

**AMPLIACIÓN DE LA BATERÍA DE MANEJO DE IMPRESIONES:  
VALIDACIÓN DE NUEVOS ÍTEMS Y RE-EVALUACIÓN DE LOS  
ANTERIORES**

---

**Trabajo de Fin de Máster**

**Máster Universitario en Gestión y Desarrollo de los Recursos Humanos**

**Alumno: Samuel José Pacheco Fernández**

**Tutorizado por: Luis Fernando Díaz Vilela**

**Curso académico 2022-2023**

## **Índice**

<b>Resumen</b>	<b>2</b>
<b>Abstract</b>	<b>2</b>
<b>Introducción teórica</b>	<b>3</b>
<b>Hipótesis</b>	<b>7</b>
<b>Método</b>	<b>9</b>
<b>Participantes</b>	<b>9</b>
<b>Instrumentos</b>	<b>10</b>
<b>Procedimiento</b>	<b>13</b>
<b>Análisis de datos</b>	<b>14</b>
<b>Resultados</b>	<b>14</b>
<b>Conclusiones</b>	<b>23</b>
<b>Discusión</b>	<b>24</b>
<b>Bibliografía</b>	<b>27</b>

## **Resumen**

Uno de los peligros más relevantes de los procesos de selección de personal es el manejo de impresiones, conocido como la distorsión intencionada de características personales propias, con el objetivo de quedar bien. Este estudio supone la continuación del trabajo de Rodríguez (2022) en el desarrollo de una batería de ítems que midan eficazmente el manejo de impresiones. La batería permitirá generar cuestionarios formados por conjuntos de ítems elegidos aleatoriamente, haciendo más difícil el entrenamiento para burlar su validez, y por lo tanto, haciendo más confiable su uso en procesos de selección. Se recogieron las respuestas de 123 estudiantes a tres conjuntos de preguntas: los usados por Rodríguez (2022) (cuestionario Q), un conjunto nuevo elegido aleatoriamente de la misma base de preguntas y los que componen la escala de manejo de impresiones del cuestionario BIDR de Paulhus. Se sometió a los participantes a dos condiciones de respuesta: quedar bien socialmente y ser totalmente honesto/a. En base a los resultados se aumenta en 17 ítems la batería desarrollada. Estas preguntas cuentan con validez convergente con la escala BIDR y el cuestionario Q. Además, se encontró una coincidencia del 80,43% del cuestionario Q entre el estudio actual y el original.

## **Abstract**

One of the most relevant threats to the personnel selection process is impression management, known as the intentional distortion of one's personal characteristics in order to make oneself look good. This study is a continuation of the work of Rodriguez (2022) in the development of a pool of items that effectively measure

impression management. The pool will make it possible to generate questionnaires made up of randomly chosen sets of items, making it more difficult to train to circumvent their validity, and thus making their use in selection processes more reliable. The responses of 123 students to three sets of questions were collected: those used by Rodriguez (2022) (questionnaire Q), a new set randomly selected from the same question base, and those corresponding to the impression management scale of the Paulhus BIDR questionnaire. Participants were subjected to two response conditions: looking good socially and being totally honest. Based on the results, the pool developed was increased by 17 items. These questions have convergent validity with the BIDR scale and the Q questionnaire. In addition, a coincidence of 80.43% of the Q questionnaire was found between the current study and the original one.

### **Introducción teórica**

La deseabilidad social se refiere a la tendencia de las personas a presentarse de manera favorable ante los demás, a fin de causar una buena impresión y evitar el rechazo social (Paulhus y Reid, 1991). Este sesgo también es conocido con otros nombres, como el sesgo de deseabilidad social, la respuesta socialmente deseable y la respuesta socialmente aceptable (Podsakoff et al., 2003).

Este fenómeno se lleva estudiando desde hace más de 60 años, tan pronto como empezaron a surgir pruebas de personalidad. Estas pruebas son susceptibles de varios sesgos, como el sesgo de aquiescencia, de respuesta extrema o de respuesta aleatoria, cobrando el sesgo de deseabilidad social especial relevancia con la introducción de escalas de validez en la prueba de personalidad del MMPI (Paulhus, 2020).

La deseabilidad social es especialmente relevante en la selección de personal y las entrevistas laborales (Ones et al., 1996). Por ejemplo, en la selección de personal, los candidatos pueden presentarse de manera más favorable de lo que realmente son para aumentar sus posibilidades de ser contratados (Levashina y Champion, 2007). Por si fuera poco, el falseamiento de las respuestas puede ir más allá, a través del entrenamiento. Este es capaz de dejar a candidatos promedio y honestos sin posibilidades de competir con los que realizaron un entrenamiento (Miller y Barret, 2008).

Entonces, ¿siempre que se habla del sesgo de deseabilidad se habla de lo mismo? O, ¿puede que en realidad comprenda diferentes fenómenos? Y por último, ¿Es posible proteger la validez de las pruebas de personalidad en los procesos de selección?

Existen diversas concepciones teóricas acerca de la deseabilidad social y desde las que se comprenden diferentes subfactores. Millham y Jacobson (1978, citado en Paulhus 1984), proponen desde sus análisis el modelo de Atribución versus Negación. El primer factor recoge la tendencia a responder atribuyéndose características deseables a uno mismo, mientras que la segunda comprende la negación de características indeseables al yo. Sin embargo, esta observación parece deberse a una mala manipulación de los datos. Usando correctamente la fórmula de Spearman-Brown se encontró una alta correlación de ambos subfactores, muy cercana a la fiabilidad del test en sí. Esto parece indicar que en realidad corresponden a medidas equivalentes del mismo constructo (Paulhus, 1984), obteniendo otros investigadores resultados similares (Ramanaiah y Martin, 1980).

La propuesta más aceptada actualmente es el modelo bifactorial del Autoengaño versus Manejo de Impresiones (Paulhus, 2020). Se ha venido proponiendo desde los años 40 por diversos autores. Algunos se refieren a estos conceptos de otra manera, como “falseamiento consciente e inconsciente”, “auto-ilusión”, “autoengaño” (self-deception) y “engaño a otros” (other-deception), pero se mantienen consistentes con el significado (Frenkel-Brunswik, 1939; Meehl y Hathaway, 1946; Sackeim y Gur, 1978; Millham y Kellogg, 1984; Paulhus, 1984). El factor Autoengaño o Alfa (Block, 1965; Wiggins, 1964), se refiere a un sesgo inconsciente de evaluación en los autoinformes, en los que se evalúan de manera más positiva de lo que son en realidad, según su propia visión del consenso social. Como antítesis, el factor Manejo de Impresiones o Gamma (Block, 1965; Wiggins, 1964) hace referencia a la tendencia a falsificar las respuestas intencionadamente (Damarin y Messick, 1965).

