
CREACIÓN Y VALIDACIÓN DE UNA ESCALA DE MANEJO DE IMPRESIONES

Alumno: MATÍAS RODRÍGUEZ RODRÍGUEZ

Tutor: LUIS FERNANDO DÍAZ VILELA

Trabajo de Fin de Máster.

Máster Universitario en Desarrollo y Gestión de Recursos Humanos

Escuela de Doctorado y Estudios de Postgrado

Universidad de la laguna, curso académico 2021-22

Contenido

Resumen.....	3
Introducción	4
Objetivos e hipótesis.....	8
Método.....	9
Participantes.....	9
Instrumentos.....	10
Procedimiento.....	11
Resultados.....	12
Discusión y conclusiones.....	15
Referencias.....	18
Anexo 1: Instrucciones de la condición de quedar bien	20
Anexo 2: Instrucciones de la condición de honestidad:.....	21

Resumen

El manejo de impresiones es la tendencia a presentar una imagen socialmente deseable de uno mismo distorsionando conscientemente la realidad. El objetivo de este estudio es generar una base de ítems que permitan captar la tendencia a gestionar impresiones. A partir de esta base se pueden generar diferentes cuestionarios, haciendo más complicado el entrenamiento en pautas de respuesta y más confiable el proceso de evaluación. Los participantes, 75 estudiantes de la Universidad de la Laguna, contestaron un conjunto de ítems creados ad hoc, y la escala de gestión de impresiones del cuestionario BIDR de Paulhus. Recibieron dos instrucciones: generar la imagen más positiva posible y ser completamente honestos. Los resultados muestran que 36 ítems son sensibles al tipo de instrucción recibida, pudiendo detectar si los sujetos evaluados intentan generar una imagen positiva o son honestos. Estos ítems se comportaron de manera convergente con el cuestionario BIDR.

Palabras clave: Deseabilidad social, Manejo de impresiones, Procesos de selección

Abstract: Impression management is the tendency to present a socially acceptable image of oneself consciously distorting reality. The objective of this study is to generate an inventory of items that allow capturing the tendency to manage impressions. From this inventory, different questionnaires can be created, making more complicated the training in response guidelines, thus the evaluation process can be more reliable. The participants, 75 students from the Universidad de la Laguna, answered a set of items created ad hoc, and the impression management scale of the Paulhus' BIDR questionnaire. They received two instructions: to generate the most positive image possible and to be completely honest. The results show that 36 items are sensitive to the type of instruction received, being able to detect if the evaluated subjects try to generate a positive image or are honest. These items behaved convergently with the BIDR questionnaire.

Keywords: Social desirability, Impression management, Selection processes

Introducción

La deseabilidad social se define como el deseo o necesidad de la persona evaluada de generar una imagen positiva frente al evaluador, modificando o modulando sus características personales para ser más aceptable socialmente (Rogers y Bender, 2003).

Dentro de los factores que componen la deseabilidad social, Millham (1974) ha propuesto los de “Atribución vs. Negación”. El polo de atribución consiste en apropiarse cualidades deseables para la persona, mientras que en el de negación se intenta ocultar aspectos negativos o no deseados. Sin embargo, los datos acerca de este constructo mostraron que eran formas distintas de medir un mismo factor, por lo que dejó de emplearse esta distinción de la deseabilidad social.

El acercamiento que ha obtenido mayor apoyo empírico y que parece captar mejor los aspectos del constructo es el de Paulhus (1986). Este autor diferencia dos factores: el manejo de impresiones y el autoengaño.

El autoengaño se puede definir como la tendencia a creer que uno tiene características, generalmente deseables socialmente, que realmente no tiene. La persona se cree realmente sus autoinformes positivos, por lo que es menor influenciado de la situación de evaluación, enmarcándose como una característica más estable y menos dependiente del contexto (Paulhus, 1984).

El manejo de impresiones se refiere a la descripción, voluntariamente distorsionada de uno mismo, con el fin de aparecer mejor de lo que uno es. Esto hace que las puntuaciones obtenidas en procesos de evaluación dependan más del contexto. Es decir, en aquellos casos en los que las respuestas van a ser determinantes para el evaluado esta tendencia a generar una imagen positiva puede aumentar, mientras que en situaciones de mayor anonimato el efecto que produce este factor es menor (Paulhus, 1984).

Entre ambos factores existe un claro componente diferenciador basado en el conocimiento o desconocimiento acerca de la modificación de las propias características (Paulhus, 1984). De la misma forma, ambos factores influyen en las respuestas en situaciones de evaluación, ya sea en formato de test, en

formato de entrevista o en cualquier otra forma en la que el evaluado es consciente de que se están analizando sus respuestas y/o comportamientos. Esta influencia es especialmente crítica en el caso del manejo de impresiones, disminuyendo la validez de las evaluaciones.

