

Propiedades psicométricas de la escala de deseabilidad social de Marlowe y Crowne en docentes peruanos

Psychometric properties of the Marlowe-Crowne social desirability scale in peruvian teachers

Andrés Burga León^{a,*}, Miguel Ecurra Mayaute^a

^aUniversidad de Lima, Perú

Recibido: 19 enero de 2017

Aceptado: 10 de octubre de 2017

Resumen

El objetivo de la investigación fue obtener evidencias de las propiedades psicométricas de las puntuaciones derivadas de aplicar la escala de deseabilidad social de Marlowe y Crowne a una población de 28 972 docentes de segundo grado de primaria evaluada en el Perú, la cual estuvo conformada por distintos estratos, como el tipo de gestión (74% estatales y 26% no estatales) y la ubicación geográfica (urbana 65% y rural 35%). Los análisis psicométricos se realizaron en dos muestras de 5000 casos cada una, con características similares a la población. Además, se describe y comparan la deseabilidad social de los docentes. Los resultados permitieron obtener evidencias de validez vinculadas a la estructura interna del instrumento aplicando análisis factoriales exploratorios y confirmatorios que apoyaron la unidimensionalidad del instrumento. Para analizar la confiabilidad del instrumento se evaluaron los modelos de medición, y se notó que el congénico era el que mejor caracterizaba a los ítems, por ello se calculó la confiabilidad a través de los coeficientes omega que alcanzaron valores entre .83 y .87, adicionalmente se incluyeron los intervalos de confianza, los errores estándar de medición, los baremos por rangos percentilares. Para realizar las comparaciones de los puntajes de la DS por estrato se analizó la invarianza del instrumento, encontrándose que no se observaron diferencias estadísticas significativas.

Palabras clave: Propiedades psicométricas, Deseabilidad Social, Escala de Marlowe Crowne, docentes de primaria.

Abstract

The objective of the research was to gather evidence of the psychometric properties of the scores derived from the Marlowe–Crowne Social Desirability Scale applied to a population of 28 972 second-grade elementary school teachers in Peru, considering different strata, such as management type (public sector 74% and private sector 26%) and geographic location (urban 65% and rural 35%). Psychometric analyses were developed in two samples of 5000 cases each, with similar characteristics to those of the study population. In addition, social desirability of the teachers was described and compared. The results allowed to gather validity evidence linked to the internal structure of the instrument by applying exploratory and confirmatory factor analyses that supported the unidimensionality of the instrument. In order to analyze the reliability of the instrument, measurement models were assessed, with the congeneric model being the best option to characterize the items. Therefore, reliability was assessed using omega coefficients that reached values between .83 and .87. Additionally, confidence intervals, standard errors of measurement, and scales based on percentile ranks were included in the research. To compare social desirability scores by stratum, the instrument invariance was analyzed and no significant statistical differences were found.

Keywords: Psychometric properties, social desirability, Marlowe–Crowne scale, elementary school teachers.

Para citar este artículo:

Burga, A., & Ecurra, M. (2017). Propiedades psicométricas de la escala de deseabilidad social de Marlowe y Crowne en docentes peruanos. *Liberabit*, 23(2), 191-212. doi: 10.24265/liberabit.2017.v23n2.03

Este es un artículo Open Access bajo la licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0



Introducción

Los resultados de las evaluaciones censales de estudiantes (ECE), realizadas por la Unidad de Medición de la Calidad de los Aprendizajes (UMC) del Ministerio de Educación del Perú, además de obtener información sobre el logro de los estudiantes, permiten obtener información sobre factores asociados a dichos logros. En el contexto de los modelos multinivel de factores asociados, también se consideran los factores propios del estudiante y del docente. De manera que se aplican cuestionarios a los docentes para recoger información sobre sus estrategias pedagógicas, autoeficacia como docente, valoración del trabajo docente y diversas variables demográficas.

El análisis estadístico de las distribuciones de las variables recogidas de los docentes muestra un sesgo positivo, en la medida que se observa una tendencia a una mayor concentración de puntajes mayores en las pruebas que, a manera de hipótesis, podría deberse a la deseabilidad social (DS) (Domínguez Espinoza, et al., 2012), además, los modelos de factores asociados muestran un tamaño del efecto pequeño al considerar las variables propias del nivel docente. Por ello, es importante contar con instrumentos estandarizados que permitan detectarla, de manera que sea incluida en los análisis estadísticos correspondientes, y puedan realizarse análisis más precisos (Soubelet & Salthouse, 2011).

La DS supone que algunas de las alternativas de respuesta en ciertos ítems orientados a medir variables de personalidad son socialmente más convenientes o deseables que otras. Por lo tanto, algunas personas tienden a elegir las en independencia de su nivel en el rasgo que el ítem pretende medir. De esta manera se distorsiona la medición de un rasgo o atributo latente, lo cual también impactaría el análisis de las relaciones entre dichos rasgos. Por ello es importante contar con instrumentos que proporcionen resultados a partir de los cuales se pueda inferir la cantidad de deseabilidad social que presenta una persona y así tomar decisiones, como

la eliminación de sus respuestas de análisis posteriores.

La DS es la tendencia de los individuos a presentar una buena impresión de acuerdo con las normas culturales actuales (Mick, 1996), lo cual supone que algunas de las alternativas de respuesta en ciertos ítems orientados a medir variables de personalidad son socialmente más convenientes o deseables que otras. Su presencia puede afectar a la validez de una investigación en diferentes ámbitos como en la psicología organizacional y la selección de personal (Furnham, 1990), el comportamiento del consumidor (Mick, 1996) y los estudios psicológicos en el campo de las ciencias de la salud (Neeley & Cronley, 2004) entre otras. King y Bruner (2000) indican que el sesgo que produce la DS juega un papel importante en la supresión u oscurecimiento de las relaciones entre variables, así como en la producción artificial de relaciones entre las variables independientes y dependientes, por lo que Sullman y Taylor (2010) recomiendan incluir una medida de la deseabilidad social siempre que la investigación lo requiera.

Uno de los instrumentos más utilizados para medir la DS es la escala de Marlowe-Crowne (MCSDS; Crowne & Marlowe, 1960), la cual fue construida por sus autores seleccionando ítems específicos de algunos inventarios y cuestionarios de personalidad más usados, de tal forma que un reactivo era incluido si cumplía los siguientes requisitos: (1) presentar un comportamiento que es culturalmente sancionado o aprobado, y (2) tener una mínima implicancia patológica o anormal, de manera que la escala contiene ítems sobre conductas que son objeto de actividades culturales aprobadas y que son de alguna manera improbables de ocurrir pero tienen un mínimo de inadaptación implícita como lo indica Crowne y Marlowe (1960).

Desde la creación del instrumento hasta la actualidad se han desarrollado una gran cantidad de estudios que analizaron sus propiedades psicométricas

como por ejemplo las evidencias de validez. En el estudio de la naturaleza dimensional de la estructura interna de la MCSDS, se halló que la escala presenta dos factores que aluden a la atribución y la negación y que predicen las puntuaciones derivadas del MMPI en las Escalas L, K y Represión-Sensibilización (Ramanaiah, Schill, & Leung, 1977), aunque estudios posteriores indican que la escala es unidimensional (Ramanaiah & Martin, 1980). Por otro lado, Beretvas, Meyer y Leite (2002), realizaron un metaanálisis y encontraron que la confiabilidad de las puntuaciones derivadas de aplicar la MCSDS es mayor en las mujeres (.79) que en los hombres (.68) y un valor insatisfactorio en los adolescentes (.53).