Ambas dimensiones pueden influir en las respuestas a una evaluación, siempre que el evaluado sea consciente de que su conducta está siendo analizada. Sin embargo, el manejo de impresiones juega un papel más relevante en los cuestionarios de personalidad, pues es un sesgo consciente capaz de cambiar según la situación (Paulhus, 1984). Es aún mayor el daño potencial a la fiabilidad cuando se van a utilizar las respuestas para su beneficio o perjuicio, como en es el caso de un proceso de selección de personal (Tracey, 2016). Se sabe que es común que los candidatos intenten causar una buena impresión. Pese a ello, una manipulación excesiva puede llevar a una evaluación inadecuada de su idoneidad y, por lo tanto, a una decisión equivocada (Hausknecht et al., 2004).

Por otro lado, algo que presenta una amenaza para la validez de las pruebas de manejo de impresiones es el entrenamiento (Hausknecht et al., 2004; Levashina y Campion, 2007). El entrenamiento en esta clase de pruebas conseguiría que las personas superen la prueba de honestidad, aunque realmente no estén siendo honestas. Esto supone que podrían contestar una prueba de personalidad de manera sesgada sin que el evaluador pueda detectarlo. En un estudio llevado a cabo por Landers et al. (2011), se analizan algunos mitos relacionados con las instrucciones dadas por los entrenadores de pruebas. Por ejemplo, algunos entrenadores sugieren responder con las puntuaciones más extremas, como un 1 o un 5, al identificar la opción "deseable". Este sesgo de respuesta extrema se utiliza en algunas pruebas para detectar el manejo de impresiones. Sin embargo, en estas pruebas, las pautas son contrarias a las utilizadas en las pruebas convencionales, ya que las respuestas extremas sugieren una tendencia a modificar la imagen. Estas pruebas suelen contener afirmaciones de "nunca" o "siempre", lo que las hace fácilmente identificables y entrenables. Estas normas son aplicables a las pruebas en las que se suman las respuestas y se obtienen puntuaciones directas de cada uno de los factores de personalidad y/o de las pruebas de manejo de impresiones.

Por otra parte, se ha estudiado la posible relación entre el sexo y el manejo de impresiones, pero no existe consenso. Algunos estudios muestran mayor nivel de manejo de impresiones por parte de las mujeres que de los hombres en un contexto laboral (Forsyth y McMillan, 2009) mientras que otros encuentran que no existen diferencias (Kazmi y Amjad, 2014). Además, desde otros focos de investigación, existen hallazgos que podrían tener relación. Varios estudios muestran mayores puntuaciones

en el rasgo de cordialidad del Big Five (Costa y McCrae, 1999) en mujeres que en hombres (ej.: Feingold, 1994; Schmitt, 2009; Weisberg, 2011). La cordialidad, también llamada amigabilidad, hace alusión a comportamientos prosociales, como ser amable o ayudar a los demás (Jones et al., 2011), lo que en relación con el manejo de impresiones, supone comportarse de manera deseable.

En conclusión, la deseabilidad social, y muy especialmente el manejo de impresiones, es un fenómeno complejo que debe tenerse en cuenta en la evaluación de la personalidad y la toma de decisiones basadas en ella. Ante este riesgo, Rodríguez (2022) ha planteado como propuesta preventiva el desarrollo de una batería amplia de ítems que midan el Manejo de Impresiones. Esto permitiría utilizar cada vez una cantidad determinada de ítems aleatorios, con el objetivo de dificultar la eficacia de un posible entrenamiento para burlar la prueba.

El presente estudio supone la continuación del trabajo de Rodríguez (2022). De esta manera, se procede al estudio de la validez de nuevos ítems que midan el Manejo de Impresiones y a la re-evaluación de los ítems propuestos anteriormente. Con el objetivo de proteger su validez, la prueba resultante, que llamaremos R, y su conjunto de preguntas no serán publicadas.

### **Hipótesis**

El cuestionario "R", formado por los nuevos ítems debe tener una buena fiabilidad, por lo tanto:

Hipótesis 1: Se espera que los nuevos ítems saturen en un único componente.



Hipótesis 2: La varianza que explica el componente principal será equiparable a la de la escala BIDR (Paulhus, 1991).

Además, en este aspecto, se contrastará también la fiabilidad de la escala de Rodríguez (2022) (Escala Q) sin el descarte de ítems realizado, explorando si existen diferencias en los ítems incluidos y descartados en ambos análisis. De esta manera:

Hipótesis 3: Los ítems descartados e incluidos en las iteraciones necesarias de análisis de componentes principales en ambas investigaciones serán los mismos para la prueba Q.

Para comprobar la validez de la prueba "R", es decir, que mida realmente el manejo de impresiones:

Hipótesis 4: Se espera que la escala R tenga una alta validez convergente tanto con la subescala de Manejo de Impresiones del cuestionario BIDR (Paulhus, 1991) como con la escala Q.

Concretamente se espera, por un lado, obtener correlaciones altas y positivas entre los diferentes cuestionarios para las mismas condiciones, es decir, responder para quedar bien o responder honestamente. Y por el otro lado, se espera obtener correlaciones nulas entre las diferentes condiciones (quedar bien/ser honesto) de los diferentes cuestionarios (nuevos ítems, ítems previos y BIDR).

Dado que las diferentes instrucciones de comportamiento de respuesta deben haber tenido efecto diferencial en las puntuaciones:

Hipótesis 5: Se espera que las puntuaciones para cada par de ítems en las tres escalas (Q, R y BDR) sean superiores en la condición de quedar bien que en la condición de honestidad.

Por último, en base a las posibles diferencias de puntuaciones entre hombres y mujeres:

Hipótesis 6: Se espera que las puntuaciones de las mujeres sean superiores a la de los hombres en todas las escalas en situación de honestidad.