La deseabilidad social es un elemento potencialmente perjudicial para la validez de las respuestas de los evaluados, sobre todo cuando es una situación en la que se van a usar las respuestas en beneficio o perjuicio de los participantes (Tracey, 2016). Dentro de estas situaciones un ejemplo claro son las evaluaciones de personalidad en los procesos de selección laboral. En estos casos, el evaluado debe responder la alternativa que considera que lo representa mejor. El evaluado es conocedor de que sus respuestas pueden provocar su admisión o rechazo. Esto hace probable que intente generar una imagen positiva para aumentar las posibilidades de contratación. Esto es, manipular sus respuestas para aparecer más ajustado al perfil buscado.

Los efectos de la deseabilidad social sobre la validez de las evaluaciones han llevado a varios autores a incluir escalas de sinceridad en los cuestionarios de Personalidad. Sin embargo, estas medidas están sujetas, a su vez, a muchos riesgos para la validez. Uno de estos es que normalmente las preguntas de estas escalas se refieren a conductas poco frecuentes o extrañas, por lo que son fácilmente identificables y los participantes pueden modificar sus respuestas con el fin de quedar bien. Otro de los problemas del uso de estas escalas como parte de pruebas de personalidad, es que existen profesionales que entrenan a los candidatos en pautas para responder. Estos profesionales tienen acceso a los cuestionarios y enseñan a los candidatos la manera de obtener puntuaciones que les brinde una ventaja en el proceso de selección (Connelly y Chang, 2015; Ellingson et al., 2001).

Una de las primeras pruebas de personalidad que incluyen la medición de la deseabilidad social es el Inventario Multifásico de Personalidad de Minnesota (MMPI), originalmente desarrollado por Hathaway y McKinley en 1942 y cuya última versión en español es de 2019 (Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom y Kaemmer, 2019). Este cuestionario contiene doce escalas de control de validez. Entre ellas, las escalas L, K y S son especialmente indicadas para su uso en selección de personal (Butcher et al., 2019), La escala L o Lie,

contiene 15 preguntas tendentes a medir si la persona evaluada presenta una tendencia a crear impresiones favorables de sí misma. La escala K o Defensividad, con 30 preguntas, detecta la tendencia a presentarse como personas excesivamente educadas y controladas cuando se obtienen puntuaciones muy altas. Por último, la escala S o escala de presentación superlativa, con 50 preguntas, mide una alta fuerza del ego. Sin embargo, puntuaciones T superiores a 65 puntos indican una tendencia al engaño en la prueba cuando la historia del participante no justifica este ajuste superior (Gordon, 2022).

Otra prueba que incluye estos aspectos es el 16PF (Cattell y Mead, 2008). Esta consta de tres escalas de control. La primera, de Manipulación de la Imagen (MI), muestra la intención de la persona a mostrar una imagen modificada. La escala de Infrecuencia (IN), que se obtiene en base al número de respuestas poco frecuentes que da el sujeto. Y por último, la escala de Aquiescencia (AQ), que evalúa la tendencia de la persona a expresar acuerdo a las afirmaciones que se presentan.

Otros instrumentos evalúan la deseabilidad social de manera concreta y sin formar parte de una prueba de personalidad más amplia. Algunos ejemplos de interés pueden ser: la “Escala de deseabilidad social (DS) de Edwards (1957), o la “Escala de deseabilidad social” de Marlowe-Crowne (MCSDS). Esta última está compuesta por 33 ítems que se dividen en comportamientos deseables o comportamientos indeseables. Los primeros hacen referencia a comportamientos socialmente aprobados pero poco comunes, mientras que los segundos se refieren a comportamientos desaprobados socialmente pero comunes. Por lo tanto, los evaluados con una puntuación alta en comportamientos deseables y baja en indeseables mostrarán una alta deseabilidad social. Esta escala consta de un solo factor llamado “necesidad de aprobación” (Crowne y Marlowe, 1964). Ha sido una de las más citadas pero tiene varios aspectos negativos, como son la redacción, que se ha quedado anticuada dada la fecha de creación y su baja fiabilidad (Beretvas et al., 2002; Leite y Beretvas, 2005).

Una de las escalas más empleadas para la evaluación de la deseabilidad social es el Inventario equilibrado de respuesta deseable (BIDR). Desarrollada

por Paulhus (1991) a partir del modelo mencionado previamente, distingue dos factores: autoengaño y manejo de impresiones. Contiene 40 elementos, 20 que evalúan autoengaño y 20 para manejo de impresiones. Este cuestionario ha mostrado una buena fiabilidad test-retest además de una elevada consistencia interna (Lanyon y Carle, 2007; Li y Li, 2008).

Este instrumento tampoco está libre de críticas, ya que algunos autores indican que los 40 ítems del inventario hacen que los evaluados puedan experimentar frustración, aburrimiento y cansancio (Schmidt et al., 2003). A esto se le suma el hecho de que normalmente se administran más pruebas, ya que el BIDR se suele emplear para controlar la fiabilidad de otras puntuaciones. Para solucionar esto, Hart y colaboradores (2015) crearon y validaron una versión reducida de 16 ítems del BIDR. Esta parece conservar una fiabilidad y validez adecuada, haciéndola una buena opción cuando sea inviable usar la escala original. Por otro lado, de la Rubia y colaboradores (2012) elaboraron una traducción y validación del BIDR, empleando tan solo 10 ítems para cada factor, mostrando buenos resultados en fiabilidad y consistencia interna.

