



Structured behavioral and conventional interviews: Differences and biases in interviewer ratings



Pamela Alonso*, Silvia Moscoso

Universidad de Santiago de Compostela, Spain

ARTICLE INFO

Article history:

Received 31 May 2017

Accepted 31 July 2017

Available online 5 November 2017

Keywords:

Conventional interview

Behavioral interview

Accuracy

Sex similarity

Interviewer confidence

ABSTRACT

This research examined three issues: (1) the degree to which interviewers feel confident about their decisions when they use a specific type of interview (behavioral vs. conventional), (2) what interview type shows better capacity for identifying candidates' suitability for a job, and (3) the effect of two biases on interview ratings: a) the sex similarity between candidate and interviewer and b) having prior information about the candidate. The results showed that the SBI made raters feel more confident and their appraisals were more accurate, that prior information negatively affects the interview outcomes, and that sex similarity showed inconclusive results. Implications for theory and practice of personnel interview are discussed.

© 2017 Colegio Oficial de Psicólogos de Madrid. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Entrevistas conductuales y convencionales estructuradas: diferencias y sesgos en las valoraciones de los entrevistadores

RESUMEN

Esta investigación examinó tres cuestiones: (1) el grado en que los entrevistadores se sienten seguros con sus evaluaciones cuando utilizan un tipo específico de entrevista (conductual o convencional), (2) qué tipo de entrevista muestra mejor capacidad para identificar la idoneidad de los candidatos y (3) el efecto de dos sesgos en las calificaciones de las entrevistas: (a) la similitud entre el sexo del candidato y el del entrevistador y (b) tener información previa sobre el candidato. Los resultados mostraron que los evaluadores se sienten más seguros de sus evaluaciones y que éstas son más precisas con la entrevista conductual estructurada - ECE, que la información previa sobre el candidato afecta negativamente a la entrevista y que la similitud en el sexo de entrevistador y entrevistado ha producido resultados no concluyentes. Finalmente, se discuten las implicaciones para la teoría y la práctica de la entrevista de selección.

© 2017 Colegio Oficial de Psicólogos de Madrid. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Decades of scientific research have established three important findings concerning personnel selection interviews. First, according to a number of surveys carried out in different countries and with all types of organizations, the employment interview is the most frequently used procedure and it is the most relevant in the

decision-making of practitioners (Alonso, Moscoso, & Cuadrado, 2015; Salgado & Moscoso, 2011). Second, research has also found that structured interviews have proven to be a valid procedure for predicting job performance (Huffcutt, Culbertson, & Weyhrauch, 2014; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Salgado & Moscoso, 1995, 2006). The third finding has been to demonstrate, across the world, that interviews are overall the instrument which is most positively regarded by candidates (Anderson, Salgado, & Hülsheger, 2010; Liu, Potočnik, & Anderson, 2016; Steiner & Gilliland, 1996).

* Corresponding author. Departamento de Organización de Empresas y Comercialización. University of Santiago de Compostela. 15782 Santiago de Compostela, Spain.

E-mail address: pamela.alonso@usc.es (P. Alonso).

A scarcely researched issue concerning the selection interview is the degree to which interviewers feel confident about their decisions when they use a specific type of interview (e.g., unstructured vs. structured). A second issue is to identify what structured interview content (e.g., conventional vs. behavioral) shows a better capacity to identify candidates' suitability for a job. A third less investigated issue is related to two biases that can affect the assessments: a) the degree to which sex similarity between candidate and interviewer affects interview decisions and b) the effect of having additional information about the candidate (e.g., test results, resume, and recommendation letters).

The objective of this research is to shed further light on these four neglected issues concerning the usefulness of the interview as a procedure for making hiring decisions.

Employment Interviews: Types and Psychometric Properties

There are three main interview types depending on their content and degree of structure (Salgado & Moscoso, 2002): (1) Conventional Unstructured Interview (CUI), which is the most used personnel interview, refers to an informal conversation between the candidate and the interviewer, who formulates the questions according to the course of the conversation and without following any previous script (Dipboye, 1992; Goodale, 1982); (2) Structured Conventional Interview (SCI), in which the interviewer works from a script or a series of guidelines about the information that must be obtained from each interviewee and it typically includes questions about credentials, technical skills, experience, and self-evaluations (Janz, Hellervik, & Gilmore, 1986); and (3) Structured Behavioral Interview (SBI), which is based on the evaluation of past behaviors (Janz, 1982, 1989; Moscoso & Salgado, 2001; Motowidlo et al., 1992; Salgado & Moscoso, 2002, 2011). Meta-analyses have shown the reliability and construct and criterion validity of the different types of interviews (e.g., Huffcutt & Arthur, 1994; Huffcutt, Culbertson, Weyhrauch, 2013, 2014; McDaniel et al., 1994; Salgado & Moscoso, 1995, 2006). Other studies have also reported on content validity (e.g., Choragwicka & Moscoso, 2007; Moscoso & Salgado, 2001).

With respect to reliability, Huffcutt et al. (2013) carried out a new meta-analysis to update the results found by Conway, Jako, and Goodman (1995). The results for low structure interviews (CUI) were .40 when they were evaluated by separate interviewers and .55 in panel interviews. For the interviews with a medium level of structure (SCI), the values increase to .48 (serial interviews) and .73 (panel of evaluators). Finally, in the category of "high structure" (SBI) they found a reliability of .61 in the case of serial interviews and .78 when the evaluation is performed by a panel of evaluators. In their meta-analysis, Salgado, Moscoso, and Gorriti (2004) found a coefficient of .83 for SBI. These results are like those found by Conway et al. (1995), that is, the higher the degree of structure, the greater the reliability among interviewers.