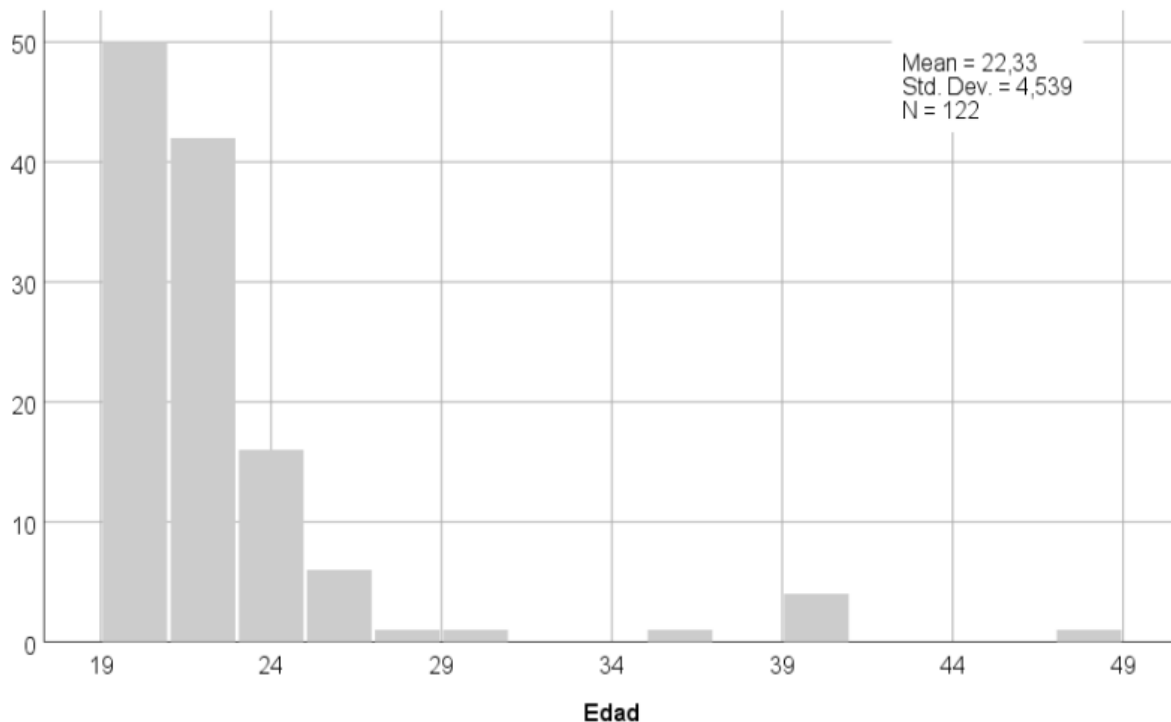
## **Método**

### **Participantes**

La muestra está formada por 123 personas pertenecientes al Grado de Psicología y al Máster de Gestión y Desarrollo de los Recursos Humanos de la Universidad de La Laguna. Concretamente fueron 100 mujeres, 22 hombres más una persona sin identificar. El rango de edad se distribuye mayoritariamente alrededor de los 21 años, como se puede observar en la Figura 1.

**Figura 1**

*Distribución de la edad*



## **Instrumentos**

Se aplicaron tres pruebas en dos condiciones cada una. Las pruebas consistieron en el Inventario equilibrado de respuesta deseable (BIDR), desarrollada por Paulhus (1991), el cuestionario Q anteriormente contrastado por Rodríguez (2022) y el nuevo cuestionario de prueba, que llamaremos "R". Por otro lado, las condiciones fueron "quedar bien" y "ser honesto".

El BIDR contiene dos subescalas, autoengaño y manejo de impresiones, con 20 ítems para cada factor, 40 en total. La fiabilidad demostrada para este cuestionario oscila entre  $\alpha = 0.67$  y  $\alpha = 0.77$  para la escala de autoengaño y entre  $\alpha = 0.77$  y  $\alpha = 0.85$  para la de manejo de impresiones. Para este cuestionario sólo se empleó la subescala que recoge el manejo de impresiones de la versión en español de Zerpa (2020). Esta

adaptación contiene 21 ítems, ya que se añadió un ítem neutral con el objetivo de que la escala no acabara con una pregunta emocionalmente negativa. El tipo de respuesta consiste en tres alternativas: verdadera/afirmativa, neutral y falsa/negativa. De esta manera, las respuestas positivas se puntúan con un 1, las neutras con un 2 y las negativas con un 3.

Los cuestionarios Q y R fueron elaborados a partir de la base de datos "International Personality Item Pool" (IPIP), desarrollada por Goldberg (1999). De la base de datos de 3320 ítems se seleccionaron 209 que podían adaptarse para evaluar el manejo de impresiones. Esta adaptación consistió en redactar las preguntas de manera que las respuestas rotundamente afirmativas o negativas, según el caso, fueran poco probables en la vida real. Por ejemplo, a partir del ítem "Hacer cosas impropias" se generó la afirmación "Nunca he hecho cosas impropias".

De esos 209 ítems adaptados Rodríguez Rodríguez (2022) seleccionó 46 al azar para conformar su cuestionario Q. En este trabajo, ese cuestionario se dividió en dos, con 23 preguntas cada uno, que llamaremos Q\_A y Q\_B, correspondientes a las hojas 1 y 2 del cuestionario original, respectivamente. Además, se generó un nuevo cuestionario, R, con 23 ítems elegidos al azar entre los 163 restantes de la base de datos (209 -46). Para cada pregunta, los participantes debían responder en base al nivel de frecuencia o de veracidad que la afirmación tiene para él/ella. La escala de respuesta consta de tres alternativas. Por lo general: "verdadero", "término medio" y "falso", pero en algunos casos estos anclajes variaron para ajustarse a la afirmación: "siempre", "a

veces” “ nunca”; “sí”, “término medio” “no”; etc. Para cada ítem, las alternativas *a* puntuaron con un 1, las *b* puntuaron con un 2 y las *c* con un 3.

Respecto a las condiciones, en la de “honestidad”, se pidió a los participantes que respondieran de la manera más sincera posible. Concretamente se les indicó: “Por favor, responde de manera HONESTA Y SINCERA. No señale sus respuestas pensando en lo que “es bueno” o “lo que interesa” para impresionar al examinador”. Por otro lado, en la condición de “quedar bien”, se pidió a los participantes que contestaran a las preguntas intentando quedar mejor con el evaluador. Concretamente: “Por favor, responde pensando en AQUELLAS ALTERNATIVAS QUE TE HARÍAN QUEDAR MEJOR CON EL ENTREVISTADOR, AUNQUE PARA ELLO TENGAS QUE MENTIR. “.

De esta manera, el instrumento utilizado consiste en dos formas del cuestionario. Estas formas dependían del tipo de cuestionario Q que le correspondiera, aleatoriamente, a cada participante. En la Tabla 1 se observa detalladamente el contenido de ambos cuestionarios.