Se ha observado en diversos estudios que la medida de la personalidad es especialmente sensible a los efectos de la deseabilidad social, sobre todo en situaciones en las que los evaluados optan a un puesto de trabajo. Un ejemplo de esto es el estudio de Birkeland y colaboradores (2006), en el que llegaron a varias conclusiones. La primera de ellas es que, si se pide a los evaluados que intenten generar una imagen positiva, aunque para ello tengan que mentir, inflan sus puntuaciones en todos los factores de personalidad, sin hacer ningún tipo de diferenciación. Además, en situaciones reales de selección de personal los candidatos eran capaces de discriminar entre unos factores de personalidad y otros, puntuando de manera desigual. En tercer lugar, inflaban sus puntuaciones sobre todo en conciencia y estabilidad emocional. Esto muestra que estos dos factores parecen ser percibidos como los más deseables en un contexto de trabajo por cuenta ajena. Por lo que los evaluados, incluso sin formación en responder a pruebas de personalidad, sabían identificar en mayor o menor medida los factores constituyentes de la prueba. Por último, también sabían identificar aquellos factores especialmente deseables para el puesto, como por ejemplo el factor de extraversión para los puestos de comercial o de ventas.

No es de extrañar que exista la posibilidad de que algunos candidatos sean entrenados en estrategias para responder a estos cuestionarios. Los entrenadores son personas con conocimientos acerca de los test de personalidad. En un estudio realizado por Landers y colaboradores (2011) se explican algunos rumores acerca de las pautas que daban estos entrenadores, como puede ser responder la puntuación extrema (por ejemplo 1 o 5) al identificar el polo “deseable”. Esto es conocido como sesgo de respuesta extrema. En las pruebas de detección de manejo de impresiones, las pautas serían las contrarias, ya que respuestas extremas indicarían una tendencia a modificar la imagen. Las preguntas de estas pruebas suelen tener afirmaciones de “nunca” o “siempre”, por lo que es relativamente fácil identificarlas y, por lo tanto, son fácilmente entrenables. Estas normas son compatibles con tipos de 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. Sin embargo, no son infalibles y, dependiendo del tipo de test y del tipo de proceso selectivo, los candidatos que sigan estas premisas pueden ver su puntuación perjudicada (Miller y Barret, 2008).

Objetivos e hipótesis

El objetivo de este trabajo es seleccionar y crear una base de ítems que permitan medir el manejo de impresiones. Además, como valor distintivo con respecto a las demás escalas existentes, el hecho de que exista un gran número de ítems disponibles y no una escala predeterminada dificulta el entrenamiento en los mismos. Esto se debe a que, en cada administración, tanto las preguntas como el orden en que aparecen, cambian, haciendo de cada pase de pruebas una selección diferente y única. Por lo tanto, más que una validación de una escala completa se busca seleccionar aquellos ítems válidos, es decir, que midan el constructo de “Manejo de Impresiones” y tengan adecuadas características psicométricas. Las hipótesis son las siguientes:

Hipótesis 1: Con respecto a la fiabilidad, se espera que todos los ítems creados saturen en un único factor.

Hipótesis 2: Se espera que los cuestionarios generados tengan una alta validez convergente con el cuestionario BIDR (Paulhus, 1991).

Hipótesis 2a: Las correlaciones entre cuestionarios en la misma condición (responder para quedar bien o responder honestamente) serán altas y positivas.

Hipótesis 2b: Las correlaciones entre condiciones para el mismo cuestionario tenderán a cero o serán ligeramente negativas.

Hipótesis 3: Con respecto al cuestionario creado las instrucciones de honestidad/falseamiento, producirán respuestas infladas en la condición de falseamiento independientemente del orden en que se produzca la condición (Honestidad-Falseamiento / Falseamiento-Honestidad).

Hipótesis 4: Con respecto al BIDR este cumplirá con la hipótesis anterior. Es decir, las instrucciones honestidad/falseamiento generará puntuaciones diferentes siendo superiores las de falseamiento, mientras que el orden de presentación de las instrucciones no afectará a las puntuaciones.

Hipótesis 5: Ambos cuestionarios tendrán la misma tendencia, de manera que las puntuaciones se verán afectadas por igual, cuando se pida a los participantes para que respondan honestamente y falseando sus respuestas para quedar bien.

Método

Participantes

La selección de la muestra ha sido mediante muestreo de conveniencia. Está formada por 75 participantes, todos estudiantes del Grado de Psicología, del Máster Universitario en Desarrollo y Gestión de Recursos Humanos y del Master Universitario en Dirección de Recursos Humanos, titulaciones pertenecientes a la Universidad de La Laguna. Dadas las características de las pruebas, en los que se les pide que den una imagen positiva, aunque tengan que mentir y también que sean honestos, se optó por hacer las pruebas totalmente anónimas, sin preguntas relativas a la edad, sexo u otros datos sociodemográficos, para garantizar la validez de las respuestas. Es por esto por lo que solo se dispone del número de sujetos y de la fuente de la que se extrajo la muestra.

