: a Summary and review of recent research. *Personnel Psychology*, 35, 281-322.
- Campion, M. A.; Palmer, D.K, y Campion, J.E. (1997). A review of structure in the selection interview. *Personnel psychology*, 50, 1-64.
- Campion, M. A.; Pursell, E. D., y Brown, B. K. (1988). Structured interviewing: Raising the psychometric properties of employment interview. *Personnel Psychology*, 41, 2542.
- Carrier, M. R.; Dalessio, A. T., y Brown, S. H. (1990). Correspondence between estimates of content and criterion-related validity values. *Personnel Psychology*, 43, 85-100.
- 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(5), 565-579.
- Flanagan, J. C. (1954). The critical incidents technique. *Psychological Bulletin*, 51, 327-358.
- Green, P. C. y Hogan, D. D. (1982). *Behavioral Interviewing*. Memphis, TN.
- Harris, M. M. (1989). Reconsidering the employment interview: a review of recent literature and suggestions for future research. *Personnel Psychology*, 42, 691-725.
- Huffcutt, A. I., y Arthur, W. Jr. (1994). Hunter y Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, 79, 184-190.
- Hunter, J. E., y Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Janz, T. (1977). The behaviour based patterned interview: an effective alternative to the warm smile selection interview. *Compendio de la Conferencia Anual de la Asociación de las Ciencias Administrativas de Canadá*.
- Janz, T. (1982). Initial comparisons of patterned behavior description interview versus unstructured interviews. *Journal of Applied Psychology*, 67, 577-580.
- 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 practise*. Beverly Hills, Sage: 158-168.

- Latham, G. P. (1989). The reliability, validity and practicality of the situational interview. En R. W., Eder, y G. R., Ferris (Eds.). *The employment interview: Theory, research and practise*. Beverly Hills, Sage, 169-182.
- Latham, G. P. et al. (1980). The situational interview. *Journal of Applied Psychology*, 65, 422-427.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563-575.
- Mayfield, E.C. (1964). The selection interview: A re-evaluation of published research. *Personnel Psychology*, 17, 239-260.
- 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 (4), 599-616.
- Motowidlo, S. J., Carter, G. W., Dunette, 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.
- Reilly, R. R., y Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology*, 35, 1-62.
- Salgado, J. F. (1995). Validez de las entrevistas conductuales estructuradas. *Psicología del Trabajo y las Organizaciones*, 11, 9-24.
- Salgado, J. F. (1999). Personnel selection methods. En Cooper C. L. y Robertson I. T (Eds.). *International Review of Industrial and Organizational Psychology*, 12, London, UK, John Wiley and Sons: 1-53.
- Salgado, J. F., y Moscoso, S. (1995). Validez de las entrevistas conductuales estructuradas. *Psicología del Trabajo y las Organizaciones*, 11, 9-24.
- Moscoso, S. y Salgado, J. F. (2001). *Entrevista Conductual Estructurada de Selección de Personal. Teoría, práctica y rentabilidad*. Madrid, Pirámide.
- Moscoso, S. y Salgado, J. F. (2001b). Psychometric properties of a structured interview to hire private security personnel. *Journal of Business and Psychology*, 16(1), 51-59.
- 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: Meta-análisis de la Fiabilidad, *Revista de Psicología de Trabajo y Organizaciones*, 20 (2), 107-139.
- Salgado, J. F., Viswesvaran, C., y Ones, D. S. (2001). Predictors used for personnel selection. En N., Anderson, D. S., Ones, H. K., Sinangil, y C., Viswesvaran (Eds.). *Handbook of industrial, work, and organizational psychology. Vol 1*. London, UK, Sage: 165-199.
- Schmidt, F. L., y Rader, M. (1999). Exploring the boundary conditions for interview validity: Meta-analytic validity findings for a new interview type. *Personnel Psychology*, 29, 79-101.
- Schmitt, N. (1976). Social and situational determinants of interviews decisions:

Implications for the employment interview. *Personnel Psychology*, 29, 79-101.

Schmitt, N., y Ostroff, C. (1986). Operationalizing the «behavioral consistency» approach: Selection test development based on a content-oriented strategy. *Personnel Psychology*, 39, 91-108.

Schüler, H., y Funke, U. (1989). The interview as a multimodal procedure. En Eder, E.W. y Ferris G. R. (Eds.). *The employment interview: Theory, research and practice*. Newbury Park, CA: Sage.

SIOP (2003). Society for Industrial and Organizational Psychology. *Principles for the validation and use of personnel selection procedures (4th ed.)*. Documento pdf., www.siop.org

Ulrich, L., y Trumbo, D. (1965). The selection interview since 1949. *Psychological Bulletin*, 63, 100-116.

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 behaviour employment interview questions. *Journal of Occupational and Organizational Psychology*, 75, 277-294.

Wagner, R. (1949). The employment interview: A critical review. *Personnel Psychology*, 2, 17-46.

Wiesner, W. H., y Cronshaw, S. F. (1988). The moderating impact of interview format and degree of structure on interview validity. *Journal of Occupational Psychology*, 61, 275-290.

Wright, O. R. Jr. (1969). Summary of research on the selection interview since 1964. *Personnel Psychology*, 22, 391-413.

Wright, P. M., Lichtenfels, P. A. y Purcell, E.D (1989). The structured interview: Additional studies and meta-analysis. *Journal of Occupational Psychology*. 62, 191-199.