**Tabla 1**  
*Contenido de las formas A y B del autoinforme*

<b>Cuestionario Austria (A)</b>	<b>Cuestionario Bahrein (B)</b>
<b>Instrucciones “honestidad”</b>	<b>Instrucciones “honestidad”</b>
Cuestionario Q (Forma A)	Cuestionario Q (Forma B)
Cuestionario R	Cuestionario R
BIDR	BIDR
<b>Instrucciones “quedar bien”</b>	<b>Instrucciones “quedar bien”</b>
Cuestionario Q (Forma A)	Cuestionario Q (Forma B)
Cuestionario R	Cuestionario R
BIDR	BIDR

## **Procedimiento**

Este estudio fue ejecutado de forma prospectiva y transversal. Además, se llevó a cabo un muestreo por conveniencia, sin cupos grupales. El autoinforme se compartió digitalmente a través de la dirección de email electrónico del tutor. Cada e-mail incluía el enlace a una de las dos formas del cuestionario, enviando a cada mitad del alumnado una de las formas aleatoriamente.

El alumnado participante consistió en estudiantes de la asignatura del profesor Luis Fernando Díaz Vilela. Estos fueron recompensados con la suma de décimas de punto en la nota final de la asignatura impartida por él. Los alumnos del Máster de Gestión y Desarrollo de los RRHH no fueron retribuidos.

El cuestionario debía cumplimentarse en la red a través de la plataforma de Google Forms ([drive.google.com](https://drive.google.com)).

Previa a su cumplimentación, el investigador se comprometió a utilizar las respuestas de manera exclusiva para los propósitos de la investigación y de acuerdo con lo establecido en la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. Además, para poder continuar, cada participante debió afirmar que participaba de forma voluntaria. Por último, fue obligatorio responder a todas las preguntas para poder entregar el cuestionario.

En cuanto al control ejercido en la investigación, se trató de un estudio observacional de encuesta. Esto significa que no se manipularon las variables estudiadas ni se controlaron factores externos como el espacio físico, la presencia de

otras personas o el tiempo de respuesta. En concreto, no se exigió que los participantes completaran el cuestionario de manera ininterrumpida.

Finalmente, las variables a estudiar fueron el sexo, como variable intergrupala, además de quedar bien y ser honesto como variables intrasujeto.

### **Análisis de datos**

Para analizar los datos se utilizó el programa SPSS (IBM SPSS versión 26) además de la plataforma de Google Drive (drive.google.com) para preparar la hoja de cálculo utilizada, que fue exportada directamente de Google Forms.

Se llevaron a cabo una serie de análisis descriptivos así como el cálculo de coeficientes de correlación de Pearson, análisis factoriales de componentes principales, coeficientes de correlación intraclase y varios MANOVA. Para todas las pruebas estadísticas se asumirá un nivel mínimo de  $\alpha = 0.05$  para considerarlas significativas.

### **Resultados**

Al menos que se indique lo contrario, las pruebas fueron realizadas sobre las respuestas de los cuestionarios en condición de honestidad. Esto es debido a que la motivación del estudio es construir una escala que mida el manejo de impresiones cuando las personas deben ser sinceras.

Previo a los análisis factoriales se comprobó la idoneidad de las matrices de correlaciones. Se comprobó que los determinantes eran diferentes de cero. Se comprobó, también, que las pruebas de Kaiser–Meyer–Olkin (KMO) tenían un valor

adecuado y, con la prueba de Bartlett, que las matrices no eran de identidad (las correlaciones fuera de la diagonal eran diferentes de cero). Como se puede ver en la Tabla 2, todos los cuestionarios superaron satisfactoriamente los tres criterios. Los determinantes resultaron muy cercanos a cero, lo que indica que las correlaciones en las matrices son altas. Las medidas de adecuación muestral (KMO) resultaron muy adecuadas, indicando que hay una alta proporción de varianza común entre las variables (Tabachnick y Fidell, 2019). Finalmente, las pruebas de esfericidad de Bartlett resultaron significativas, indicando que las matrices no son de identidad.

**Tabla 2**

*Resultados de los contrastes KMO y esfericidad de Bartlett*

Cuestionario	Factorial 1					Factorial 2				
	Det.	KMO	Bartlett			Det.	KMO	Bartlett		
			$\chi^2$	***	g.l.			$\chi^2$	***	g.l.
<b>R</b>	0.01E-5	0,860	1123,709	***	253	0.002	0,909	742,957	***	105
<b>Q (Forma A)</b>	6.80E-6	0,806	660,366	***	253	8.85E-5	0,841	536,614	***	136
<b>Q (Forma B)</b>	1.40E-6	0,816	653,831	***	253	4.73E-6	0,833	602,852	***	210
<b>BIDR</b>	5.84E-5	0,842	1112,971	***	210	0.000	0,858	1050,651	***	190

\*\*\*p < .001

A la vista de estas propiedades, se procedió a realizar los análisis de componentes principales. Para cada cuestionario se siguió un proceso iterativo mediante el que se fueron descartando preguntas para las que el componente principal explicaba menos de un 10% de su varianza, o lo que es lo mismo, que saturaran por debajo de 0,33. Para cada uno de los cuestionarios bastó con dos análisis para alcanzar este criterio.



En primer lugar, para el cuestionario R, el componente principal hallado en el primer análisis explicó el 31,83% de la varianza común de los 23 ítems. Siguiendo el criterio anterior se excluyeron 6 ítems, por lo que se realizó un segundo análisis sobre los 17 ítems restantes. Se obtuvo un componente principal que explica el 42,71% de la varianza común. Este componente explica más del 10% de la varianza de todos y cada uno de los ítems, por lo que no fue necesario excluir ninguno. Por lo tanto, la Hipótesis 1, que afirma la saturación de todos los ítems en un componente principal se confirma parcialmente.

En segundo lugar, para el cuestionario BIDR se halló un componente principal que explica el 36,44 % de la varianza. Se excluyó un ítem y se realizó de nuevo el análisis. Esta vez se obtuvo un componente principal que explica el 37,95% de la varianza, sin ser necesario excluir más preguntas. Dado que la varianza explicada por la escala R no solo es similar sino superior a la BIDR se confirma la Hipótesis 2.

En tercer lugar, para la prueba Q, se halló un componente principal en cada una de sus formas, explicando un 31,90% de la varianza de la Forma A y un 35,56% de la varianza de la Forma B. Se descartaron 5 y 2 ítems respectivamente, manteniendo un total de 17 en la Forma A y 21 en la Forma B. Se realizó un segundo análisis de componentes principales para cada forma, obteniéndose un componente principal en cada caso, que explican el 40,31% y 38,60% de la varianza común respectivamente.