Instrumentos

Se aplicaron dos pruebas en dos condiciones cada una. Las pruebas consistieron en el Inventario equilibrado de respuesta deseable (BIDR). Desarrollada por Paulhus (1991) y un cuestionario de prueba que llamaremos “Q”.

El BIDR se compone de dos factores, autoengaño y manejo de impresiones. Su versión original contiene 40 elementos, 20 para evaluar cada factor. Este cuestionario ha mostrado una fiabilidad 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 estudio se empleó tan solo el factor de manejo de impresiones de una versión en español de la prueba (Zerpa 2020). Esta versión consta de 21 ítems en lugar de 20 ya que esta escala terminaba con una pregunta que probablemente evocaría emociones negativas (“Tengo algunas costumbres que me parecen horribles”), y para no terminar el pase de las pruebas con esto se añadió otra de contenido más neutro. La escala de respuesta es de tres alternativas, una afirmativa/verdadera, una neutra y una negativa/falsa, estas alternativas se adaptaban a cada uno de los ítems. Las respuestas positivas se puntúan con un uno, las neutras con un 2 y las negativas con un 3.

El cuestionario “Q” fue desarrollado para este trabajo seleccionando ítems a partir de la base de datos “International Personality Item Pool” (IPIP), desarrollada por Goldberg (1999). Esta base de datos contiene 3320 ítems, muchos de los cuales evalúan constructos ya establecidos. Se seleccionaron 209 preguntas 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”.

A partir de la selección y adaptación de estas preguntas, se seleccionaron al azar 46 de ellas para conformar un cuestionario. 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 estas respuestas varían para ajustarse a la afirmación. Esta se puntúa con un punto a la respuesta “a”, dos puntos a la “b” y tres puntos a la “c” en los elementos directos y al contrario en los elementos inversos.

Una vez generado el cuestionario, se elaboraron dos cuadernillos de respuesta. Ambos cuadernillos contenían las mismas pruebas (el nuevo cuestionario Q y la escala de manejo de impresiones del cuestionario BIDR), repetidas. La mitad de los cuadernillos empezaban con instrucciones de quedar bien, seguidas de las dos pruebas. A continuación, se presentaban instrucciones de honestidad y las dos pruebas otra vez. La mitad de la muestra recibió primero las instrucciones de “quedar bien”, mientras que la otra mitad recibió primero las instrucciones de “honestidad”. La composición esquemática de los cuadernillos se expone en la *Tabla 1*. Además, los cuadernillos contenían una medida de Personalidad para cada condición, que no será utilizada en este estudio.

Tabla 1: Composición y orden de los cuadernillos de respuesta.

Cuadernillo 1	Cuadernillo 2
Instrucciones “quedar bien”	Instrucciones “honestidad”
Cuestionario Q	Cuestionario Q
BIDR	BIDR
Instrucciones “honestidad”	Instrucciones “quedar bien”
Cuestionario Q	Cuestionario Q
BIDR	BIDR

Procedimiento

En este estudio se siguió una metodología de encuestas, con un diseño transversal.

Tras crear los cuadernillos, se solicitó a los/as profesores/as de los cursos anteriormente citados el permiso para pasar los cuestionarios, consensuando la fecha y hora para su administración. Esta se realizó en las aulas de cada curso durante aproximadamente 30 minutos a los/as alumnos/as que voluntariamente quisieran participar.

Resultados

Con el fin de comprobar la Hipótesis 1, se realizaron varios análisis de componentes principales. Para estos análisis se usaron los datos correspondientes a la condición de “quedar bien” debido a que el interés de este estudio es el de generar ítems que permitan detectar esta tendencia. Además, se limitaron los resultados a un único componente.

Tras someter a los 46 ítems de ambos cuestionarios al análisis componentes principales tan solo se mantuvieron aquellos que saturaron por encima de 0,33. Es decir, aquellos cuya varianza fuese explicada por el componente en al menos un 10%. Se eliminaron trece ítems por no cumplir con este criterio.

Posteriormente se repitió el análisis de componentes principales con los 33 ítems seleccionados. Como resultado de este análisis, un ítem mostró saturaciones por debajo de 0,33, por lo que se excluyó también. Un tercer análisis sobre la matriz de correlaciones de los 32 ítems restantes, resultó en un componente principal que explicaba más del 10% de la varianza de cada uno de ellos. Estos 32 ítems fueron seleccionados para mantenerlos en la base de datos y para los análisis posteriores. De esta manera, se pasó de 46 ítems a un conjunto de 32. Por lo tanto, la hipótesis uno, que se refería a la saturación de todos los ítems en un único factor se cumple parcialmente.