Debido al tamaño de la escala original se han elaborado formas cortas del MCSDS, las cuales fueron desarrolladas a fin de disminuir el tiempo de administración, debido que se pudo demostrar que algunos ítems contribuían poco a la escala general (Strahan & Gerbasi, 1972). Adicionalmente los estudios en los que se compararon diversas formas, cortas y largas, de la escala indican que las formas cortas se pueden utilizar con tanto éxito como la versión original, (Loo & Thorpe, 2000; Sârbescub, Rustu, & Costea, 2012;); aunque por otro lado existen reportes en los cuales se indica que no hay un buen ajuste del modelo de las formas cortas (Barger, 2002). Por otro lado, en el contexto de las investigaciones transculturales se ha observado que la forma C del MCSDS no presentó equivalencia escalar, por lo que no se recomienda su uso en ese tipo de investigaciones (Verardi et al., 2010).

El estudio de la DS ha originado la necesidad de realizar la adaptación lingüística y psicométrica de la escala MCSDS a contextos diferenciados, por lo que se desarrollaron diferentes adaptaciones a diversos lenguajes tanto de la versión original como de las diferentes formas cortas desarrolladas. En Europa Rudmin (1999) tradujo la escala al idioma noruego, desarrollando además una versión corta de 10 ítems; Ferrando y Chico (2000) desarrollaron la versión española, estudiando sus propiedades psicométricas

desde el marco de la teoría de respuesta al ítem (TRI) aplicando el modelo logístico de dos parámetros, demostrando que la versión adaptada fue unidimensional, con propiedades psicométricas similares a las obtenidas en la versión original. Por su parte Scagliusi et al. (2004), tradujeron la escala al portugués, estableciendo la adecuación lingüística con el procedimiento de doble traducción, alcanzando una adecuada confiabilidad con el método test-retest. En tanto que Sârbescub, Rusu y Costea (2012) desarrollaron la versión rumana, aplicando análisis factoriales confirmatorios, demostrando que las versiones cortas lograron mejores ajustes que la versión larga siendo la mejor de todas la que incluye 13 ítems. Aunque la más confiable fue la versión de 21 ítems.

En el África subsahariana, Vu, Tran, Pham y Ahmed (2011) tradujeron la escala para ser usada en, Etiopia, Kenia, Mozambique y Uganda. En Oceanía, Ray (1984) aplicó en Australia dos formas cortas de la escala Marlowe-Crowne propuestas por Greenwald y Satow (1970) y por Strahan y Gerbasi (1972) las cuales presentaron coeficientes alfas satisfactorias en las muestras estudiadas.

En América latina se han realizado dos estudios en Argentina. Cosentino y Castro (2008) adaptaron la escala presentando evidencias de validez convergente, divergente y predictiva, así como confiabilidad por consistencia interna. En tanto que Pérez, Labiano y Brusasca (2010) analizaron las propiedades psicométricas de la escala evaluando la consistencia interna, la capacidad discriminativa de los ítems y calculando las correlaciones entre variables sociodemográficas (edad, escolaridad, sexo) y los puntajes de DS. En el Perú no se han encontrado reportes de estudios psicométricos con la escala de Marlowe-Crowne.

La MCSDS en el contexto de la investigación, ha tenido diferentes aplicaciones, por un lado, Stöber (2001) y Tatman y Kreamer (2014) la han utilizado como un instrumento para obtener evidencias de

validez vinculadas a las relaciones con otras variables, de tipo convergente y discriminante de otros instrumentos que también evaluaban la deseabilidad social. Por otro lado, Enríquez y Domínguez (2010) la utilizaron para analizar los efectos que producen las instrucciones socialmente deseables en los resultados de los cuestionarios que evalúan la capacitación. Además, Attray y Jones (2007) han sugerido que en el proceso de construcción de instrumentos psicológicos es conveniente utilizar una medida de la deseabilidad social como la Marlowe-Crowne para analizar la tendencia a responder a los ítems.

El objetivo principal del estudio fue evaluar las propiedades psicométricas de la escala de deseabilidad social de Marlowe y Crowne, aplicando las técnicas de la teoría clásica de los test. Trabajando con una población de docentes de segundo grado de primaria del Perú.

El segundo objetivo, implicó analizar las diferencias en la DS según la ubicación geográfica y el tipo de gestión de la institución educativa (IIEE) de los docentes.

Método

Tipo y diseño de investigación

Balnaves y Caputi (2001) indican que la investigación de tipo exploratoria busca, entre otras cosas, conocer las características de un instrumento de medición. Por lo tanto, la presente investigación corresponde a la categoría antes mencionada. De acuerdo con Ato, Lopez y Benavente (2013), el estudio corresponde a una investigación instrumental, en la medida que se analizan las propiedades psicométricas de la escala de DS.

En cuanto al diseño de esta investigación se puede ubicar en la categoría que Howitt y Cramer (2011) denominan no manipulativa, no se realiza la manipulación de las variables. Concretamente se trata de un diseño de investigación por encuestas de tipo transversal, pues busca describir la distribución de un fenómeno haciendo inferencias a la población a partir de una muestra recogida en un único momento en el tiempo (Fraenkel, Wallen, & Hyun, 2011).

Participantes

La población objetivo de la ECE realizada el 2010, son todas las IIEE que tienen cinco o más estudiantes matriculados en segundo grado de primaria (Ministerio de educación del Perú, 2009). La evaluación se realiza a nivel nacional y recoge información de todos los estudiantes que asisten a dichas IIEE, así como de sus docentes.

En el caso de la ECE 2010, la cobertura fue del 95% (Ministerio de educación del Perú, 2011). Del total de casos válidos (28,972), el 22.8% de los casos son varones y el 77.2% mujeres. Respecto a la edad, los varones presentan ligeramente una mayor edad ($M_{edad} = 43$; $DE_{edad} = 10$) que las mujeres ($M_{edad} = 40$; $DE_{edad} = 9$).

Para el desarrollo de los análisis psicométricos se trabajaron con dos muestras de 5000 casos cada una, las cuales fueron obtenidas de forma probabilística, conservando la estructura y composición de la población, por lo cual se realizaron dos muestreos estratificados, aplicando el módulo de muestras complejas del programa Spss v24 (IBM, 2016).

Tabla 1

Distribución de los docentes, según la Dirección Regional de Educación (DRE), gestión y zona geográfica en la que se ubica la IIEE donde labora.

	n	%
Amazonas	616	2.1
Ancash	1375	4.7
Apurímac	556	1.9
Arequipa	1147	4.0
Ayacucho	522	1.8
Cajamarca	1118	3.9
Callao	740	2.6
Cusco	1377	4.8
Huancavelica	595	2.1
Huánuco	1341	4.6
Ica	679	2.3
Junín	1437	5.0
La Libertad	1958	6.8
Lambayeque	919	3.2
Lima provincias	900	3.1
Lima Metropolitana	6005	20.7
Loreto	1526	5.3
Madre de Dios	144	0.5
Moquegua	147	0.5
Pasco	342	1.2
Piura	1994	6.9
Puno	1184	4.1
San Martín	1177	4.1
Tacna	236	0.8
Tumbes	202	0.7
Ucayali	502	1.7
En blanco	233	0.8
Tipo de gestión		
Estatal	21411	73.9
No Estatal	7561	25.3
Área Geográfica		
Urbana	18707	64.6
Rural	10265	35.4
Total	28972	100.0

Instrumento

La Escala de discapacidad social (MCSDS) fue construida por Marlowe y Crowne en 1960 y tenía por finalidad medir la tendencia a dar una imagen favorable de sí mismo, ante diversas situaciones propuestas. La escala está conformada por 33 ítems, los cuales presentan enunciados acerca de diferentes situaciones cotidianas, ante las cuales se debe responder usando un formato de tipo dicotómico (verdadero-falso). Se presentan 17 ítems de forma directa y 16 de forma inversa, los cuales se deben recodificar, de manera que se expresa la presencia de la DS. Las puntuaciones totales que se pueden obtener fluctúan entre cero (baja DS) y 33 (alta DS). Las puntuaciones derivadas de aplicar este instrumento en distintas poblaciones han mostrado elevadas estimaciones de confiabilidad, aplicando coeficientes de estabilidad, así como de consistencia interna (Crowne & Marlowe, 1960); y ha presentado diferentes evidencias de validez (Leite & Beretvas, 2005; Loo & Loewen, 2004; Loo & Thorpe, 2000).