Several studies have found that structure is also an important moderator of validity since as the level of structure increases, the interview validity increases. Recently, Huffcutt et al. (2014) found higher validity coefficients. Specifically, their results showed a coefficient of .20 for non-structured interviews (CUI), .46 for conventional structured interviews (SCI) and .70 for those with a higher level of structure (SBI). This last result is very similar to the value of .68 found by the meta-analysis of Salgado and Moscoso, 1995, 2006, in which they concluded that the SBI was valid for all occupations with validity ranging from .52 for managers to .80 for clerical occupations.

Other relevant studies have found that the SBI is more resistant to adverse impact (Alonso, 2011; Alonso, Moscoso, & Salgado, 2017; Levashina, Hartwell, Morgeson, & Campion, 2014; Rodríguez,

2016). There is also evidence of the economic utility of the SBI (Salgado, 2007). As a whole, the results of the meta-analytical reviews performed supported the use of SBIs for hiring decisions.

Research vs. Practice Gap

Despite the empirical evidence on the psychometric properties of the SBI, there is still a gap between research findings and professional practice (Alonso et al., 2015; Anderson, Herriot, & Hodkinson, 2001). Nowadays, most medium and small companies continue using unstructured interviews rather than structured behavioral ones.

In this regard, there are some issues related to professional practices that have been insufficiently researched. For instance, research is scarce concerning the degree to which interviewers feel confident about the decisions based on SBI or SCI. Two small-sample studies carried out by Salgado and Moscoso (1997, 1998) found that the interviewers have more confidence in their assessments with SBI than with SCI. However, additional studies are necessary.

Research has also shown that access to previous information about candidates (e.g., resume, recommendation letters, academic record, and test scores) can produce impression bias in appraisals (Campion, 1978; Paunonen, Jackson, & Oberman, 1987). For example, Macan and Dipboye (1990) found that the interviewer's prior impressions on candidates correlated .35 with the ratings given to interviewees. The frequency of this kind of bias seems to be larger for unstructured interviews than for structured ones (Dipboye, 1997). In fact, research on highly structured interviews recommends against having access to the candidate's prior information (Campion, Palmer, & Campion, 1997; Latham, Saari, Pursell, & Campion, 1980). This recommendation has been supported by the meta-analytical studies of McDaniel et al. (1994) and Searcy, Woods, Gatewood, and Lace (1993), who found higher criterion validity when the interviewers did not have access to cognitive test scores.

Another scarcely researched issue is the degree to which sex similarity between candidate and interviewer can bias interview decisions. Elliott (1981) found that the female candidates were assessed slightly higher by male interviewers ($d=0.28$) and that the male candidates were rated similarly by female and male interviewers in a SCI. Using a campus recruitment interview, Graves and Powell's (1996) findings showed that sex similarity of interviewer and candidate correlated .08 with the overall appraisal. In a third study, Sacco, Scheu, Ryan, and Schmitt (2003) found that the ratings for the candidate were higher when interviewer and candidate sex were matched ($d=0.09$). More recently, McCarthy, Van Iddekinge, and Campion (2010) examined the effects of sex similarity on the evaluations for three types of highly structured interviews (experience-based, situational, and behavioral). They concluded that the effects of sex similarity were non-significant. Therefore, as a whole, the findings of these three studies are inconclusive, although they suggest that SBIs can be more robust against sex-similarity bias than SCIs and UCIs.

Aims of the Study

The first objective of this study is to compare the effectiveness of each interview in identifying the candidate's suitability for a job. Considering that the SBI has more validity than the SCI, the following hypotheses are considered:

Hypothesis 1: the SBI identifies candidates' capacities more accurately, which implies that it discriminates better between qualified and unqualified candidates than the SCI.

Hypothesis 1a: Accuracy for identifying qualified and unqualified candidates will be greater for the SBI than for the SCI.

Hypothesis 1b: Qualified candidates will receive higher scores in the SBI than in the SCI, while non-qualified candidates will receive lower scores in the SBI than in the SCI.

The second objective is the study of the degree to which interviewers feel confident about their decisions according to the type of interview used, SBI or SCI. Considering that the SBI allows for a more precise evaluation and that the information acquired through the SCI is more susceptible to different interpretations, the following hypothesis is proposed:

Hypothesis 2: Interviewers will be more confident about their decisions when using the SBI than when using the SCI.

The third aim of this research is to analyze and compare SCI and SBI resistance to two biases that may influence the interview decision: (a) the effect of having additional information about the candidate and (b) the effect of the similarity of sex between evaluator and candidate. With regard to these two biases, we make the following two hypotheses:

Hypothesis 3: Additional information affects assessments made with both SCIs and SBIs.

Hypothesis 4: SCIs are more affected by interviewer-interviewee sex similarity than SBIs.

Method

Sample

The sample consisted of 241 university students aged between 18 and 59 (mean was 24.53 and *SD* was 6.7); 78.4% were studying a subject related to personnel selection and, therefore, had theoretical knowledge about the different types of interview that are used in selection contexts; 57.4% of the sample was female. The study was presented to the students as an academic exercise in which they had to evaluate different candidates.

Design

We used a 2 x 2 x 2 design. The independent variables were: (a) type of interview, i.e., SCI or SBI; (b) candidate qualification level, i.e., qualified or unqualified; and (c) interviewee sex. The dependent variable was the raters' assessment of the candidates.

Experiment Preparation

Scripts creation. Before the video-recording of interviews, four scripts were developed for a HR technician job. The content of these scripts detailed both the questions that the interviewer should ask and the exact answers that the interviewee should give. Scripts 1 and 2 were for an SCI and the other two for an SBI.

For the two interview types, we created two different scenarios with exactly the same questions. Nevertheless, the candidate's responses varied substantially, so that the answers corresponded to a qualified candidate in scripts 1 and 3, and the answers

INTERVIEW	CANDIDATE		
	QUALIFIED		UNQUALIFIED
	SCI	SCRIPT 1	SCRIPT 2
	SBI	SCRIPT 3	SCRIPT 4

Figure 1. Interview Scripts.

corresponded to an unqualified candidate in scripts 2 and 4. The participants rated the candidates in four dimensions, including organization and planning, teamwork, problem-solving, and overall score. Figure 1 shows the four experimental combinations.