Una vez seleccionados estos 32 ítems se invirtieron aquellos con saturación negativa. Es decir, cuanto mayor era la puntuación mayor era la tendencia a ser honesto, mientras que cuánto menor era la puntuación, existía una tendencia a generar una imagen positiva. El resto de ítems tenían una tendencia contraria, por lo que estos debían ser invertidos para que todos fueran en la misma dirección y las puntuaciones pudieran sumarse sin que se contrarrestaran.

Este proceso se repitió con el BIDR. Tras el primer análisis restringido a un componente principal, se eliminaron tres ítems que saturaban por debajo de 0,33. En un segundo análisis, el componente principal explicó más del 10% de la varianza de cada uno de los 18 ítems restantes. De igual manera también se

invertieron aquellos ítems con saturaciones negativas y se calcularon las medias para cada participante.

Así, se generaron cuatro puntuaciones correspondientes a las cuatro condiciones: cuestionario (Q) en situación de Honestidad (H) y en situación de “quedar bien” (F) y BIDR en situación de honestidad (H) y en situación de “quedar bien” (F).

En la Tabla 2 se puede observar que las correlaciones entre cuestionarios de la misma condición son altas y positivas. Además, las correlaciones entre condiciones para el mismo cuestionario son negativas, aunque mucho menores.

Tabla 2: Correlación entre el cuestionario creado (Q) y el BIDR en condición de honestidad (H) y en la de “quedar bien” (F).

	BIDRF	BIDRH	QF
BIDRH	-0,247 * (74)		
QF	0,782 *** (74)	-0,175 (74)	
QH	-0,217 (72)	0,806 *** (73)	-0,117 (72)

Entre paréntesis se muestran los grados de libertad.

Como puede verse, estos resultados apoyan claramente las hipótesis 2a y 2b.

Para comprobar la tercera hipótesis, se aplicó un ANOVA sobre las puntuaciones en el cuestionario Q en función de la condición y el orden (Tabla 3).. El efecto de la condición fue significativo y considerable ($F(1, 70) = 230,412$; $p < ,001$; $\eta^2 = ,767$). Los participantes obtienen puntuaciones mayores en la condición de honestidad que en la de “quedar bien”. El orden en el que aparecieron las condiciones no influyó en las respuestas ($F(1, 70) = 0,031$; $p = ,86$; $\eta^2 < ,001$), ni se produjo interacción alguna entre condición y orden ($F(1, 70) = 0,362$; $p = ,549$; $\eta^2 = ,005$) por lo que se confirma la Hipótesis 3

Tabla 3: Descriptivos del cuestionario Q.

Condición	Orden	Media	Dt
Quedar bien	Primero quedar bien	1,26	0,31
	Primero honestidad	1,23	0,25
Honestidad	Primero quedar bien	2,03	0,30
	Primero honestidad	2,07	0,35

N = 36

Para comprobar la Hipótesis 4, en la que se espera que BIDR sea sensible a la condición, pero no al orden, se aplicó un ANOVA similar al anterior. En este caso, las variables independientes fueron las mismas y la variable dependiente fueron las respuestas al cuestionario BIDR, cuyas distribuciones se describen en la Tabla 4. En primer lugar, se obtuvo un efecto considerable (73,6%) y significativo de la condición ($F(1, 72) = 200,666$; $p < ,001$; $\eta^2 = ,736$). Mientras que el orden de presentación no tuvo efecto sobre las puntuaciones ($F(1, 72) = 0,092$; $p = 0,762$; $\eta^2 = ,001$), como tampoco lo tuvo la interacción entre condición y orden de presentación ($F(1, 72) = 0,316$; $p = 0,575$; $\eta^2 = ,004$). Por lo tanto, la tendencia en los resultados fue similar a los del cuestionario desarrollado, confirmándose la hipótesis cuatro al comportarse el BIDR de una manera muy similar al cuestionario desarrollado para este trabajo.

Tabla 4: Descriptivos del BIDR.

Condición	Orden	Media	Dt	N
Quedar bien	Primero quedar bien	1,22	0,29	36
	Primero honestidad	1,20	0,30	38
Honestidad	Primero quedar bien	2,07	0,36	36
	Primero honestidad	2,12	0,41	38

Para comprobar la Hipótesis 5 se aplicó un ANOVA sobre los datos de la Tabla 5, con dos factores intra, Cuestionario, con dos niveles (Q y BIDR) y Condición, con dos niveles también (quedar bien y honestidad).

Tabla 5: Medias y desviaciones típicas de ambos cuestionarios según condición.

Condición	Cuestionario	Media	Dt
Quedar bien	Q	1,24	0,28
	BIDR	1,21	0,29
Honestidad	Q	2,05	0,32
	BIDR	2,10	0,39

n = 72

El efecto de condición sobre las respuestas a ambos cuestionarios, representado en la *Figura 1*, resultó significativo y considerable (76.3%). Tanto el cuestionario propio como el BIDR son capaces de captar las diferencias

debidas a las instrucciones de honestidad y de “quedar bien” ($F(1, 71) = 228,503$; $p < ,001$; $\eta^2 = ,763$).