Para el desarrollo del estudio un grupo de 7 expertos, revisaron de forma independiente la versión original en inglés de Crowne y Marlowe (1960) y las versiones en español desarrolladas por Ferrando y Chico (2000) en España y por Cosentino y Castro (2008) y por Pérez et al. (2010) en Argentina, de manera que se integró la información y se realizaron las modificaciones pertinentes a fin de garantizar la adecuación lingüística del instrumento.

Estrategias de recolección de datos

La Oficina de Medición de la Calidad de los Aprendizajes (UMC) del Ministerio de Educación viene realizando desde el año 2007 las ECE. En dicha evaluación se recoge información a nivel nacional, sobre el rendimiento de los estudiantes de segundo grado de primaria en comprensión lectora y matemáticas, así como de algunas variables que son utilizadas en los modelos multinivel de factores asociados (Christiansen, Garrett, & Marcos, 2016).

Durante la evaluación realizada en el año 2011 entre los instrumentos de factores asociados se aplicó un cuestionario al docente que recogía, entre otras variables, las respuestas a la escala de Marlowe-Crowne (Crowne & Marlowe, 1960). La recolección de datos de la ECE 2011 fue desarrollada el 29 y 30 de noviembre. Las aplicaciones se realizaron de forma tanto individual como grupal en los docentes evaluados. Al finalizar las aplicaciones la empresa encargada del proceso de lectura de los instrumentos entregó la base de datos en formato ASCII que fue utilizada para todos los análisis de datos aplicados como parte de esta investigación.

Procedimiento

El desarrollo del estudio supuso aplicar varias estrategias psicométricas basadas en la TCT, orientadas a obtener evidencias de validez, confiabilidad, y la elaboración de baremos en rangos percentilares. Inicialmente se intentó trabajar con la base de datos completa (28 972 casos), pero se optó por aplicar la metodología de la Validación cruzada como lo proponen Price (2017) y Hair et al. (2014), por lo que se trabajaron con 2 muestras obtenidas de forma estratificada de 5000 casos cada una. En esta investigación se buscaron evidencias de validez vinculadas a la estructura interna, mediante el análisis factorial confirmatorio (AFC) y el análisis factorial exploratorio (AFE), que según Thompson (2008) y Preacher y MacCallum (2003) son procedimientos que permiten analizar estructura latente de un conjunto de variables.

En la primera muestra se realizaron cuatro análisis factoriales confirmatorios a fin de establecer si en la muestra seleccionada se podía replicar la estructura latente de los principales hallazgos reportados en la bibliografía, para ello se utilizó el programa Mplus 7 (Muthén & Muthén, 2012), se aplicó el método de estimación de los mínimos cuadrados ponderados diagonalmente (WLSMV), debido a que se analizaron variables dicotómicas (0, 1) que corresponde a las variables categoriales (Li, 2016; Muthén & Muthén,

2012). Como criterios para analizar los modelos se aplicaron: RMSEA ($< .08$), el CFI (Índice de Bondad de Ajuste Comparativo, $\geq .95$) y el TLI (Índice de Tucker-Lewis, $\geq .95$) de acuerdo a lo propuesto por Bentler (1990), Ruiz, Pardo y San Martín (2010).

Luego se desarrollaron los análisis factoriales exploratorios (AFE), con la finalidad de establecer la unidimensionalidad del instrumento (Hattie, 1985). Dado que el instrumento está conformado por ítems dicotómicos, la factorización se realizó con la matriz de correlaciones tetracóricas (Burga, 2006). Además, Wilson (2005) señala que la evidencia de validez basada en la estructura interna requiere tener claridad conceptual sobre el constructo de interés. Considerando que en las investigaciones de Ferrando y Chico (2000), y Sârbescub, Rusu y Costea (2012) se encontraron evidencias que apoyan la unidimensionalidad de las medidas derivadas de la aplicación del test, se partió de dicho supuesto para analizar los resultados de la presente investigación. Es importante señalar que, en la práctica, ningún instrumento es perfectamente unidimensional, por lo que se busca tener instrumentos que represente a través de sus puntuaciones un solo factor dominante (Wright & Linacre, 1989). Para el desarrollo de los AFE se aplicó el programa Factor v 10.3.01 (Lorenzo-Seva & Ferrando, 2015).

Para verificar si se cumplía con el supuesto normalidad bivariada entre dos variables, se aplicó la prueba RMSEA de Jöreskog (2005), la cual señala que si los valores son menores de .1 no hay serios problemas vinculados a la violación de dicho supuesto. Además, se puede contrastar la hipótesis nula referida a que su valor es menor o igual a .1; al factorizar la matriz de correlaciones tetracóricas entre los 33 ítems con el método MINRES, que busca minimizar los residuos de los mínimos cuadrados ordinarios (Jöreskog, 2003), ninguno de los 528 pares de variables tuvo un valor RMSEA superior a .1, por lo que las evidencias fueron consideradas como suficientes para aceptar el supuesto de una distribución normal bivariada subyacente (Jöreskog,

2005). Para desarrollar sus cálculos se utilizó el programa Lisrel (Du Toit, Du Toit, Mels, & Cheng, 2005).

En la segunda muestra se realizó, en primer lugar, un análisis factorial confirmatorio a fin de verificar la unidimensionalidad de los ítems obtenidos en la primera muestra, de manera que se podía obtener evidencias de validez basadas en la estructura interna.

También se desarrollaron estudios factoriales confirmatorios a fin de establecer qué modelo de medición predominaba en los datos y seleccionar el indicador de confiabilidad que resultara adecuado, para ello se analizaron los modelos de medición de ítems tau equivalentes, congénicos y paralelos (Dunn, Baguley, & Brunnsden, 2014). Para la evaluación del ajuste de los modelos evaluados se utilizó la prueba general χ^2 , RMSEA ($\leq .05$), CFI ($\geq .95$), WRMR (Raíz ponderada media residual cuadrada, $\leq .05$) y para determinar qué modelo era el adecuado, se compararon los modelos aplicando los índices incrementales Δ CFI, Δ RMSEA y Δ WRMR (Chen, 2007). De manera que la confiabilidad del instrumento se calculó a través del coeficiente Omega de McDonald (McDonald, 1999; Trizano-Hermosilla & Alvarado, 2016).

Finalmente, para realizar las comparaciones de acuerdo a las variables sociodemográficas gestión del colegio y la ubicación geográfica se procedió a verificar la invariancia de la medición, por lo que se procedió secuencialmente, desde el modelo con menos restricciones hasta el modelo con más restricciones (Brown, 2006), a las cuales se aplicaron a las pruebas de invariancia de los ítems, de acuerdo al número de variables latentes (invarianza configuracional), cargas factoriales (invarianza métrica) y los interceptos (invarianza escalar). Para evaluar la invarianza, de acuerdo con Vandenberg y Lance (2000). Se analizaron los índices de bondad de ajuste, a nivel general CFI, RMSEA, y los índices incrementales de bondad de ajuste (Δ CFI y $\Delta\chi^2$), en cuales se establecieron como criterios, que las

diferencias de χ^2 , no sean estadísticamente significativas ($p < .05$) y las diferencias entre los CFI ($< .01$) siguiendo a Byrne (2011) y Usher y Pajares (2008).

En la medida que los resultados indicaron que eran invariantes de acuerdo con las variables demográficas seleccionadas, el análisis comparativo por estratos implicó obtener estadísticos descriptivos a nivel nacional considerando estratos, como la gestión (pública / privada) de la institución educativa en la que labora el docente y su ubicación geográfica (urbana / rural) y el departamento. Estas variables han sido consideradas para contrastar las hipótesis referidas a la presencia de diferencias estadísticamente significativas en las medidas promedio calculadas en el marco del análisis factorial de varianza (Field, 2009), también se calcularon los tamaños de los efectos, aplicando el coeficiente ω^2 para el efecto global y los efectos secundarios (Ferguson, 2009).