To verify that the scripts fulfilled the purpose of this research, seven personnel selection experts were asked to evaluate the candidates represented in the four designed scripts. A written copy of each of the scripts was given to them along with an evaluation sheet with the interview dimensions. A 5-point Likert scale was used. Table 1 summarizes the assessments made by the expert group.

As can be seen in Table 1, in both the SBI and the SCI, the qualified candidate obtains higher scores than the unqualified candidate in all dimensions; in all cases these differences were significant. Scores are more extreme in the case of SBI, that is, the differences between the qualified and the unqualified candidate are much more pronounced in the SBI than in the SCI.

Video recordings. A man and a woman were selected to play the role of interviewees. To avoid the image of the interviewees making an impression on the raters which could affect the ratings, an effort was made to ensure that the two candidates had a similar image and both appeared dressed in the same way during the interview: a black jacket and a white shirt. The role of interviewers was played by two personnel selection experts.

Finally, eight interviews were recorded, four in which the man played the role of interviewee and four in which it was the woman who did. The scripts were the same for both actors in each condition. The duration of the videos is approximately 10 minutes in the case of the SCI, and in the SBI, script 3 (qualified candidate) lasted 35 minutes and script 4 (unqualified candidate) 22 minutes.

Measures

Interviewee assessment. Participants assessed the candidate in the dimensions of organization and planning, teamwork, problem-solving, and overall assessment on a scale of 1 to 5 (1 = *insufficient* and 5 = *excellent*). Although one of the characteristics of the SBI is the use of behavioral anchor scales (BARS) for the assessment of candidates, in this study it was decided to use the same rating scale

Table 1
Means and Standard Deviations of the Experts' Assessments.

	SCI Qualified		SCI Unqualified		<i>F</i>	SBI Qualified		SBI Unqualified		<i>F</i>
	$\bar{X}$	<i>SD</i>	$\bar{X}$	<i>SD</i>		$\bar{X}$	<i>SD</i>	$\bar{X}$	<i>SD</i>	
Organization and planning	3.57	0.53	2.00	0.00	60.50***	4.86	0.38	1.43	0.53	192.00***
Teamwork	3.14	0.69	2.14	0.69	7.35**	4.71	0.49	1.86	0.38	150.00***
Problem-solving	3.29	0.76	1.57	0.53	24.00***	4.71	0.49	1.43	0.53	144.27***
Overall score	3.29	0.76	1.86	0.38	20.00***	4.79	0.39	1.71	0.49	168.09***

Note. $\bar{X}$ = mean; *SD* = standard deviation.

** $p < .01$, *** $p < .001$

	Candidate	SCI	SBI
GROUP 1	<i>Qualified</i>	<i>Woman</i>	<i>Man</i>
GROUP 2	<i>Qualified</i>	<i>Man</i>	<i>Woman</i>
GROUP 3	<i>Unqualified</i>	<i>Woman</i>	<i>Man</i>
GROUP 4	<i>Unqualified</i>	<i>Man</i>	<i>Woman</i>

Figure 2. Candidates Rated by the First Sub-sample.

		Woman	Man
GROUP 1	SCI + SBI	<i>Qualified</i>	<i>Unqualified</i>
GROUP 2	SCI + SBI	<i>Unqualified</i>	<i>Qualified</i>

Figure 3. Candidates Rated by the Second Sub-sample.

in both interviews. In this way, the potential effect of the rating system was neutralized.

Internal consistency reliability, calculated from Cronbach's alpha, ranges from .77 ($n = 143$) in the case of the qualified interviewee to .86 ($n = 154$) for the unqualified candidate using the SBI. In SCI, the values obtained were .80 ($n = 121$) and .78 ($n = 153$) for the qualified and the unqualified applicant respectively.

Confidence about the assessment. Raters indicated the degree of confidence about the assessment they had made. A 5-point Likert scale was used (1 = *not confident* and 5 = *totally confident*).

Procedure

Participants had to adopt the role of raters. They were instructed so as to believe that the interviews were part of a real selection process for a HR technician. After seeing each interview, raters assessed the candidates, using the scale described in the previous section.

A sub-sample ($n = 175$) was divided into four groups, in which sex and candidate type (qualified vs. unqualified) were alternatively presented, as can be seen in Figure 2. Raters watched an SCI and an SBI. In the first group, the female candidate was interviewed with an SCI and the male candidate with an SBI. The two candidates were qualified. In the second group, the female candidate was interviewed with the SBI and the male candidate with an SCI. In this case, the two candidates were qualified, too. In the third group, the female candidate was interviewed with an SCI and the male candidate with an SBI. In the fourth group, the female candidate was interviewed with SBI and the male candidate with an SCI. In these last two groups, the candidates were unqualified. Furthermore, in each group, raters were divided randomly, one half of the group first watched the SBI and then the SCI and the other half watched them in the reverse order.

Another sub-sample ($n = 66$) watched both interview types for the same candidate. That is, they watched two different interviews with the same candidate. As shown in Figure 3, raters were divided into two groups, in which the sex and candidate type were alternated. Thus, in one of the groups, the qualified candidate (both in SCI and SBI) corresponded to the woman and the unqualified candidate to the man and in the other group, the qualified candidate corresponded to the man and the nonqualified to the woman.

Results

Table 2 shows the results for the four interview ratings. Firstly, we report the differences between qualified and unqualified candidates, who have been interviewed using the SCI and, secondly, the

differences using the SBI. Finally, the differences according to the type of interview appear.

Regarding the SCI, the mean of the qualified candidate ranged from 3.27 for the teamwork dimension to 3.85 for organization and planning. The overall score was 3.55. For the unqualified candidate, the mean ranges from 2.04 for the problem-solving dimension to 3.20 for teamwork. In this case, the overall score was 2.64. The differences between the two candidates were significant for the dimensions of organization and planning, problem-solving, and overall score ($p < .001$). The effect sizes (ES) ranged from $d = 1.13$ for the overall score to $d = 1.62$ for the problem-solving dimension.