Posteriormente, a partir de una comunicación personal (M. Rodríguez, 10 de abril de 2023) se elabora la Tabla 3 para contrastar los ítems rechazados y aceptados de la prueba Q en este estudio y en el análisis anterior de Rodríguez (2022). Ambos análisis

coinciden en la estimación del 80,43 % de los ítems, quedando 9 ítems de los 46 iniciales en contradicción. Por lo tanto, respecto la Hipótesis 3: “los ítems descartados e incluidos a partir de los análisis de componentes principales en ambas investigaciones serán los mismos para la prueba Q”, queda, mayoritariamente, confirmada.

**Tabla 3**

*Rechazo/aceptación de los ítems por investigación*

Ítem	FORMA A		FORMA B	
	Rodríguez	Actual	Rodríguez	Actual
1			X	X
2	X	X	X	X
3	X	X	X	X
4				X
5		X		X
6	X	X		X
7				X
8	X	X	X	X
9	X	X	X	
10			X	X
11	X	X	X	X
12		X		
13	X	X	X	X
14	X	X	X	X
15		X	X	X
16	X	X		X
17	X	X	X	X
18	X	X	X	X
19	X	X	X	X
20	X	X	X	X
21	X	X	X	X
22			X	X
23	X	X	X	X

*X = Aceptado*

Para contrastar la validez concurrente del cuestionario R, se realizaron múltiples análisis de correlación de Pearson entre las escalas R, Q y BIDR para las dos condiciones (honestidad y quedar bien). Tal como se puede observar en la Tabla 4, las puntuaciones

de la escala R correlacionaron positiva y significativamente con los otros cuestionarios en la misma condición. Por lo tanto, se confirma la Hipótesis 4, en cuanto a la validez convergente con las prueba BIDR y Q.

**Tabla 4**

*Correlaciones del cuestionario R con Q (A y B) y BIDR, dentro de cada condición.*

	<b>Honestidad</b>			<b>Quedar bien</b>		
	BIDR	Q_A	Q_B	BIDR	Q_A	Q_B
<b>R</b>						
Honestidad	0,89 ***	0,88 ***	0,874 ***	0,03 n.s.	0,22 n.s.	0,06 n.s.
Quedar bien	0,05 n.s.	0,07 n.s.	-0,1 n.s.	0,75 ***	0,75 ***	0,79 ***
<b>g.l.</b>	123	65	58	123	65	58

\*\*\* p < .001

Para comprobar el efecto esperado de las instrucciones sobre las respuestas a cada pregunta, se realizaron pruebas t de Student para muestras apareadas. Estas pruebas contrastaron las puntuaciones de cada pregunta entre condiciones. Como se muestra en la Tabla 5, se obtuvo diferencias significativas para todos los pares, con puntuaciones menores en la condición de honestidad que en la de quedar bien salvo para las preguntas inversas. De esta manera, se confirma la Hipótesis 5, que afirma las mayores puntuaciones de los ítems en la condición de quedar bien que en la de honestidad.

**Tabla 5***Pruebas t entre condiciones para los ítems componentes de la puntuación en R.*

<b>Honestidad vs. Quedar bien</b>		
	<b>t (122)</b>	
Ítem 02	-10.674	***
Ítem 07	-11.687	***
Ítem 09	-8.957	***
Ítem 11	-7.648	***
Ítem 12	-8.113	***
Ítem 13	-9.141	***
Ítem 15	-10.649	***
Ítem 16	-8.311	***
Ítem 17	-12.285	***
Ítem 04 <sup>a</sup>	8.390	***
Ítem 06 <sup>a</sup>	6.555	***
Ítem 08 <sup>a</sup>	6.816	***
Ítem 14 <sup>a</sup>	10.205	***
Ítem 18 <sup>a</sup>	9.869	***
Ítem 19 <sup>a</sup>	10.019	***

<sup>a</sup> Ítems inversos\*\*\*  $p < .001$ 

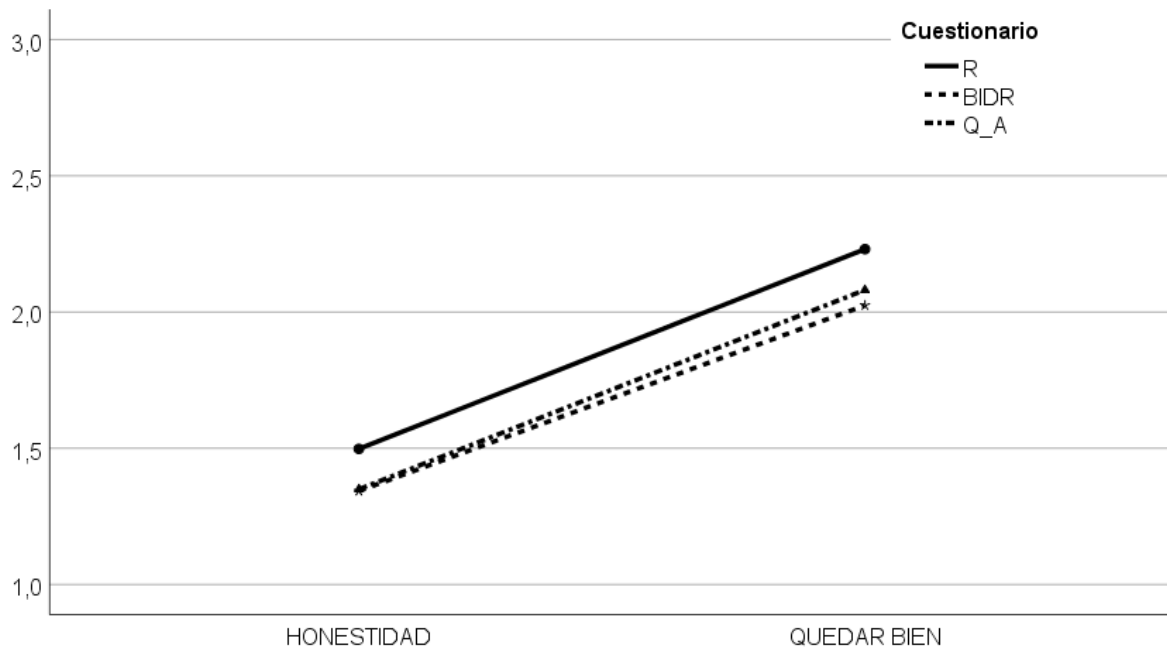
Con el fin de analizar los efectos simples y de interacción del cuestionario y la condición sobre las respuestas, se realizaron dos MANOVA 3x2, uno sobre cada forma del macro-cuestionario en línea. Así, se utilizó como variables intrasujeto, el cuestionario, con tres niveles (Q, R o BIDR) y la condición, con dos niveles (honestidad vs. quedar bien). En cada análisis se emplearon los datos de los participantes que contestaron cada forma del cuestionario Q (Q\_A y Q\_B) respectivamente.