Finalmente, los baremos han sido trabajados mediante el rango percentilar (RP) que es definido por Frisbie (2005) como el porcentaje de puntuaciones menores a un valor determinado, para su cálculo se aplicó el algoritmo desarrollado por el programa Spss v24 (IBM, 2016).

Resultados

Antes de presentar los resultados de los análisis, debe señalarse que se ha seguido el enfoque moderno sobre validez (Kane, 2001, 2006), la cual es definida como el grado en el que la teoría y evidencias sustentan las interpretaciones de las puntuaciones implicadas en los usos propuestos del test (American Educational Research Association [AERA], American Psychological Association [APA], y National Council of Measurement in Education [NCME], 2014).

Análisis en la muestra 1

Inicialmente con la primera muestra se trató de establecer a partir de las estructuras factoriales encontradas en diferentes estudios previos cual se replicaba mejor en la muestra de docentes para ellos se aplicaron análisis factoriales confirmatorios, por ello se establecieron como modelos a contrastar el modelo 1 de Ferrando y Chico (2000) y de Sârbescub, Rustu y Costea (2012), quienes indicaban que existía una estructura latente unidimensional. A continuación, se incluyó el modelo 2 de Pérez et al. (2010) con dos factores ortogonales. Luego el modelo 3 de Lara Cantú (1990) con factores oblicuos y finalmente el modelo 4 de Lara Cantú y Reed (1988) con tres factores oblicuos.

Los índices de ajuste y los criterios que se usaron para evaluar los modelos analizados propuestos por Ruiz, Pardo y San Martín (2010) y por Bentler (1990), fueron el RMSEA (raíz del residuo cuadrático medio de aproximación) menor de .08; el CFI (Índice de Bondad de Ajuste Comparativo), mayor de .95 y el TLI (Índice de Tucker-Lewis) mayor de .95.

Los resultados (Ver Tabla 2) indican con claridad que los modelos presentan adecuados RMSEA (.03 - .05), pero los indicadores de bondad de ajuste CFI alcanzaron valores menores que los esperados y al presentar los TLI ($< .95$), los modelos no se ajustan adecuadamente para la muestra de estudio, por lo que los modelos probados no son válidos y se procedió a ejecutar los análisis factoriales exploratorios a fin de identificar cual es la estructura latente que subyace en la muestra estudiada.

Tabla 2
Comparación de los modelos derivados de la DS

Modelo	χ^2	GL	RMSEA	CFI	TLI
Modelo 1 Modelo unidimensional de Ferrando y Chico (2000), y Sârbescub, Rustu y Costea (2012)	5141.15 ***	495	.04	.75	.74
Modelo 2 Modelo de dos factores ortogonales de Pérez, Labiano, M. y Brusasca (2010)	2509.54 ***	188	.05	.74	.71
Modelo 3 Modelo de 2 factores de Lara Cantú (1990)	1731.80 ***	398	.03	.88	.88
Modelo 4 Modelo de tres factores de Lara Cantú y Reed (1988)	1695.30 ***	203	.04	.85	.86

Nota: *** $p < .001$,

Análisis factoriales exploratorios

Se desarrollaron tres análisis factoriales exploratorios sucesivos, por lo que se procedió a evaluar la pertinencia de su aplicación, a través de la medida de adecuación muestral de Kaiser, Meyer y Olkin, y el test de esfericidad de Bartlett (Field, 2009; Furr & Bacharach, 2008; Hair et al., 2014). Los resultados señalaron que en el primer análisis existen asociaciones estadísticamente significativas entre los ítems ($\chi^2[528, N = 5000] = 12923.8, p < .001$) y las cuales son de tamaños medios siendo lo suficientemente grandes como para justificar su factorización ($KMO = .846$). Se utilizó el método de Hull (Lorenzo-Seva, Timmerman, & Kiers, 2011) para determinar el número de factores necesarios

para explicar las covariaciones entre los ítems. Los resultados indicaron que se debían retener un factor con coeficiente GFI entre .96 y .94.

En la tabla 3 se muestran los resúmenes de las cargas factoriales de los 33 ítems, dada una solución unidimensional que en el primer análisis factorial explica el 19% de la varianza. Como puede apreciarse, los ítems 1, 2, 16, 18, 24, 29 y 31 tienen cargas inferiores a .30, que es el mínimo recomendado para considerar que un ítem corresponde a un factor (Hair et al., 2014). Por ello se decidió realizar un segundo análisis factorial, eliminando aquellos ítems con una carga menor a .30.

Tabla 3
Cargas factoriales de los ítems que conforma la prueba, en cada análisis

ítem	Primer AFE		Segundo AFE		Tercer AFE	
	F	h ²	F	h ²	F	h ²
it_01	.26	.07				
it_02	.19	.04				
it_03	.47	.22	.49	.24	.62	.39
it_04	.32	.10	.31	.10	.55	.30
it_05	.56	.31	.58	.33	.65	.42
it_06	.44	.20	.47	.22	.65	.42
it_07	.39	.15	.39	.15	.55	.32
it_08	.46	.21	.45	.20	.56	.32
it_09	.53	.28	.54	.29	.60	.36
it_10	.50	.25	.52	.27	.62	.39
it_11	.49	.24	.52	.27	.68	.45
it_12	.58	.33	.58	.34	.69	.46
it_13	.33	.11	.30	.09	.61	.38
it_14	.48	.23	.51	.26	.59	.33
it_15	.64	.41	.64	.41	.68	.45
it_16	.26	.07				
it_17	.31	.09	.26	.07		
it_18	.11	.01				
it_19	.56	.32	.57	.32	.71	.48
it_20	.05	.00				
it_21	.41	.17	.09	.01		
it_22	.42	.18	.37	.14	.66	.43
it_23	.43	.19	.44	.20	.64	.41
it_24	.24	.06				
it_25	.38	.15	.46	.22	.67	.44
it_26	.46	.21	.33	.11	.63	.40
it_27	.33	.11	.42	.17	.62	.39
it_28	.60	.36	.30	.09	.61	.38
it_29	.29	.08	.62	.39	.65	.42
it_30	.53	.28	.54	.29	.67	.44
it_31	.15	.02				
it_32	.37	.14	.38	.15	.65	.42
it_33	.40	.16	.37	.14	.62	.39

Nota: F: Carga Factorial, h²: comunalidad.

En el segundo análisis factorial, los resultados preliminares fueron adecuados como es el caso del test de esfericidad de Bartlett ($\chi^2[276, N = 5000] = 10171.6, p < .001$) y la medida de adecuación muestral de Kaiser, Meyer y Olkin (.855). El método de Hull (Lorenzo-Seva, Timmerman, & Kiers, 2011) apoya la propuesta de una solución unifactorial, tal y como se aprecia en la respectiva. Las cargas factoriales de los 26 ítems, dada una solución de un factor explica el 23% de la varianza. Como puede apreciarse en dicha tabla, los ítems 17 y 21 tienen cargas inferiores a .30. Por ello se decidió realizar un nuevo análisis factorial eliminándolos.

En el tercer análisis factorial exploratorio, los resultados tanto del test de esfericidad de Bartlett ($\chi^2[253, N = 5000] = 9087.4, p < .001$) y la medida de adecuación muestral de Kaiser, Meyer y Olkin (.838), indicaron la pertinencia de aplicar el análisis, en este caso el método de Hull (Lorenzo-Seva et al., 2011) una vez más apoya la propuesta de una solución unifactorial. Las cargas factoriales de los 24 ítems, dada una solución satisfactoria de un factor, que explica el 40% de la varianza. Como puede apreciarse en la tabla todos los ítems tienen cargas iguales o superiores a .50, que supera el mínimo recomendado para considerar que un ítem corresponde a un factor (Hair et al., 2014).