Por otro lado, no resultó significativa la diferencia de medias entre los cuestionarios, por lo que ambos obtuvieron medias similares en iguales condiciones ($F(1, 71) = 0,290$; $p = 0,592$; $\eta^2 = ,004$). Sin embargo, la interacción entre condición y cuestionario sí resultó significativa, aunque con un tamaño del efecto muy pequeño (6,4%), por lo que los resultados parecen mostrar que el BIDR es ligeramente más sensible a la hora de captar diferencias en la condición de “quedar bien” y honestidad ($F(1, 71) = 4,825$; $p < ,05$; $\eta^2 = ,064$). Es por esto que la hipótesis cinco se cumple de manera parcial, ya que, efectivamente, ambos cuestionarios son sensibles al tipo de instrucción recibida, pero el BIDR parece algo más sensible.

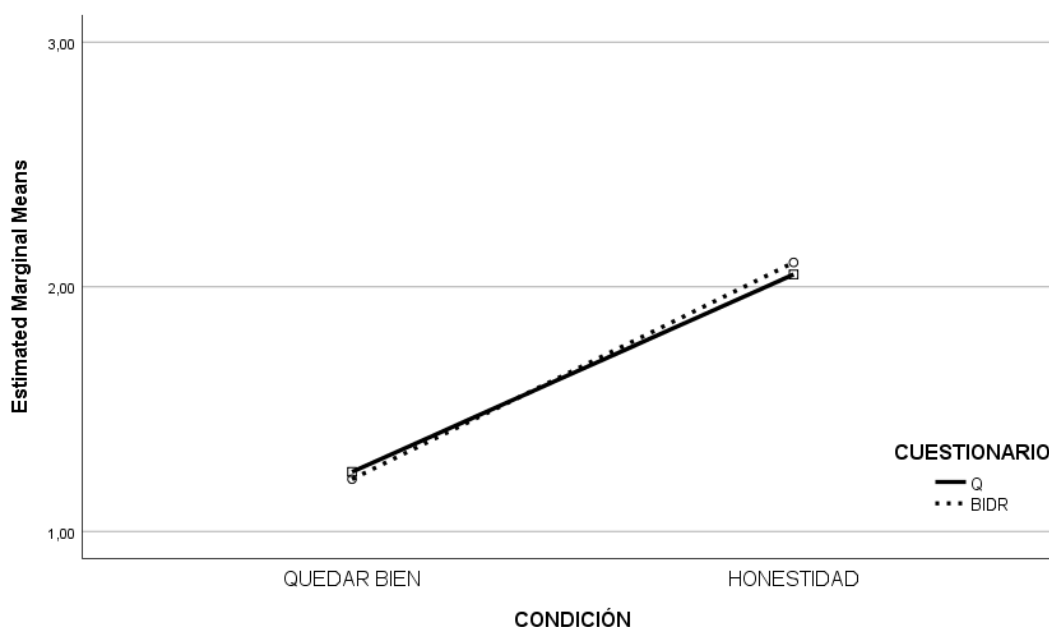


Figura 1: Medias de cada cuestionario en las dos condiciones.

Discusión y conclusiones

En general, las hipótesis planteadas en el estudio se cumplieron total o parcialmente.

En primer lugar, la mayoría de los ítems seleccionados (36 de 46) saturaron en un único factor con más de 0.33, explicando como mínimo un 10% de la varianza de ese factor, que se corresponde con el constructo de manejo de impresiones, que es el que se pretende detectar.

De igual manera, existe validez convergente con el cuestionario BIDR de Paulhus (1991), con correlaciones significativas, por lo que se puede afirmar que ambos están midiendo constructos similares.

Por último, tanto el cuestionario desarrollado como el BIDR son sensibles al tipo de instrucción que reciben los participantes, mostrando diferencias significativas. Además, no son sensibles al orden en el que se presentan las instrucciones, tal y como se esperaba.

Los resultados muestran que el BIDR es ligeramente más sensible que el cuestionario desarrollado, mostrando mayores diferencias en las medias de la condición de “quedar bien” y honestidad, aunque con tamaños de efecto leves. Sin embargo, se esperaba que ambos fueran igualmente sensibles. Es esta la hipótesis que no se cumple totalmente.

Dada la naturaleza del constructo, todas las pruebas de deseabilidad social (y en especial las de manejo de impresiones) corren el riesgo de ser poco fiables, no por las características de la prueba en sí, sino por la intención de los sujetos a mostrar una imagen positiva. Esto se debe a que, generalmente, estas pruebas se usan en procesos de selección de personal en los que se evalúan características de personalidad. En estos casos, la motivación por modificar las respuestas para generar una imagen socialmente deseable es muy alta (Tracey, 2016). Además de esto, existe el riesgo de que los evaluados hayan recibido entrenamiento en pautas para identificar y contestar a estas pruebas. Esto hace especialmente importante poner en marcha todos los mecanismos de los que se disponga para evitar los efectos de este entrenamiento.