With respect to the SBI, the differences between candidates' scores were more extreme. The mean of the qualified candidate ranged from 4.50 for organization and planning to 4.19 for teamwork, while the mean of the unqualified candidate was considerably smaller. The mean ranged from 2.15 for teamwork to 1.88 for organization and planning. The differences between the interviewees were statistically significant for the four dimensions ($p < .001$), which confirms Hypothesis 1. The ES ranged from $d = 3.40$ for overall scores to $d = 2.64$ for teamwork. This result suggests that the differentiation between qualified and unqualified candidates becomes more accurate with the SBI.

The last two columns of Table 2 show the results of the comparison between the two candidates according to the type of interview. The qualified candidate obtained higher scores when evaluated with the SBI. The differences are statistically significant ($p < .001$) for the four dimensions. They ranged from $d = 1.24$ for the teamwork dimension to $d = .85$ for organization and planning. In the case of the unqualified candidate, the scores were lower when interviewed with an SBI, and the differences were statistically significant for organization and planning, teamwork, and overall score ($p < .001$). The ES ranged from $d = -1.15$ for teamwork to $d = -0.80$ for the overall score. These results support H1a and H1b, as the SBI allows for better discrimination between qualified and unqualified candidates than SCI.

With regard to the confidence degree of raters, the means were 4.13 and 2.97 for the SBI and SCI, respectively, in the case of the qualified candidate. The difference was statistically significant ($p < .001$, $d = 1.22$). For the unqualified candidate, the degree of confidence was 4.12 and 3.47 for the SBI and the SCI, respectively ($p < .001$, $d = 0.8$). Therefore, the SBI interviewers reported a similar degree of confidence for qualified and unqualified candidates, while SCI interviewers reported to be more confident when the candidate was an unqualified one ($p < .001$, $d = 0.51$). In other words, SCI interviewers report to be less confident about their rating when the candidate is qualified. These results supported Hypothesis 2.

Table 3 shows the comparison between the candidate scores when the raters watched one interview type only or both. Positive signs indicate that candidates received a higher score when the raters watched one interview only. For SCI, the ratings were higher for all dimensions in the case of the qualified candidate when the raters have available information on the other interview. However, for the unqualified candidate, the ratings were higher only for the problem-solving dimension.

As far as the SBI is concerned, in the case of the qualified candidate, there were non-significant differences for some of the dimensions assessed. For the unqualified candidate, the scores were lower in all dimensions, and the differences were statistically significant. These results partially supported Hypothesis 3.

Tables 4 and 5 report the results for the effects of interviewer-interviewee sex similarity. Table 4 shows that there were no differences when the qualified candidate was interviewed with a SCI. However, sex-similarity had significant effects for male rater-female candidate combination for the unqualified candidate. In this last case, female candidates obtained higher scores than the male candidates in three of the rated dimensions.

Table 2
Means, Standard Deviations, and Differences in Interview Dimensions and in the Degree of Confidence.

									Candidate Type Comparison				Interview Type Comparison			
									Qualified/Unqualified				SBI/SCI			
	SCI Qualified		SCI Unqualified		SBI Qualified		SBI Unqualified		SCI		SBI		Qualified Candidate		Unqualified Candidate	
	$\bar{X}$	<i>SD</i>	$\bar{X}$	<i>SD</i>	$\bar{X}$	<i>SD</i>	$\bar{X}$	<i>SD</i>	<i>F</i>	<i>d</i>	<i>F</i>	<i>d</i>	<i>F</i>	<i>d</i>	<i>F</i>	<i>d</i>
Organization and planning	3.85	0.86	2.67	0.98	4.50	0.68	1.88	0.85	110.24***	1.26	860.40***	3.38	48.00***	0.85	56.97***	-0.85
Teamwork	3.27	0.81	3.20	0.96	4.19	0.68	2.15	0.85	0.47	0.08	522.31***	2.64	101.90***	1.24	103.45***	-1.15
Problem-solving	3.54	0.83	2.04	0.99	4.26	0.74	1.96	0.83	180.18***	1.62	647.35***	2.93	57.29***	0.93	0.76	-0.10
Overall score	3.55	0.73	2.64	0.86	4.29	0.59	1.99	0.75	86.11***	1.13	856.91***	3.40	83.32***	1.13	49.67***	-0.80
Interviewer confidence about the evaluations	2.97	1.11	3.47	0.88	4.13	0.78	4.12	0.72	17.96***	-0.51	0.01	0.01	99.12***	1.22	50.01***	0.80
	n = 123		n = 158		n = 145		n = 156									

Note. $\bar{X}$ = mean; *SD* = standard deviation; *n* = sample size.
*** *p* < .001

Table 3
Means, Standard Deviations, and Differences in the Interview Dimensions.