Análisis en la muestra 2

Evidencia de validez basada en la estructura interna

A fin de establecer si los resultados del AFE se corroboraban en la segunda muestra se procedió a desarrollar a manera de validación cruzada el AFC del modelo unidimensional, los resultados obtenidos indicaron que se obtuvo un $\chi^2 = 2556.87$ $Gl = 252$, con un adecuado RMSEA (.043), y elevados indicadores de bondad de ajuste CFI (.96) y TLI (.97), por lo que se puede concluir que el modelo unidimensional se ajusta a la muestra de estudio.

Estimación de la confiabilidad de las puntuaciones

La confiabilidad de las puntuaciones derivadas obtenidas al aplicar un instrumento de medición en una muestra implica que, al ser replicado el proceso de recolección de datos, los resultados serán semejantes pues las puntuaciones presentan poco error de medición (Brennan, 2001). En esta investigación dicha propiedad se estimó mediante el coeficiente de omega (Dunn, Baguley, & Brunnsden, 2014). Sijtsma (2009) señala que el coeficiente alfa de Cronbach a pesar de ser ampliamente utilizado en la investigación psicométrica, en la medida que muchas veces se le considera como una estimación del límite inferior de la confiabilidad. Siguiendo con la propuesta de este mismo autor, el problema del coeficiente alfa es que tiende a subestimar dicho límite, por lo cual también se reporta el coeficiente omega. Además, se reportan los intervalos al 95 % de confianza (Romano, Kromrey, & Hibbard, 2010). Finalmente, y en congruencia con los estándares de la AERA, APA y NCME (2014) se calculó el error estándar de medida.

A fin de establecer la pertinencia de aplicar el coeficiente Alfa de Cronbach como estimación de la confiabilidad, se procedió a evaluar el modelo de medición que sustentaban las puntuaciones derivadas, pues este coeficiente tiene como requisito el ser un mejor estimador en el caso de corresponder al modelo de medición de ítems tau-equivalentes (Trizano-Hermosilla & Alvarado, 2016; Cho & Kim, 2014; Dunn, Baguley, & Brunnsden, 2014; Zhang & Yuan, 2016). El modelo, asume que, todos los ítems de la prueba tienen la misma puntuación verdadera, o cargas de factores iguales en todos los elementos en un modelo factorial, éste es un requisito para aplicar el coeficiente alfa (Cronbach, 1951). Si se viola este supuesto, el valor de la verdadera confiabilidad será subestimado (Raykov, 1997; Graham, 2006) en una cantidad que puede variar entre 0.6 y 11.1% según la gravedad de la violación (Green & Yang, 2009). Como indican Teo y Fan (2013) por lo general los datos no cumplen con esta suposición generalmente.

Por otro lado, el modelo paralelo es más restrictivo para medir la consistencia interna, pues asume que todos los ítems presentan iguales, medias, varianzas y varianzas de error. El modelo congénico es el menos restrictivo de los modelos de medición de manera que en los ítems, las varianzas y las varianzas de error pueden ser diferentes, por ello es adecuado para un constructo unidimensional, en la medida que no utiliza suposiciones poco realistas sobre los ítems (Jöreskog & Sörbom, 1979). Por ello se recomienda aplicar el coeficiente ω de McDonald (Dunn, Baguley, & Brunsten, 2014).

Los resultados (ver tabla 7) indican que el modelo de medición congénico mostró ser el que alcanzaba un mejor ajuste ($\chi^2[104] = 584.40$, $p < .001$; CFI = .97; RMSEA = .01, IC 90% [.073, .109]; WRMR = .01) y, tomando como referencias las variaciones mínimas en el CFI ($\Delta CFI = -.12$), RMSEA ($\Delta RMSEA = .04$) y WRMR ($\Delta WRMR = .06$) respecto al modelo tau equivalente, puede concluirse que las cargas factoriales son estadísticamente diferentes y el uso del coeficiente α no está justificado, por lo que sería más conveniente aplicar el coeficiente ω (Trizano-Hermosilla & Alvarado, 2016; Dunn et al., 2014; McDonald, 1999; Jöreskog & Sörbon, 1979).

Tabla 4
Comparación de modelos de medición

	χ^2	GL	RMSEA	CFI	WRMR
Modelo congénico	584.40	152	.01	.97	.01
Modelo tau equivalente	1079.74	169	.05	.85	.07
Modelo paralelo	1351.09	188	.17	.70	.08

La estimación de la confiabilidad de las puntuaciones mediante el coeficiente ω indica que en la muestra total presenta un valor de .89 con IC 95% [.88 - .90] y un EEM = .94, Al analizar la confiabilidad según estratos demográficos se encuentra que la muestra de $IIEE_{estatales}$ el valor alcanzado fue de .85 (IC95% [.84 - .86], EEM = 1.31), en la muestra de $IIEE_{no estatales}$ presentaron un valor de .84 (IC95% [.83 - .85], EEM = 1.33), en relación con la ubicación geográfica las $IIEE_{urbanas}$ presentan un coeficiente de .85 (IC95% [.84 - .86], EEM = 1.32), en la muestra de $IIEE_{no estatales}$ presentaron un valor de .83 (IC95% [.82 - .84], EEM = 1.47).

Diferencias de discapacidad social según tipo de colegio y área geográfica

Los resultados indican que la invarianza configuracional es satisfactoria, por lo que se asumió como línea base, se analizó la invarianza métrica

(cargas factoriales) que es un nivel con más restricciones que el modelo general (Milfont & Fischer, 2010). Se encontró además un aceptable funcionamiento, pues los índices de ajuste no modificaron significativamente. Luego se analizó la invarianza escalar (el intercepto de los ítems), obteniendo resultados adecuados. Fundamentalmente, los resultados reportan primero, que los índices de ajuste SRMR ($< .064$) y RMSEA ($< .053$) alcanzaron valores aceptables, además de producirse solo ligeros cambios entre el CFI de los modelos ($< .005$), diferencias que en el CFI son similares o más pequeñas que los criterios de cambio del CFI propuestos por Chen (2007) ($< .01$) y Cheung y Rensvold (2002) ($< .005$); estos resultados indican que las propiedades de medición se mantienen satisfactoriamente invariantes en los criterios de invarianza configuracional, métrica y escalar de acuerdo a los tipo de IIEE y las áreas geográficas (ver Tabla 5).

Tabla 5
Invarianza de la deseabilidad social según tipo de Institución educativa y área geográfica

Invarianza	χ^2	GL	CFI	SRMR	RMSEA	$\Delta\chi^2$	gl	p	Δ CFI
Gestión									
Configuracional	5428.08	505	.994	.026	.061				
Métrica	5448.94	520	.995	.027	.060	20.86	15	.141	.001
Escalar	5462.14	528	.996	.027	.060	13.2	8	.105	.002
Área Geográfica									
Configuracional	5378.96	505	.980	.026	.061				
Métrica	5399.31	520	.980	.027	.060	20.35	15	.159	<.001
Escalar	5410.95	528	.980	.027	.060	11.64	8	.168	<.001

El análisis comparativo por estratos implicó obtener estadísticos descriptivos a nivel nacional considerando estratos, como la gestión (pública / privada) de la institución educativa en la que labora el docente y su ubicación geográfica (urbana / rural) y el departamento. Estas variables han sido consideradas para contrastar las hipótesis referidas a la presencia de diferencias estadísticamente significativas en las medidas promedio calculadas en el marco del análisis factorial de varianza (Field, 2009), también se calcularon los tamaños de los efectos, aplicando el coeficiente ω^2 para el efecto global y los efectos secundarios (Ferguson, 2009).

En la tabla 6 se muestran los resultados de aplicar el análisis factorial de varianza y se llega a la conclusión de que no existen diferencias estadísticamente significativas en las puntuaciones de la escala de deseabilidad social según el tipo de gestión, área geográfica o la interacción de ambas variables.