La solución que se plantea en este trabajo es la de generar una base de ítems, todos igualmente válidos para medir el manejo de impresiones. Este trabajo sería el primero de una serie que llevaría a crear una base lo suficientemente extensa como para seleccionar de manera aleatoria un número de ítems en cada aplicación de la escala. Por lo tanto, cada vez que se respondiera a la prueba el orden y el contenido de los ítems sería diferente, haciendo que los entrenadores no puedan dar unas pautas tan claras para responder de manera deseable y así ser detectados por estas pruebas.

Como ya se ha comentado anteriormente, el entrenamiento en pautas para responder afecta tremendamente a los resultados y a su validez para hacer predicciones. Esta modificación de los resultados se da tanto en pruebas de personalidad (Martin et al., 2002; Viswesvaran y Ones, 1999) como en pruebas de deseabilidad social (Landers et al., 2011; Miller y Barret, 2008). Por lo tanto, el hecho de que sea menos sensible se verá compensado por la dificultad para identificar la prueba y para entrenarse en cómo responder a la misma, dándole una utilidad a tener en cuenta en los procesos de selección de personal.

Este estudio tiene algunas limitaciones. En primer lugar, se debe realizar este proceso de validación de ítems varias veces, ya que tan solo se cuenta con 36, y para generar una base de datos numerosa hace falta repetir el procedimiento de recogida y análisis de datos. Sin embargo, a pesar de ser un primer acercamiento, los resultados son prometedores para continuar generando y validando ítems.

Otra limitación se refiere a las condiciones de recogida de datos, ya que los resultados obtenidos son producto de una situación controlada cuasi-experimental, en el que los estudiantes siguen las instrucciones que se les plantean. Los participantes no tienen ninguna motivación para generar respuestas en un sentido positivo o socialmente deseable, ya que los datos son anónimos y con una finalidad científica. Por lo tanto, se hace necesaria una validación de este cuestionario en situaciones de selección reales, para analizar los posibles cambios que puedan surgir. En este sentido ya se han podido recoger algunos datos preliminares en estas situaciones, mostrando resultados prometedores.

De igual manera, se hace necesario aumentar la muestra de sujetos para que los datos tengan mayor evidencia y estén menos sujetos a variaciones que se puedan dar debidas al azar.

Referencias:

- Beretvas, S. N., Meyers, J. L., & Leite, W. L. (2002). A Reliability Generalization Study of the Marlowe-Crowne Social Desirability Scale. *Educational & Psychological Measurement*, 62(4), 570–589. <https://doi-org.accedys2.bbt.ull.es/10.1177/0013164402062004003>
- Butcher, J.N., Graham, J.R., Ben-Porath, Y.S., Tellegen, A., Dahlstrom, W.G. y Kaemmer, B. (2019). *MMPI-2 Inventario Multifásico de Personalidad de Minnesota-2*. Madrid: Tea Ediciones.
- Cattell, H. E. P., y Mead, A. D. (2008). The Sixteen Personality Factor Questionnaire (16PF). En G. J. Boyle, G. Matthews, y D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment, Vol. 2. Personality measurement and testing* (pp. 135–159). Sage Publications, Inc. <https://doi.org/10.4135/9781849200479.n7>
- Connelly, B. S., y Chang, L. (2015). A Meta-Analytic Multitrait Multirater Separation of Substance and Style in Social Desirability Scales. *Journal of Personality*, 84(3), 319-334. doi:10.1111/jopy.12161
- Crowne, D. P., y Marlowe, D. (1964). *The approval motive*. New York, NY: Wiley
- de la Rubia, J. M., Cadena, C. H. G., y Casas, C. J. A. (2012). Traducción y validación del Inventario Balanceado de Deseabilidad Social al Responder en una muestra probabilística de estudiantes universitarios mexicanos. *Revista de Psicología GEPU*, 3(2), 54-72.
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. Dryden Press.
- Ellingson, J. E., Smith, D. B., & Sackett, P. R. (2001). Investigating the influence of social desirability on Personality factor structure. *Journal of Applied Psychology*, 86(1), 122-133. doi:10.1037//0021-9010.86.1.122
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe*, Vol. 7 (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.
- Gordon, R. M. (2011). *ROBERT M. GORDON, PH.D. ABPP*. Recuperado el 28 de 04 de 2022, de MMPI-2: <https://www.mmapi-info.com/validity-scales>
- Hart, C. M., Ritchie, T. D., Hepper, E. G., y Gebauer, J. E. (2015). The balanced inventory of desirable responding short form (BIDR-16). *Sage Open*, 5(4), 2158244015621113.
- 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*, 202-210. doi:10.1037/a0020375.sup
- Lanyon, R. I., y Carle, A. C. (2007). Internal and External Validity of Scores on the Balanced Inventory of Desirable Responding and the Paulhus