En el primer análisis, con los datos de los participantes que contestaron Q\_A, se halló un efecto significativo del tipo de cuestionario ( $F(2, 128) = 37.689, p < .001, \eta^2 = 0.371$ ) y de la condición ( $F(1, 64) = 173.851, p < .001, \eta^2 = 0.731$ ), pero no se halló significación en la interacción cuestionario x condición ( $F(2, 128) = 1,904, n.s., \eta^2 = 0.153$ ). Concretamente, se obtuvieron puntuaciones mayores en la condición de quedar

bien que en la de ser honesto (Figura 2). Los contrastes indicaron que las puntuaciones en R son superiores a las de cada uno de los otros dos cuestionarios, aunque las diferencias sean pequeñas (ver la Tabla 6, *Muestra con Q\_A*).

## Figura 2

*Efectos de Cuestionario y Condición con la versión A de Q.*



Al aplicar el mismo análisis a los datos de los participantes que contestaron Q\_B, también se halló un efecto significativo del tipo de cuestionario ( $F(2, 114) = 24.912, p < .001, \eta^2 = 0.304$ ) y, sobre todo, de la condición,  $F(1, 57) = 116.597, p < .001, \eta^2 = 0.672$ . En este caso tampoco se encontró significación en la interacción entre tipo de cuestionario y condición, indicando que los tres cuestionarios son igualmente sensibles a las instrucciones. En la Tabla 6 puede verse que, también con esta submuestra se cumple que el cuestionario R obtiene puntuaciones, aunque ligeramente, significativamente mayores que BIDR y Q\_B.

**Tabla 6**

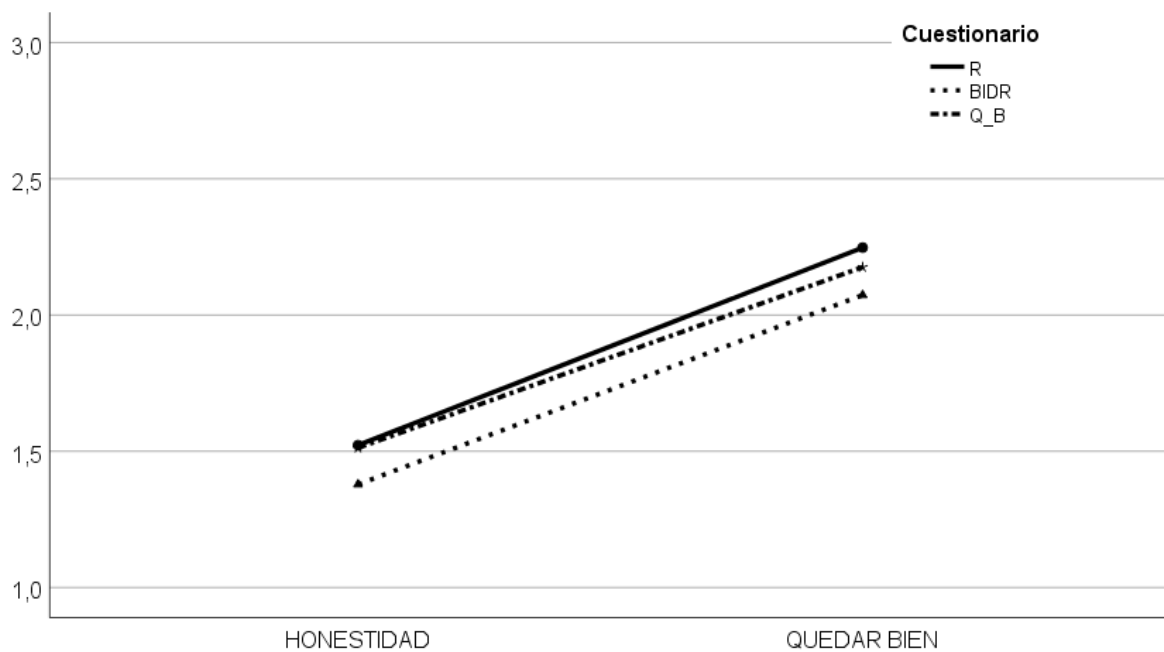
Contrastes de las puntuaciones del cuestionario R con cada uno de los otros dos cuestionarios.

		g.l.	F		$\eta^2$
Muestra con Q_A (Figura 2)	BIDR	1, 64	67.922	***	0,515
	Q_A	1, 64	81.825	***	0,561
Muestra con Q_B (Figura 3)	BIDR	1, 57	45.496	***	0.444
	Q_B	1, 57	30.948	***	0.352

\*\*\*  $p < .001$

**Figura 3**

Efectos de Cuestionario y Condición con la versión B de Q.



El efecto principal de la condición hallada en ambas muestras apoya la hipótesis 5, respecto a las mayores puntuaciones en la condición de quedar bien que en condición de honestidad.

Para analizar la posible diferencia en el manejo de impresiones entre hombres y mujeres se realizaron pruebas t de Student para cada cuestionario con el sexo como variable intergrupar. Como se puede observar en la Tabla 7, solo se hallaron diferencias estadísticamente significativas en el cuestionario BIDR en condición de quedar bien (t

(120) = 2,416;  $p < .001$ ). Concretamente, puntuaron los Hombres ( $M = 2,21$ ;  $SD = 0.324$ ) por encima de las Mujeres ( $M = 2.01$ ;  $SD = 0.55$ ). Por lo tanto, se rechaza la Hipótesis 6, en la que se afirma una puntuación de las mujeres superior a la de los hombres en condición de honestidad.

**Tabla 7**  
*Diferencias por género.*

			$\bar{X}$	SD	t	g.l.
R	Honestidad	Hombres	1,57	0,447	0,630	120
		Mujeres	1,50	0,437		
	Quedar bien	Hombres	2,26	0,304	0,360	120
		Mujeres	2,24	0,292		
BIDR	Honestidad	Hombres	1,51	0,508	1,585	26.043 <sup>a</sup>
		Mujeres	1,33	0,368		
	Quedar bien	Hombres	2,21	0,314	<b>2,416 *</b>	<b>120</b>
		Mujeres	2,01	0,355		
Q_A	Honestidad	Hombres	1,46	0,418	1,031	62
		Mujeres	1,33	0,377		
	Quedar bien	Hombres	2,13	0,281	0,644	62
		Mujeres	2,07	0,300		
Q_B	Honestidad	Hombres	1,59	0,360	0,681	56
		Mujeres	1,50	0,406		
	Quedar bien	Hombres	2,26	0,139	1,387	56
		Mujeres	2,16	0,228		

<sup>a</sup>T de Welch para muestras con varianzas diferentes.