A pesar de que dichas diferencias no han sido estadísticamente significativas, se calcularon los tamaños del efecto que fueron $\omega^2 = .04$ y $\omega^2 = .02$ para gestión y área respectivamente. Estos valores pueden ser considerados pequeños (Ferguson, 2009), al igual que el tamaño del efecto calculado para el modelo total, cuyo valor fue menor que $\omega^2 = 01$.

Tabla 6
Análisis Factorial de Varianza de las puntuaciones de deseabilidad social según estratos.

Estrato	n	M	e.e.	DE	F	p	ω^2
Gestión					.137	.711	.04
Estatad	21411	17.03	.02	3.49			
No Estadad	7328	17.17	.04	3.33			
Área					.219	.640	.02
Urbana	18707	17.09	.02	3.40			
Rural	10032	17.03	.04	3.55			
Gestión x Área				.220	.639	.01	

Elaboración de baremos de las puntuaciones

Los baremos son instrumentos de aplicación profesional por lo que fueron desarrollados como rangos percentilares y se presentan en el anexo 1.

Discusión

Los estudios de la dimensionalidad de escala de Deseabilidad Social de Marlowe y Crowne, se han desarrollado en el contexto del AFE por lo que se proponen diferentes estructuras factoriales como es el caso de Ferrando y Chico, (2000); Sârbescub, et al., (2012); Pérez et al., (2010); Lara Cantú, (1990); Lara Cantú y Reed, (1988). Por ello se plantea como primer objetivo del presente estudio evaluar desde una óptica del AFC cuál de las estructura se correspondía con los resultados del análisis en los docentes de segundo grado de primaria del Perú. Para ello se extrajo una primera muestra de 5000 docentes en los cuales se efectuaron los análisis correspondientes, los cuales permitieron establecer que ninguno de los resultados previos era replicado, por lo que se optó por desarrollar una nueva estrategia aplicando el AFE en la muestra seleccionada.

Se desarrollaron tres análisis factoriales exploratorios, sucesivos, de manera que de los 33 ítems originales que conforman la prueba, sólo 24 de ellos sustentaban una solución unidimensional, lo cual fue verificado de acuerdo con los análisis estadísticos preliminares como el test de esfericidad de Bartlett y la prueba de Kaiser, Meyer y Olkin (Hair et al., 2014) y el método de Hull para establecer la significación estadística de los factores a retener (Lorenzo-Seva et al., 2011). Este hallazgo concuerda con los resultados reportados en los estudios de Ferrando y Chico (2000) en estudiantes argentinos, y Sârbescub, Rusu y Costea (2012) en estudiantes de rumanos, lo cual sirve de sustento para interpretar el constructo desde una visión unitaria, estableciendo uno de los requerimientos más importantes de un instrumento de medición como es el de medir un único rasgo o atributo latente (Wilson, 2005).

Con la finalidad de realizar la validación cruzada y corroborar si los resultados alcanzados eran consistentes, se decidió trabajar con una segunda muestra extraída de la población, en la cual aplicó el AFC, los resultados fueron favorables a la estructura unidimensional, de manera que se puede aportar

evidencias de validez vinculadas a la estructura interna, aunque una limitación del estudio podría ser el no aportar otras evidencias de validez

En relación con el estudio de la confiabilidad se aplicó el método de consistencia interna, los resultados permitieron establecer que el coeficiente a utilizarse debería ser el coeficiente ω (McDonald, 1999) el cual alcanzó un valor elevado de .89, lo cual indicaría que los ítems son homogéneos (Muñiz, 1999). Además, en concordancia con los estándares de la AERA, APA y NCME (2015) se calculó el error estándar de medida a fin de establecer la desviación estándar de los errores de medida asociados a las puntuaciones observadas de un test, para una muestra de examinados. Este valor puede ser utilizado para calcular los intervalos de confianza de las puntuaciones observadas, y estimar las puntuaciones verdaderas.

El segundo objetivo fue realizar la comparación de los resultados de la DSen los docentes de segundo grado de primaria, según variables demográficas, de manera que los análisis de la invarianza de la estructura factorial unidimensional, permiten concluir que las puntuaciones fueron invariantes. La comparación poblacional de la DS indicó que no existen diferencias estadísticamente significativas según el tipo de gestión, el área geográfica y la interacción de ambas variables. Es decir, este fenómeno se presenta de forma similar, en independencia del contexto sociodemográfico analizado. Dada la ausencia de estudios que analicen las diferencias de deseabilidad social según estas variables, este resultado no ha podido ser comparado con otras investigaciones antecedentes.

Finalmente debe señalarse que los resultados del estudio han demostrado la importancia de medir la deseabilidad social, es decir la tendencia de auto-attribuirse características de personalidad socialmente deseables y rechazar aquellas indeseables, al realizar investigaciones psicológicas, en la medida que en el contexto educativo se encuentra presente en los

procesos de interacción social, por lo que resulta necesario tomarlo en cuenta especialmente cuando se realizan evaluaciones de los factores propios de los estudiantes y de los docentes con la finalidad de obtener resultados más precisos en el contexto de los modelos multinivel de factores asociados.

Considerando los resultados de esta investigación, se recomienda utilizar la escala de deseabilidad Social de Marlowe y Crowne (MCSDS) en docentes a nivel nacional. Dado que únicamente se trabajó con esta población, se sugiere analizar las propiedades psicométricas de las puntuaciones derivadas de aplicar este instrumento en otras poblaciones, poniendo énfasis en el modelo de medición que la sustenta y el análisis de su invarianza.

Referencias

- American Educational Research Association. (AERA), American Psychological Association (APA), & National Council of Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Ato, M., López, J. J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 9(3), 1038-1059. Doi: 10.6018/analesps.29.3.178511.
- Attray, J., & Jones, M. C., (2007). Essential elements of questionnaire design and development. *Journal of Clinical Nursing*, 16, 234-243. doi: 10.1111/j.1365-2702.2006.01573.x.
- Balnaves, M., & Caputi, P. (2001). *Introduction to quantitative research methods. An investigative approach*. London: Sage Publications.
- Barger, S. D. (2002). The Marlowe-Crowne Affair: Short Forms, Psychometric Structure and Social Desirability. *Journal of Personality Assessment*, 79(2), 286-305. doi: 10.1207/S15327752JPA7902_11
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.
- Beretvas, S. N., Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurements*, 62, 570-589. doi: 10.1177/0013164402062004003
- Brennan, R.L. (2001). An essay on the history and future of reliability from the perspective of replications. *Journal of Educational Measurement*, 38(4), 295-317. doi: 10.1111/j.1745-3984.2001.tb01129.x
- Brown, T.A. (2006) *Confirmatory factor analysis for applied research*. New York: Guilford.
- Burga, A. A. (2006). La unidimensionalidad de un instrumento de medición, perspectiva factorial. *Revista de Psicología. Pontificia Universidad Católica*, 25(1), 53-80.
- Byrne, B. M. (2011). *Structural Equation Modeling with Mplus: Basic Concepts, Applications, and Programming*. New York, NY: Routledge.
- Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*. 14(3), 464-504. doi: 10.1080/10705510701301834
- Christiansen, A., Garrett, P., & Marcos, D. (2016). *Regiones en perspectiva: la influencia de los factores asociados al aprendizaje al término de la educación primaria (Estudios Breves No. 2)*. Lima: Ministerio de educación del Perú. Recuperado de http://umc.minedu.gob.pe/wp-content/uploads/2016/07/EB02_Regiones_en_perspectiva_VF.pdf
- Cosentino, A. C., & Castro, A. (2008). Adaptación y validación de la Marlowe-Crowne Social Desirability Scale. *Interdisciplinaria*, 25(2), 197-216.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. doi:10.1007/BF02310555.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349-354. doi: 10.1037/h0047358.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255.