	SCI												SBI											
	Qualified candidate						Unqualified candidate						Qualified candidate						Unqualified candidate					
	Only SCI		SCI + SBI				Only SCI		SCI + SBI				Only SBI		SBI + SCI				Only SBI		SBI + SCI			
	$\bar{X}$	<i>SD</i>	$\bar{X}$	<i>SD</i>	<i>F</i>	<i>d</i>	$\bar{X}$	<i>SD</i>	$\bar{X}$	<i>SD</i>	<i>F</i>	<i>d</i>	$\bar{X}$	<i>SD</i>	$\bar{X}$	<i>SD</i>	<i>F</i>	<i>d</i>	$\bar{X}$	<i>SD</i>	$\bar{X}$	<i>SD</i>	<i>F</i>	<i>d</i>
Organization and planning	3.62	0.87	4.29	0.64	19.13***	-0.84	2.79	0.91	2.50	1.06	3.41	0.30	4.51	0.67	4.48	0.70	0.04	0.03	2.03	0.87	1.67	0.78	7.07**	.43
Teamwork	3.13	0.80	3.54	0.78	7.07**	-0.51	3.26	1.01	3.11	0.88	1.00	0.16	4.17	0.71	4.23	0.64	0.25	-0.08	2.33	0.93	1.91	0.66	9.70**	.51
Problem-solving	3.41	0.83	3.83	0.80	6.96**	-0.50	1.85	0.88	2.32	1.06	9.35**	-0.50	4.25	0.79	4.27	0.66	0.03	-0.03	2.07	0.86	1.80	0.76	4.03*	.33
Overall score	3.38	0.72	3.85	0.61	13.23***	-0.70	2.64	0.86	2.64	0.86	0.00	0.00	4.28	0.61	4.30	0.56	0.02	-0.02	2.14	0.79	1.78	0.63	9.37**	.50
Interviewer confidence about the evaluations	2.66	1.10	3.61	0.86	23.34***	-0.92	3.33	1.00	3.67	0.64	5.77***	-0.39	4.16	0.74	4.08	0.84	0.35	0.10	4.07	0.66	4.19	0.80	1.13	-.17
	<i>n</i> = 82		<i>n</i> = 41				<i>n</i> = 91		<i>n</i> = 66				<i>n</i> = 83		<i>n</i> = 62				<i>n</i> = 92		<i>n</i> = 64			

Note. $\bar{X}$ = mean, *SD* = standard deviation, *n* = sample size.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4
Means, Standard Deviations, and Differences due to Sex-Similarity: SCL.

	QUALIFIED CANDIDATE																			
	Female raters		Male raters										Female raters				Male raters			
	Woman		Man		Woman		Man		Woman		Man		Woman		Man					
	$\bar{X}$	SD	$\bar{X}$	SD	F	d	$\bar{X}$	SD	$\bar{X}$	SD	F	d	$\bar{X}$	SD	$\bar{X}$	SD	F	d		
Organization and planning	3.72	0.86	4.02	0.83	3.47	-0.35	3.48	0.63	3.88	0.97	3.78	-0.47	4.03	0.83	4.00	0.85	0.01	0.04		
Teamwork	3.23	0.80	3.33	0.83	0.47	-0.13	3.21	0.68	3.24	0.88	0.03	-0.04	3.41	0.83	3.13	0.83	1.14	0.34		
Problem-solving	3.42	0.80	3.73	0.87	4.13*	-0.37	3.31	0.71	3.50	0.86	0.95	-0.24	3.68	0.92	3.87	0.74	0.51	-0.22		
Overall score	3.46	0.70	3.63	0.74	1.68	-0.24	3.43	0.63	3.49	0.75	0.12	-0.09	3.65	0.75	3.60	0.74	0.05	0.07		
Interviewer confidence	2.80	1.04	3.21	1.19	4.10*	-0.37	2.59	1.02	2.95	1.04	2.18	-0.35	3.38	1.23	2.80	1.01	2.58	0.49		
	$n = 71$		$n = 52$				$n = 29$		$n = 42$				$n = 37$		$n = 15$					
	UNQUALIFIED CANDIDATE																			
	Female raters		Male raters										Female raters				Male raters			
	Woman		Man		Woman		Man		Woman		Man		Woman		Man					
	$\bar{X}$	SD	$\bar{X}$	SD	F	d	$\bar{X}$	SD	$\bar{X}$	SD	F	d	$\bar{X}$	SD	$\bar{X}$	SD	F	d		
Organization and planning	2.66	0.99	2.68	0.98	0.01	-0.01	2.51	1.12	2.82	0.82	1.95	-0.31	2.65	0.95	2.70	1.02	0.06	-0.05		
Teamwork	3.19	1.00	3.21	0.92	0.02	-0.02	3.12	0.99	3.26	1.02	0.38	-0.14	3.43	0.90	2.97	0.90	4.85*	0.51		
Problem-solving	1.90	0.89	2.19	1.06	3.57	-0.30	1.90	0.83	1.89	0.95	0.00	0.01	2.48	1.04	1.89	1.02	6.16*	0.57		
Overall score	2.59	0.91	2.68	0.80	0.42	-0.10	2.53	0.96	2.67	0.87	0.47	-0.15	2.87	0.73	2.49	0.84	4.57*	0.48		
Interviewer confidence	3.41	0.88	3.53	0.88	0.73	-0.14	3.33	0.93	3.50	0.83	0.71	-0.19	3.33	0.92	3.76	0.80	4.83*	-0.50		
	$n = 81$		$n = 77$				$n = 42$		$n = 39$				$n = 40$		$n = 37$					

Note. $\bar{X}$ = mean; SD = standard deviation; n = sample size.

* p < .05

Table 5
Means, Standard Deviations, and Differences due to Sex-Similarity: SBI.