- Deception Scales. *Educational and Psychological Measurement*, 67(5), 859–876.
- Leite, W. L., y Beretvas, S. N. (2005). Validation of scores on the Marlowe-Crowne social desirability scale and the balanced inventory of desirable responding. *Educational and Psychological Measurement. A Bimonthly Journal Devoted to the Development and Application of Measures of Individual Differences*, 65(1), 140.
- Li, F., y Li, Y. (2008). The Balanced Inventory of Desirable Responding (BIDR): A factor analysis. *Psychological Reports*, 103(3), 727–731. <https://doi-org.accedys2.bbtck.ull.es/10.2466/PR0.103.7.727-731>
- Martin, B. A., Bowen, C. C., y Hunt, S. T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences*, (32), 247–256. [http://dx.doi.org/10.1016/S0191-8869\(01\)00021-6](http://dx.doi.org/10.1016/S0191-8869(01)00021-6)
- Miller, C. E., y Barret, G. V. (2008). The Coachability and Fakability of Personality-Based Selection Tests Used for Police Selection. *Public Personnel Management*, 37(3), 339-351.
- Millham, J. (1974). Two components of need for approval score and their relationship to cheating following success and failure. *Journal of Research in Personality*, 8(4), 378–392. [https://doi-org.accedys2.bbtck.ull.es/10.1016/0092-6566\(74\)90028-2](https://doi-org.accedys2.bbtck.ull.es/10.1016/0092-6566(74)90028-2)
- Paulhus, D. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46(3), 598-609.
- Rogers, R., y Bender, S. D. (2003). Evaluation of malingering and deception. In A. M. Goldstein (Ed.), *Handbook of psychology: Forensic psychology*, Vol. 11. (pp. 109–129). John Wiley & Sons Inc.
- Schmidt, F. L., Le, H., y Ilies, R. (2003). Beyond alpha: an empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs. *Psychological Methods*, 8(2), 206–224. <https://doi-org.accedys2.bbtck.ull.es/10.1037/1082-989x.8.2.206>
- Tracey, T. J. (2016). A note on socially desirable responding. *Journal of Counseling Psychology*, 63(2), 224-232. doi:10.1037/cou0000135
- Viswesvaran, C., y Ones, D. (1999). Meta-Analyses of Fakability Estimates: Implications for Personality Measurement. *Educational and Psychological Measurement*, 59(2), 197- 210. <https://doi.org/10.1177/00131649921969802>
- Zerpa, A. E., (2020). Efectos del manejo de la imagen personal sobre las decisiones de selección. *Directorio de la Universidad de La Laguna*

Anexo 1: Instrucciones de la condición de quedar bien

A continuación, se le presentan una serie de preguntas. Por favor, responda pensando en **AQUELLAS ALTERNATIVAS QUE TE HARÍAN QUEDAR MEJOR CON EL ENTREVISTADOR, AUNQUE PARA ELLO TENGA QUE MENTIR.**

A la hora de puntuar su nivel de acuerdo con las frases que le mostramos recuerde lo siguiente:

- 1 = Si la frase es completamente falsa en su caso, si está en total desacuerdo con ella
- 2 = Si la frase es tan cierta como falsa, si no puede decidirse o si Usted se considera neutral en relación con lo que se dice en ella
- 3 = Si la frase es completamente cierta, si está totalmente de acuerdo con ella.

Al contestar tenga en cuenta lo siguiente:

- No piense demasiado las cuestiones ni emplee mucho tiempo en decidirse. Las frases son muy cortas para darle todos los detalles que Vd. necesita
- Evite señalar la respuesta 3, excepto cuando le sea imposible decidirse por las otras; lo corriente es que esto le ocurra sólo cada tres o más cuestiones.
- Procure no dejar ninguna cuestión sin contestar.

RECUERDE: responda pensando en **AQUELLAS ALTERNATIVAS QUE TE HARÍAN QUEDAR MEJOR CON EL ENTREVISTADOR, AUNQUE PARA ELLO TENGAS QUE MENTIR.**

Anexo 2: Instrucciones de la condición de honestidad:

A continuación, se le presentan una serie de preguntas. Por favor, responda de manera **HONESTA Y SINCERA**. No señale sus respuestas pensando en lo que “es bueno” o “lo que interesa” para impresionar al examinador.

No hay respuestas correctas o incorrectas, pues pretendemos conocer cuál es su opinión, sea cual sea, para así poder tomar decisiones de forma más ajustada a cómo es usted.

- 1 = Si la frase es completamente falsa en su caso, si está en total desacuerdo con ella
- 2 = Si la frase es tan cierta como falsa, si no puede decidirse o si Usted se considera neutral en relación con lo que se dice en ella
- 3 = Si la frase es completamente cierta, si está totalmente de acuerdo con ella.

A la hora de puntuar su nivel de acuerdo con las frases que le mostramos recuerde lo siguiente:

- No piense demasiado las cuestiones ni emplee mucho tiempo en decidirse. Las frases son muy cortas para darle todos los detalles que Vd. necesita
- Evite señalar la respuesta 2, excepto cuando le sea imposible decidirse por las otras; lo corriente es que esto le ocurra sólo cada tres o más cuestiones.
- Procure no dejar ninguna cuestión sin contestar.

RECUERDE: Conteste sinceramente. No señale sus respuestas pensando en lo que “es bueno” o “lo que interesa” para impresionar al examinador.