- Cho E., & Kim S. (2014). Cronbach's coefficient alpha: well-known but poorly understood. *Organizational Research. Methods*, 18(2), 207–230. doi: 10.1177/1094428114555994.
- Domínguez Espinoza, A., Aguilera Mijares, S., Acosta Canales, T., Navarro Contreras, G., & Ruiz Paniagua, Z. (2012). La deseabilidad social revalorada: más que una distorsión, una necesidad de aprobación social. *Acta de investigación psicológica*, 2(3), 808-824. Recuperado de: <http://www.scielo.org.mx/pdf/aip/v2n3/v2n3a5.pdf>
- Du Toit, S.H.C., Du Toit, M., Mels, G., & Cheng, Y. (2005). *Analysis of complex survey data using LISREL*. Lincolnwood, IL: Scientific Software International, Inc. (Unpublished report).
- Dunn, T. J., Baguley, T., & Brunsdén, V. (2014). From alpha to Omega: A practical solution to the pervasive problema of internal consistency estimation. *British Journal of Psychology*, (3), 399-412. doi : 10.1111/bjop.12046-
- Enríquez, F. J., & Domínguez, A. C. (2010). Influencia de la deseabilidad social (DS) en reportes de capacitación. *Psicología Iberoamericana*, 18(1), 69-79.
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40(5), 532–538. doi: 10.1037/a0015808
- Ferrando, P. J., & Chico, E. (2000). Adaptación y análisis psicométrico de la escala de deseabilidad social de Marlowe Crowne. *Psicothema*, 12(3), 383-389. Recuperado de <http://www.psicothema.com/pdf/346.pdf>.
- Field, A. (2009). *Discovering statistics using SPSS (and sex and drugs and rock «n» roll)* (3a. ed.). Londres: Sage Publications.
- Fraenkel, J., Wallen, N., & Hyun, H. (2011). *How to Design and Evaluate Research in Education*. Atlanta: McGraw-Hill Education.
- Frisbie, D. A. (2005). Measurement 101: some fundamentals revisited. *Educational Measurement: Issues and Practice*, 24(3), 21-28. doi: 10.1111/j.1745-3992.2005.00016.x
- Furnham, A. (1990). *The Protestant Work Ethic: The psychology of work-related beliefs and behaviours*, London: Routledge.
- Furr, R. M., & Bacharach, V. R. (2008). *Psychometrics, an introduction*. Los Angeles: Sage.
- Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability what they are and how to use them. *Educational and Psychological Measurement*, 66(6), 930-944. doi: 10.1177/0013164406288165.
- Green, S. B., & Yang, Y. (2009). Reliability of summed item scores using structural equation modeling: an alternative to coefficient Alpha. *Psychometrika* 74(1), 155–167. doi: 10.1007/s11336-008-9099-3.
- Greenwald, H. J., & Satow, Y. (1970). A Short Social Desirability Scale. *Psychological Reports*, 27, 131-135.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7ª ed.). Essex: Pearson.
- Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9(2), 139-164. doi: 10.1177/014662168500900204
- Howitt, D., & Cramer, D. (2011). *Introduction to research methods in psychology* (3ª ed.). Essex: Pearson Education.
- Jöreskog, K. G., & Sörbom, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt Books.
- IBM. (2016). *Manual Sample Complex 24*. Armonk, NY: IBM corporation.
- Jöreskog, K. G. (2003). *Factor analysis by MINRES*. Recuperado del sitio de Internet de Scientific Software International (SSI). Recuperado de <http://www.ssicentral.com/lisrel/techdocs/minres.pdf>
- Jöreskog, K. G. (2005). *Structural equation modeling with ordinal variables using LISREL*. Scientific Software International (SSI). Recuperado de <http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf>
- Kane, M. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319-342. doi: 10.1111/j.1745-3984.2001.tb01130.x

- Kane, M. (2006). Validation. En R. L. Brennan (Eds.), *Educational measurement*, (4a ed., pp. 17-64). Westport, CT: American Council on Education y Praeger Publishers.
- King, M. F., & Bruner, G. C. (2000). Social desirability bias: A neglected aspect of validity testing. *Psychology and Marketing*, 17(2), 79-103. doi: 10.1002/(SICI)1520-6793(200002)17:2%3C79::AID-MAR2%3E3.0.CO;2-0
- Lara-Cantú, M. A. (1990). Validez y confiabilidad de la Escala de Deseabilidad Social de Marlowe y Crowne en una población de adultos. *Salud Mental*, 13, 35-38.
- Lara-Cantú, M. A., & Reed, Z. M. (1988). La escala de deseabilidad social de Marlowe y Crowne: Un estudio psicométrico. *Salud Mental*, 11(3), 25-29.
- Leite, W. L., & 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*, 65, 140-154. doi: 10.1177/0013164404267285
- Li, Ch. (2016) Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavioral Research Methods*, 48(3), 936-49. doi: 10.3758/s13428-015-0619-7
- Loo, R. & Thorpe, K. (2000). Confirmatory factor analyses of the full and short versions of the Marlowe-Crowne Social Desirability Scale. *The Journal of Social Psychology*, 140(5), 628-635. doi: 10.1080/00224540009600503
- Loo, R., & Loewen, P. (2004). Confirmatory factor analysis of scores from full and short versions of the Marlowe-Crowne Social Desirability Scale. *Journal of Applied Social Psychology*, 34, 2343-2352. doi: 10.1111/j.1559-1816.2004.tb01980.x
- Lorenzo-Seva, U., & Ferrando, P. J. (2015). *Factor v.10.3.01* [programa informático]. Recuperado del sitio de internet de la Universitat Rovira i Virgili. Recuperado de <http://psico.fcep.urv.es/utilitats/factor/soft/factor10.3.64bits.zip>
- Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research*, 46, 340-364. doi: 10.1080/00273171.2011.564527
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Nueva Jersey: L. Erlbaum Associates.
- Mick, D.G. (1996). Are studies of dark side variables confounded by socially desirable responding? The case of materialism. *The Journal of Consumer Research*, 23(2), 106-119. doi: 10.1086/209470
- Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3, 112-131.
- Ministerio de Educación del Perú. (2009). *Evaluación Censal de Estudiantes (ECE). Segundo grado de primaria y cuarto grado de primaria de IE EIB. Marco de Trabajo*. Recuperado de http://umc.minedu.gob.pe/wp-content/uploads/2014/07/Marco_de_Trabajo_ECE.pdf.
- Ministerio de Educación del Perú. (2011). *Resultados de la Evaluación Censal de Estudiantes 2010 – ECE 2010. Segundo grado de Primaria (2P)* [Diapositivas de PowerPoint]. Recuperado de http://www2.minedu.gob.pe/umc/ece2010/Resultados_ECE2010Segundo_grado.pdf
- Muthén, L. A., & Muthén, B. O. (2012). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.
- Muñiz, J. (1999). *Teoría Clásica de los Tests*, (2ª. ed.). Madrid: Pirámide.
- Neeley, S. M., & Cronley, M. L. (2004). When research participants don't tell it like it is: pinpointing the effects of social desirability bias using self vs. indirect-questioning. *Advances in Consumer Research*, 31, 432-433.
- Pérez, M. J., Labiano, M., & Brusasca, C. (2010). Escala de Deseabilidad Social: análisis psicométrico en muestra argentina. *Evaluar*, 10, 53-67.
- Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, 2(1), 13-43. Recuperado de http://quantpsy.org/pubs/preacher_maccallum_2003.pdf