	QUALIFIED CANDIDATE																	
	Female raters		Male raters		Female raters								Male raters					
					Woman				Man				Woman				Man	
	$\bar{X}$	SD	$\bar{X}$	SD	F	d	$\bar{X}$	SD	$\bar{X}$	SD	F	d	$\bar{X}$	SD	$\bar{X}$	SD	F	d
Organization and planning	4.55	0.65	4.44	0.71	0.84	0.15	4.41	0.71	4.66	0.58	2.81	-0.39	4.39	0.72	4.49	0.71	0.35	-0.14
Teamwork	4.04	0.73	4.35	0.59	7.69**	-0.46	4.00	0.80	4.07	0.69	0.18	-0.09	4.23	0.62	4.44	0.55	2.39	-0.36
Problem-solving	4.26	0.73	4.26	0.75	0.00	0.00	4.16	0.72	4.34	0.73	1.17	-0.25	4.13	0.85	4.37	0.66	1.78	-0.32
Overall score	4.22	0.58	4.36	0.59	1.98	-0.24	4.06	0.62	4.34	0.53	4.30*	-0.49	4.27	0.58	4.43	0.59	1.24	-0.27
Interviewer confidence	3.96	0.81	4.29	0.72	6.78**	-0.43	3.88	0.83	4.03	0.80	0.60	-0.18	4.45	0.57	4.17	0.80	2.75	0.39
	$n = 73$		$n = 72$				$n = 32$		$n = 41$				$n = 31$		$n = 41$			
	UNQUALIFIED CANDIDATE																	
	Female raters		Male raters		Female raters								Male raters					
					Woman				Man				Woman				Man	
	$\bar{X}$	SD	$\bar{X}$	SD	F	d	$\bar{X}$	SD	$\bar{X}$	SD	F	d	$\bar{X}$	SD	$\bar{X}$	SD	F	d
Organization and planning	2.09	0.90	1.67	0.74	10.06**	0.51	2.12	0.91	2.03	0.89	0.17	0.10	1.85	0.84	1.47	0.56	5.02*	0.53
Teamwork	2.26	0.92	2.04	0.76	2.61	0.26	2.25	0.94	2.27	0.91	0.00	-0.02	2.05	0.69	2.03	0.85	0.02	0.03
Problem-solving	2.05	0.89	1.85	0.75	2.19	0.24	2.08	0.85	2.00	0.98	0.14	0.09	1.97	0.81	1.72	0.66	2.16	0.34
Overall score	2.14	0.81	1.84	0.64	6.43*	0.41	2.20	0.78	2.03	0.87	0.74	0.21	1.95	0.65	1.71	0.62	2.51	0.38
Interviewer confidence	3.88	0.70	4.38	0.66	21.23***	-0.74	3.80	0.78	4.00	0.53	1.51	-0.29	4.31	0.61	4.46	0.70	0.96	-0.23
	$n = 81$		$n = 75$				$n = 51$		$n = 30$				$n = 39$		$n = 36$			

Note. $\bar{X}$ = mean; SD = standard deviation; n = sample size.

* p < .05, ** p < .01, *** p < .001.

Table 5 shows that there were differences in overall scores for female rater-male candidate combination for the qualified candidate. In this last case, male candidates obtained higher scores than the female candidates. For the unqualified candidate, sex-similarity had significant effects for male rater- female candidate combination for the unqualified candidate in the organization and planning dimension.

Discussion

This study contributes to the research and practice of personnel interviews by shedding further light on four neglected issues concerning the usefulness of selection interviews as a procedure for making hiring decisions. The first issue was whether SBI is a more accurate method for distinguishing between qualified

and unqualified candidates than SCI. The second issue was to examine what interview type produces more self-confidence in ratings and decision made by the interviewers. The third research issue refers to the potential bias of having prior information about candidates. Finally, the fourth issue refers to the effects of interviewer-interviewee sex similarity on interviewer appraisals and decisions.

The first contribution of this study is that, in accordance with Hypothesis 1, our findings showed that SBI allows for a clearer and more accurate differentiation between qualified and unqualified candidates. In other words, the findings supported the hypothesis that SCIs are a weaker method than SBIs for identifying what candidate is a better fit for the job conditions and requirements.

The second contribution of this study is to support previous findings of Salgado and Moscoso (1997, 1998), which showed that interviewers feel more self-confident about their appraisal and decisions when they use a SBI than when they use a SCI. This finding confirmed our Hypothesis 2.

The third contribution was to show how prior information about the candidate can produce biased interview decisions. We found that this only occurs in the evaluation of one of the candidates, which partially confirms Hypothesis 3. However, our results may be due to sampling error and, also, we have not used behaviorally anchored rating scales (BARS) due to experimental design needs. However, SBI is characterized by the use of BARS, which facilitate the accuracy in evaluations (Motowidlo et al., 1992). So, this bias could be reduced for the SBI when the BARS are used (Blackman, 2017).

With regard to the fourth aim of this research, results show that interviewer-interviewee sex similarity produces differences for some dimensions, for the two interview types, and for different rater-candidate sex combinations. Therefore, we can only conclude that the results were inconclusive. This finding suggests that additional studies are needed.

Implications for Practice and Future Research

The findings of this study have implications for the practice of personnel selection interviews. The results suggest that, from an applied point of view, SBI is a more robust method than SCI for identifying a candidate's suitability for a job. The SCI shows some limitations in differentiating between qualified and unqualified candidates. These results converge with the empirical evidence of the superior operational validity of SBI (Huffcutt & Arthur, 1994; McDaniel et al., 1994; Salgado & Moscoso, 2006). Moreover, it is likely that in practice the differences between the candidates are less obvious than in this study, which could contribute to an increase in the limitations of SCI in differentiating between qualified and unqualified candidates. Consequently, we recommend practitioners use SBI rather than SCI when possible.

From a practical point of view it is also relevant to know that interviewers feel more confident when they use a SBI than a SCI. This point is relevant in connection with the finding that SBIs are less frequently used than SCI (Alonso et al., 2015). Taking into account that SBIs are more valid and accurate for making personnel decisions, researchers can also recommend SBI to practitioners because they produce greater self-confidence in their appraisals and decisions.

Our third recommendation to interviewers is to avoid using the information collected previously with other methods (e.g., cognitive tests, personality inventories, and letters of recommendation) during the interview process and decisions. We found that prior information has a bias effect on the interviewer evaluations. This recommendation concurs with meta-analytic findings that showed larger criterion validity for both types of interviews when

interviewers do not have access to prior information about the candidate (McDaniel et al., 1994; Searcy et al., 1993).

With regard to future research, the significant growth in the use of new information technologies (IT) in the selection and assessment processes (e.g., e-recruitment, phone-based interviews, online interviews) suggests that new studies should be conducted to verify the way in which the findings of the present research can be transported to new assessment methods (Aguado, Rico, Rubio, & Fernández, 2016; Bruk-Lee et al., 2016; García-Izquierdo, Ramos-Villagrasa, & Castaño, 2015; Schinkel, van Vianen, & Ryan, 2016). This is especially relevant in the case of the interview, given that more and more companies are conducting online interviews, in which the interviewer and the interviewee communicate only through a computer and in which there is no real interaction between the interviewer and the interviewee (Grieve & Hayes, 2016; Silvester & Anderson, 2003).