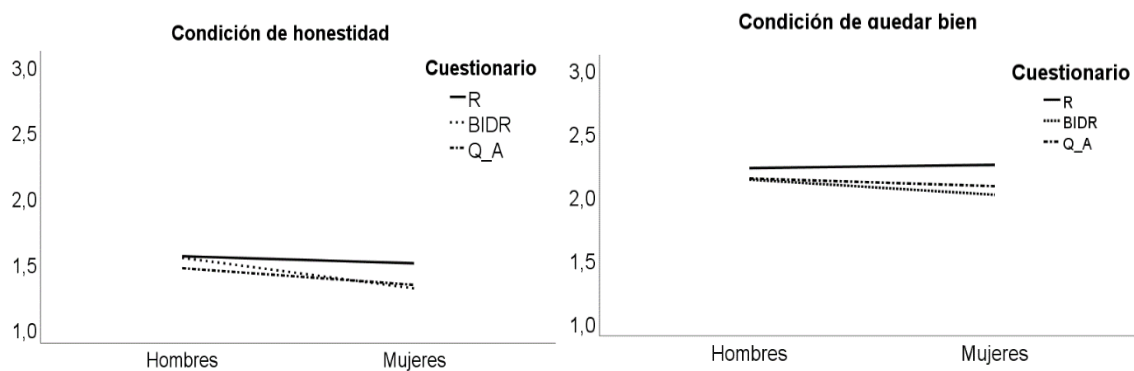
\* $p < .001$

Con el fin de explorar un posible efecto de interacción entre las variables sexo, escala y condición, se realizaron dos MANOVA incluyendo, en cada uno, una de las formas del cuestionario Q. Estos análisis incluyeron dos factores intra, cuestionario, con tres niveles (Q, R, BIDR), y condición, con dos niveles (honestidad y quedar bien), y un factor inter, sexo.

En el primer MANOVA (con R, BIDR y Q\_A) se encontró una interacción significativa entre cuestionario y sexo ( $F(2, 124) = 4.18, p < .05, \eta^2 = 0.063$ ). Concretamente, afectando al contraste entre R y BIDR ( $F(1, 62) = 9.317; p < .01; \eta^2 = 0.131$ ) y entre R y Q\_A ( $F(1, 62) = 7.351; p < .01; \eta^2 = 0.106$ ). Como se puede observar en la Figura 4, las puntuaciones en BIDR y Q\_A se ven más afectadas por el género que las de R.

#### Figura 4

*Puntuaciones medias de R, BIDR y Q\_A en ambas condiciones para hombres y mujeres*



En el segundo MANOVA (con R, BIDR y Q\_B) ninguna de las interacciones posibles resultó significativa.

#### Conclusiones

En su mayoría se cumplieron total o parcialmente las hipótesis planteadas.

En primer lugar, los ítems saturaron en un único factor. Después de descartar los ítems cuya varianza quedaba explicada menos de un 10% por el componente principal, se obtuvieron 17 ítems válidos.

En segundo lugar, la varianza explicada por el componente principal en la escala



R en comparación con la escala BIDR (Paulhus, 1991), no solo la iguala sino que la supera (42,706% y 37,953% respectivamente).

En tercer lugar, los ítems conservados y descartados para la escala Q en el análisis de componentes principales anterior de Rodríguez (2022) y en el actual coinciden mayoritariamente, en concreto, en el 80,43% de los ítems.

En cuarto lugar, la escala R muestra una validez convergente con la escala ya estandarizada del BIDR (2022) y la escala Q, a través de los contrastes de correlaciones.

En quinto lugar, las instrucciones de las condiciones crearon efectivamente el contexto diferencial de ser honesto vs. quedar bien, obteniéndose mayores puntuaciones en la situación de falseamiento que en la de honestidad.

Por último, no se cumplió la hipótesis que esperaba diferencias significativas en las respuestas en condición de honestidad según el sexo, en concreto, con puntuaciones de las mujeres mayores que las de los hombres. Sin embargo, se halló un efecto de interacción del sexo y la escala, en la que las escalas Q\_A y BIDR se muestran más sensibles al sexo que la escala R. Esto se dió para la muestra del primer cuestionario online, pero no en el segundo.

## **Discusión**

Como se ha descrito anteriormente, el sesgo de deseabilidad social puede mermar la validez de algunas pruebas, incluso cuando las personas no tienen el deseo consciente de mentir. Sin embargo, en un contexto de selección de personal, la motivación consciente hacia contestar de la manera más deseable posible es muy alta

(Tracey, 2016), dado que contestar lo que se desea que se conteste es lo que va a conseguir el objetivo final de ser contratado. Para ello, esta escala añade numerosos ítems que permiten detectar cuando las personas pasan de ser honestas, aunque sujetas al sesgo de deseabilidad social, a mentir para quedar bien conscientemente, lo que se denomina como manejo de impresiones (Paulhus, 2020).

Sin embargo, existen numerosos profesionales que saltándose la ética intentan enseñar a sus clientes a burlar la validez de las pruebas de manejo de impresiones, siendo capaces de, por ejemplo, como sería frecuente en procesos de selección, contestar de manera manipulada una prueba de personalidad, pero pasando sin problemas la prueba de manejo de impresiones. Para prevenir el éxito de este entrenamiento, este estudio añade ítems válidos a una batería de preguntas que permitirá extraer cada vez un conjunto de ítems aleatorio, de manera que formen una escala de manejo de impresiones. Además, ninguno de los ítems serán publicados.

Por lo tanto, aunque es necesario continuar validando ítems y contrastando los ítems anteriores, este estudio supone un avance en el desarrollo de una batería de ítems que mide el manejo de impresiones, y permite elaborar pruebas con extracción aleatoria de conjuntos de ítems no publicados y diferentes cada vez. De esta manera, se espera dificultar significativamente su entrenamiento, y por lo tanto, siendo de gran utilidad para aplicar en procesos de selección sin perder su validez.

Se deben mencionar algunas limitaciones que posee este estudio. Respecto a la muestra, se encuentra limitada en número y en representatividad de la población general. Especialmente los análisis realizados en relación a la escala Q cuentan con

aproximadamente la mitad de la muestra total, al haberse dividido sus ítems a la mitad para cada cuestionario. Este tamaño reducido de la muestra puede estar afectando especialmente a los análisis sobre el efecto del sexo pues, además, en los estudios de psicología la ratio de alumnado es bastante desigualado, con muchas más mujeres que hombres.