Research Limitations

This study has some limitations. As mentioned previously, one limitation is that the samples sizes in some of the experimental conditions are relatively small. Although the overall sample is large ($n = 241$), in some conditions the sample size was thirty individuals. A second limitation is that the interview scripts were acted out by the same two people (one man and one woman). Additional interviewees would be desirable but this would require increasing the experimental sample size accordingly.

In summary, the objective of this research is to shed further light on these four neglected issues concerning the usefulness of the personnel interview as a procedure for making hiring decisions. Our findings suggest that SBI makes raters feel more confident and their appraisals are more accurate. We also found that prior information on the candidate negatively affects the interview outcomes.

Financial Support

The two authors contributed equally. The research reported in this manuscript was supported by Grant PSI2014-56615-P from the Spanish Ministry of Economy and Competitiveness and Grant 2016 GPC GI-1458 from the Consellería de Cultura, Educación e Orden, Xunta de Galicia.

Conflict of Interest

The authors of this article declare no conflict of interest.

References

- Aguado, D., Rico, R., Rubio, V. J., & Fernández, L. (2016). Applicant reactions to social network web use in personnel selection and assessment. *Journal of Work and Organizational Psychology*, 32, 183–190. <http://dx.doi.org/10.1016/j.rpto.2016.09.001>
- Alonso, P. (2011). ¿Producen resultado adverso de género las entrevistas estructuradas de selección de personal? *Journal of Work and Organizational Psychology*, 27, 43–53. <http://dx.doi.org/10.5093/tr2011v27n1a5>
- Alonso, P., Moscoso, S., & Cuadrado, D. (2015). Procedimientos de selección de personal en pequeñas y medianas empresas españolas. *Journal of Work and Organizational Psychology*, 31, 79–89. <http://dx.doi.org/10.1016/j.rpto.2015.04.002>
- Alonso, P., Moscoso, S., & Salgado, J. F. (2017). Structured behavioral interview as a legal guarantee for ensuring equal employment opportunities for women: A meta-analysis. *European Journal of Psychology Applied to Legal Context*, 9, 15–23. <http://dx.doi.org/10.1016/j.ejpal.2016.03.002>
- Anderson, N., Herriot, P., & Hodgkinson, G. P. (2001). The practitioner-researcher divide in Industrial Work and Organizational (IWO) psychology: Where are we now, and where do we go from here? *Journal of Occupational and Organizational Psychology*, 74, 391–411. <http://dx.doi.org/10.1348/096317901167451>
- Anderson, N., Salgado, J. F., & Hülsheger, U. R. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment*, 18, 291–304. <http://dx.doi.org/10.1111/j.1468-2389.2010.00512.x>