Por otro lado, dado que la muestra se compone en su mayoría de estudiantes de psicología, y de resto, de estudiantes de máster en recursos humanos, características particulares de este grupo como pueden ser por ejemplo la edad promedio, la personalidad y los intereses, pueden estar afectando a los puntuaciones obtenidas en las escalas y a los análisis realizados de manera diferencial a la población general.

Otra limitación que se debe resaltar es el contexto o método de recogida de datos. En este contexto los participantes siguieron las instrucciones motivados, posiblemente, por la petición del profesor, contribuir al conocimiento científico, colaborar con un compañero y/o beneficiarse de una pequeña subida de nota en una asignatura. Además, bajo el anonimato o estricta privacidad. Sin embargo, en el contexto real, además de conocerse su identidad, tienen el objetivo global de conseguir el puesto de trabajo, a lo que se podría interponer el resultado de esta prueba. Por lo tanto, es necesario poner a prueba esta escala desde en contextos más naturales (por ejemplo, ajustando la variable del anonimato) además de no experimentales, permitiendo el avance en el entendimiento, ajuste y desarrollo de ítems que midan el manejo de impresiones en los procesos de selección de personal.

## Bibliografía

- Block, J. (1965). *The challenge of response sets. unconfounding meaning, acquiescence, and social desirability in the MMPI*. Appleton-Century-Crofts.
- Costa, P. T. Jr., y McCrae, R. R. (1999). *NEO PI-R Inventario de Personalidad NEO Revisado: Manual*. TEA Ediciones.
- Damarin, E, y Messick, S. (1965). Response styles as personality variables: A theoretical integration of multivariate research. *ETS Research Bulletin Series*, 1965: i-116. <https://doi.org/10.1002/j.2333-8504.1965.tb00967.x>
- Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, 116, 429–456.
- Forsyth, D. R., y McMillan, J. (2009). Envy and Gratitude as Mediators of the Relations between Social Undermining and Self-esteem, Depression, and Stress. *Journal of Social and Clinical Psychology*, 28(8), 1035-1070.
- Frenkel-Brunswik, E. (1939). Mechanisms of self-deception. *Journal of Social Psychology*, JO, 409-420.
- Hausknecht, J. P., Day, D. V., y Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57(3), 639-683.
- Jones, S., Miller, J. D., y Lynam, D. R. (2011). Personality, antisocial behavior, and aggression: A meta-analytic review. *Journal of Criminal Justice*, 39(4), 329–337. <https://doi.org/10.1016/j.jcrimjus.2011.03.004>
- Kazmi, S. S., y Amjad, N. (2014). The Relationship of Impression Management with Job Satisfaction and Organizational Commitment. *Pakistan Journal of Commerce and Social Sciences*, 8(1), 75-86.
- Landers, R. N., Sackett, P. R. y Tuzinski, K. A. (2011). Retesting after initial failure, coaching rumors, and warnings against faking in online personality measures for selection. *Journal of Applied Psychology*, 96(1), 202–210. doi:10.1037/a0020375
- Levashina, J., y Campion, M. A. (2007). Measuring faking in the employment interview: Development and validation of an Interview faking behavior scale. *Journal of Applied Psychology*, 92(6), 1638-1656.
- Meehl, P. E., y Hathaway, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology*, 30, 525-564.

- Millham, J., y Kellogg, R. W. (1980). Need for social approval: Impression management or self-deception? *Journal of Research in Personality*, *14*, 445-457.
- Ones, D. S., Viswesvaran, C. y Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, *81*(6), 660-679. doi:10.1037/0021-9010.81.6.660
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*(3), 598-609.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, y L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). Academic Press. <https://doi.org/10.1016/B978-0-12-590241-0.50006-X>
- Paulhus, D. L., y Reid, D. B. (1991). Enhancing accuracy in self-report scales by including social desirability scales: The impact of denial and acquiescence response styles. *Journal of Personality*, *60*(4), 699-721.
- Paulhus, D.L. (2020). Socially Desirable Responding on Self-Reports. En: Zeigler-Hill, V., Shackelford, T.K. (eds). *Encyclopedia of Personality and Individual Differences*. Springer, Cham. [https://doi.org/10.1007/978-3-319-24612-3\\_1349](https://doi.org/10.1007/978-3-319-24612-3_1349)
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.Y., Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879-903. doi:10.1037/0021-9010.88.5.879
- Ramanaiah, N., y Martin, H. (1980). On the Two-Dimensional Nature of the Marlowe-Crowne Social Desirability Scale. *Journal of Personality Assessment*, *44*(5), 507-514.
- Rodríguez Rodríguez, M. (2022). *Creación y validación de una escala de manejo de impresiones* (Trabajo de Fin de Máster, Universidad de La Laguna). <http://riull.ull.es/xmlui/handle/915/31701>
- Rogers, R., y Bender, S. D. (2003). Evaluation of malingering and deception. En A. M. Goldstein (Ed.), *Handbook of psychology: Forensic psychology*, *11*, 109-129. John Wiley y Sons Inc.
- Sackeim, H. A., y Gur, R. C. (1978). Self-deception, self-confrontation, and consciousness. En G. E. Schwartz y D. Shapiro (Eds.). *Consciousness and self-regulation: Advances in research*, *2*, 139-197). New York: Plenum Press.

- Schmitt, D., Realo, A., Voracek, M., y Allik, J. (2008). Why Can't a Man Be More Like a Woman? Sex Differences in Big Five Personality Traits Across 55 Cultures. *Journal of Personality and Social Psychology*, 94(1), 168-182.
- Sigel, M. H. (1948). Nonattitude research methods in social psychology. *Psychological Bulletin*, 45(2), 175-199.
- Tabachnick, B. G., y Fidell, L. S. (2019). *Using multivariate statistics (7th ed.)*. Pearson.
- Tracey, Terence J. G. (2016). A note on socially desirable responding. *Journal of Counseling Psychology*, 63(2), 224-232. doi:10.1037/cou0000135
- Weisberg, Y., DeYoung, C., y Hirsch, J. (2011). Gender differences in personality across the ten aspects of the Big Five. *Frontiers in Psychology*, 2(178), 1-11.
- Wiggins, J. S. (1964). Convergences among stylistic response measures from objective personality tests. *Educational and Psychological Measurement*, 24(3), 551-562. <https://doi.org/10.1177/001316446402400310>
- Zerpa, A. E. (2020). Efectos del manejo de la imagen personal sobre las decisiones de selección. *Directorio de la Universidad de La Laguna*.