- Blackman, M. (2017). Using interviewing in selection. In H. W. Goldstein, E. D. Pulakos, J. Passmore, & C. Semedo (Eds.), *The Wiley Blackwell Handbook of the Psychology of Recruitment, Selection and Employee Retention* (pp. 182–201). New York: John Wiley & Sons, Ltd.
- Bruck-Lee, V., Lanz, J., Drew, E. N., Coughlin, C., Levine, P., Tuzinski, K., & Wrenn, K. (2016). Examining applicant reactions to different media types in character-based simulations for employee selection. *International Journal of Selection and Assessment*, 24, 77–91. <http://dx.doi.org/10.1111/ijisa.12132>
- Campion, M. A. (1978). Identification of variables most influential in determining interviewers' evaluations of applicants in a college placement center. *Psychological Reports*, 42, 947–952. <http://dx.doi.org/10.2466/pr0.1978.42.3.947>
- Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology*, 50, 655–702. <http://dx.doi.org/10.1111/j.1744-6570.1997.tb00709.x>
- Choragwicka, B., & Moscoso, S. (2007). Validez de contenido de una Entrevista Conductual Estructurada. *Revista de Psicología del Trabajo y de las Organizaciones*, 23, 75–92.
- Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology*, 80, 565–579. <http://dx.doi.org/10.1037/0021-9010.80.5.565>
- Dipboye, R. L. (1992). *Selection interviews: Process perspectives*. South-Nashville, TN: Southwestern Publishing Group.
- Dipboye, R. L. (1997). Structured selection interviews: Why do they work? Why are they underutilized? In N. Anderson, & P. Herriott (Eds.), *International Handbook of Selection and Assessment* (pp. 455–474). London: J Wiley.
- Elliott, A. G. P. (1981). Sex and decision making in the selection interview: A real-life study. *Journal of Occupational Psychology*, 54, 265–273. <http://dx.doi.org/10.1111/j.2044-8325.1981.tb00613.x>
- García-Izquierdo, A. L., Ramos-Villagrasa, P. J., & Castaño, A. M. (2015). e-Recruitment, gender discrimination, and organizational results of listed companies on the Spanish Stock Exchange. *Journal of Work and Organizational Psychology*, 31, 155–164. <http://dx.doi.org/10.1016/j.rpto.2015.06.003>
- Goodale, J. G. (1982). *The fine art of interviewing*. Englewood Cliffs, NJ: Prentice Hall.
- Graves, L. M., & Powell, G. N. (1996). Sex similarity, quality of the employment interview and recruiters' evaluation of actual applicants. *Journal of Occupational and Organizational Psychology*, 69, 243–261. <http://dx.doi.org/10.1111/j.2044-8325.1996.tb00613.x>
- Grieve, R., & Hayes, J. (2016). Employment testing online, offline, and over the phone: Implications for e-assessment. *Journal of Work and Organizational Psychology*, 32, 95–101. <http://dx.doi.org/10.1016/j.rpto.2016.04.001>
- Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, 79, 184–190. <http://dx.doi.org/10.1037/0021-9010.79.2.184>
- Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2013). Employment interview reliability: New meta-analytic estimates by structure and format. *International Journal of Selection and Assessment*, 21, 264–276. <http://dx.doi.org/10.1111/ijisa.12036>
- Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2014). Moving forward indirectly: Reanalyzing the validity of employment interviews with indirect range restriction methodology. *International Journal of Selection and Assessment*, 22, 297–309. <http://dx.doi.org/10.1111/ijisa.12078>
- Janz, T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology*, 67, 577–580. <http://dx.doi.org/10.1037/0021-9010.67.5.577>
- Janz, T. (1989). The patterned behavior description interview: The best prophet of the future is the past. In R. W. Eder, & G. R. Ferris (Eds.), *The Employment Interview: Theory, Research, and Practice* (pp. 158–168). Thousand Oaks, CA: Sage.
- Janz, T., Hellervik, L., & Gilmore, D. C. (1986). *Behavior description interviewing: New, accurate cost effective*. Englewood Cliffs, NJ: Prentice Hall.
- Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology*, 65, 422–427. <http://dx.doi.org/10.1037/0021-9010.65.4.422>
- Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, 67, 241–293. <http://dx.doi.org/10.1111/peps.12052>
- Liu, X., Potočník, K., & Anderson, N. (2016). Applicant reactions to selection methods in China. *International Journal of Selection and Assessment*, 24, 296–303. <http://dx.doi.org/10.1111/ijisa.12148>
- Macan, T. H., & Dipboye, R. L. (1990). The relationship of interviewers' preinterview impressions to selection and recruitment outcomes. *Personnel Psychology*, 43, 745–768. <http://dx.doi.org/10.1111/j.1744-6570.1990.tb00681.x>
- McCarthy, J. M., Van Iddekinge, C. H., & Campion, M. A. (2010). Are highly structured job interviews resistant to demographic similarity effects? *Personnel Psychology*, 63, 325–359. <http://dx.doi.org/10.1111/j.1744-6570.2010.01172.x>
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, 79, 599–616. <http://dx.doi.org/10.1037/0021-9010.79.4.599>
- Moscoso, S., & Salgado, J. F. (2001). Psychometric properties of a structured behavioral interview to hire private security personnel. *Journal of Business and Psychology*, 16, 51–59. <http://dx.doi.org/10.1023/A:1007835704733>
- Motowidlo, S. J., Carter, G. W., Dunnette, M. D., Tippins, N., Werner, S., Burnett, J. R., & Vaughan, M. J. (1992). Studies of the structured behavioral interview. *Journal of Applied Psychology*, 77, 571–587. <http://dx.doi.org/10.1037/0021-9010.77.5.571>
- Paunonen, S. V., Jackson, D. N., & Oberman, S. M. (1987). Personnel selection decisions: Effects of applicant personality and the letter of reference. *Organizational Behavior and Human Decision Processes*, 40, 96–114. [http://dx.doi.org/10.1016/0749-5978\(87\)90007-0](http://dx.doi.org/10.1016/0749-5978(87)90007-0)
- Rodríguez, A. (2016). Predictive validity and adverse impact of the structured behavioral interview in the public sector. *Journal of Work and Organizational Psychology*, 32, 75–85. <http://dx.doi.org/10.1016/j.rpto.2016.04.003>
- Sacco, J. M., Scheu, C. R., Ryan, A. M., & Schmitt, N. (2003). An Investigation of Race and Sex Similarity Effects in Interviews: A Multilevel Approach to Relational Demography. *Journal of Applied Psychology*, 88, 852–865. <http://dx.doi.org/10.1037/0021-9010.88.5.852>
- Salgado, J. F. (2007). Análisis de utilidad económica de la Entrevista Conductual Estructurada en la selección de personal de la administración general del País Vasco. *Revista de Psicología del Trabajo y de las Organizaciones*, 23, 139–154.
- Salgado, J. F., & Moscoso, S. (1995). Validez de las entrevistas conductuales estructuradas. *Revista de Psicología del Trabajo y las Organizaciones*, 11, 9–24.
- Salgado, J. F., & Moscoso, S. (1997). Seguridad de las valoraciones en la entrevista conductual estructurada y en la entrevista convencional. San Sebastián, Spain.
- Salgado, J. F., & Moscoso, S. (1998). *Interviewer self-confidence in assessments using behavior description and conventional interviews*. San Francisco, CA.
- Salgado, J. F., & Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work and Organizational Psychology*, 11, 299–324. <http://dx.doi.org/10.1080/13594320244000184>
- Salgado, J. F., & Moscoso, S. (2006). Utiliser les entretiens comportementaux aux structures pour la sélection du personnel? In C. Lévy-Leboyer, C. Louche, & J. P. Rolland (Eds.), *RH: Les apports de la psychologie du travail* (1) (pp. 195–207). Paris: Éditions d'Organisation.
- Salgado, J. F., & Moscoso, S. (2011). *La entrevista conductual estructurada de selección de personal: Teoría práctica y rentabilidad* (3ª edición). Madrid: Pirámide.
- Salgado, J. F., Moscoso, S., & 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 del Trabajo y de las Organizaciones*, 20, 107–139.
- Schinkel, S., van Vianen, A. E., & Ryan, A. M. (2016). Applicant reactions to selection events: Four studies into the role of attributional style and fairness perceptions. *International Journal of Selection and Assessment*, 24, 107–118. <http://dx.doi.org/10.1111/ijisa.12134>
- Searcy, C. A., Woods, P. N., Gatewood, R., & Lace, C. (1993). *The validity of structured interviews: A meta-analytical search for moderators*. San Francisco, CA.
- Silvester, J., & Anderson, N. (2003). Technology and discourse: A Comparison of face-to-face and telephone employment interviews. *International Journal of Selection and Assessment*, 11, 206–214. <http://dx.doi.org/10.1111/1468-2389.00244>
- Steiner, D. D., & Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology*, 81, 134–141. <http://dx.doi.org/10.1111/j.0965-075X.2004.00265.x>