- Price, L. (2017). *Psychometric methods: theory into practice*. New York: The Guilford Press.
- Ramanaiah, N. V., & Martin, H. J. (1980). On the two dimensional nature of the Marlowe-Crowne Social Desirability Scale. *Journal of Personality Assessment*, 44(5), 505-514.
- Ramanaiah, N.V., Schill, T., & Leung, L. S. (1977). A test of the hypothesis about the two-dimensional nature of the Marlowe-Crowne Social Desirability Scale. *Journal of Research in Personality*, 11, 251-259. doi: 10.1016/0092-6566(77)90022-8
- Ray, J. J. (1984). The reliability of short Social Desirability Scales. *The Journal of Social Psychology*, 123, 133-134. doi: 10.1080/00224545.1984.9924522
- Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21, 173-184.
- Romano, J. L., Kromrey, J. D., & Hibbard, S. T. (2010). A Monte Carlo study of eight confidence interval methods for coefficient alpha. *Educational and Psychological Measurement*, 70(3), 376-393. doi: 10.1177/0013164409355690.
- Rudmin, F. W. (1999). Norwegian short form of the Marlowe-Crowne Social Desirability Scale. *Scandinavian Journal of Psychology*, 40, 229-233. doi: 10.1111/1467-9450.00121.
- Ruiz, M., Pardo, A., & San Martín, R. (2010). Modelos de Ecuaciones Estructurales. *Papeles del Psicólogo*, 31(1), 34-45. Recuperado de <http://www.papelesdelpsicologo.es/pdf/1794.pdf>.
- Sârbescub, P., Rustu, S., & Costea, L. (2012). The Marlowe-Crowne Social Desirability Scale in Romania: the development of a 13-item short form. *Romanian Journal of Experimental Applied Psychology*, 3(2), 37-44. Recuperado de <http://www.rjeap.ro/files/vol3no2/6art5.pdf>.
- Scagliusi, F. B., Cordás, T. A., Polacow, V. O., Coelho, D., Alvarenga, M., Philippi, S. T., & Lancha Jr, A. H. (2004). Tradução da escala de desejo de aceitação social de Marlowe y Crowne para a língua portuguesa Translation into Portuguese of the Marlowe-Crowne Social Desirability Scale. *Revista de Psiquiatria Clínica*, 31(6), 272-278. doi: 10.1590/S0101-60832004000600001.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 4(1), 107-120. doi: 10.1007/s11336-008-9101-0
- Soubelet, A., & Salthouse, T. A. (2011). Influence of social desirability on age differences in self-reports of mood and personality. *Journal of Personality*, 79(4), 741-762. Doi: 10.1111/j.1467-6494.2011.00700.x.
- Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment*, 17, 222-232. doi:10.1027//1015-5759.17.3.222.
- Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlow-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, 28(2), 191-193. doi: 10.1002/1097-4679(197204)28:2%3C191::AID-JCLP2270280220%3E3.0.CO;2-G
- Sullman, M. J. M., & Taylor, J. E. (2010). Social desirability and self-reported driving avoidance: Should we be worried?. *Transportation Research*, 13, 215-221. doi: 10.1016/j.trf.2010.04.004
- Tatman, A. W., & Kreamer, S. (2014). Psychometric properties of the Social Desirability Scale-17 with individuals on probation and parole in the United States. *International Journal of Criminal Justice Sciences*, 9(1), 122-130. Recuperado de <http://www.sascv.org/ijcjs/pdfs/Tatman&Kreamerijcjs2014vol9issue1.pdf>.
- Teo, T., & Fan, X. (2013). Coefficient alpha and beyond: issues and alternatives for educational research. *Asia Pacific Educational Researcher*, 22(2), 209-213. doi: 10.1007/s40299-013-0075-z.
- Thompson, B. (2008). *Confirmatory and Confirmatory Factor Analysis: understanding concepts and applications*. Washington, DC: American Psychological Association.
- Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best Alternatives to Cronbach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements. *Frontiers in Psychology*, 7, 769. doi: 10.3389/fpsyg.2016.00769.

- Usher, E. L., & Pajares, F. (2008). Self-efficacy for self-regulated learning a validation study. *Educational and Psychological Measurement*, 68(3), 443-463.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-70.
- Verardi, S., Dahourou, D., Ah-Kion, J., Bhowon, U., Tseung, C. N., Amoussou-Yeye, D., ... Rossier, J. (2009). Psychometric properties of the Marlowe-Crowne Social Desirability Scale in eight african countries and switzerland. *Journal of Cross-Cultural Psychology*, 41(1), 19-34. doi: 10.1177/0022022109348918
- Vu, A., Tran N., Pham K., & Ahmed, S. (2011). Reliability of the Marlowe-Crowne social desirability scale in Ethiopia, Kenya, Mozambique, and Uganda. *BMC Medical Research Methodology*, 11, 1-7. doi: 10.1186/1471-2288-11-162
- Wilson, M. (2005). *Constructing measures. An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wright, B., & Linacre, J. (1989). Observations are always ordinal; measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation*, 70(12), 857-860.
- Zhang, Z., & Yuan, K. (2016). Robust Coefficients Alpha and Omega and Confidence intervals with outlying observations and missing data: Methods and Software. *Educational and Psychological Measurement* 76(3), 1-25. doi: 10.1177/0013164415594658

Andrés Burga León

Universidad de Lima, Perú

Asesor en temas de psicometría, estadística y metodología de la investigación cuantitativa de la Oficina de Medición de la Calidad de los Aprendizajes del Ministerio de Educación del Perú. Docente Asociado de la Universidad de Lima. Miembro del equipo ganador del concurso 2016 organizado por el Consorcio de Investigación Económica y Social (CIES). Autor de diversos artículos dentro de su especialidad

Autor corresponsal: aburga@ulima.edu.pe

Luis Ecurra Mayaute

Universidad de Lima, Perú

Psicólogo, Magister en problemas de aprendizaje, Docente universitario y de postgrado. Investigador afiliado al Registro Nacional de Investigadores en Ciencia y Tecnología (REGINA) con el número 6408, Realiza estudios sobre temas psicométricos, análisis estadísticos y psicología educativa.

lescurra@ulima.edu.pe

ANEXO 1

Tabla 7

Baremos de las puntuaciones, expresados mediante rangos percentilares.

puntaje	RP
23	99
22	95
21	88
20	79
19	68
18	56
17	45
16	35
15	26
14	19
13	13
12	9
11	6
10	4
9	2
7-8	1
0-6	0

ANEXO 2

Escala de deseabilidad social de Marlowe-Crowne

A continuación, encontrará una serie de frases que están relacionadas con actitudes personales. Lea atentamente cada una de ellas y decida si su forma habitual de ser o actuar se parece (V) o no (F) al contenido de la frase. Por favor, no deje ninguna frase sin responder.

01. Antes de votar me informo detalladamente de la capacidad de todos los candidatos.
02. Siempre dejo lo que estoy haciendo para ayudar a alguien con problemas.
03. A veces me cuesta ponerme a trabajar si no me encuentro con ánimos.
04. Nunca me ha caído realmente mal una persona.
05. Algunas veces dudo de mi habilidad para triunfar en la vida.
06. A veces estoy descontento cuando no puedo hacer las cosas a mi manera.
07. Siempre soy muy cuidadoso con mi manera de vestir.
08. En casa, me comporto tan bien en la mesa como cuando voy a un restaurante.
09. Si pudiera entrar en una sala de cine sin pagar y estuviera seguro de que no me verán, probablemente lo haría.
10. En algunas ocasiones he renunciado a hacer algo porque pensaba que me faltaba la habilidad necesaria para hacerlo.
11. A veces me gusta chismear un poco.
12. Ha habido veces en que he tenido sentimientos de rebeldía contra personas con autoridad, sabiendo que ellos tenían la razón.
13. Independientemente de con quién esté hablando, siempre escucho atentamente.
14. Alguna vez me «he hecho el loco» para quitarme a alguien de encima.
15. Alguna vez me he aprovechado de alguien.
16. Cuando cometo un error siempre estoy dispuesto a admitirlo.
17. Siempre intento practicar lo que predico.
18. No encuentro particularmente difícil relacionarme con gente escandalosa y detestable.
19. A veces trato de vengarme en lugar de perdonar y olvidar lo que me han hecho.
20. Cuando no sé algo no me importa admitirlo.
21. Siempre soy cortés, incluso con gente desagradable.
22. A veces insisto en hacer las cosas a mi manera.
23. En alguna ocasión he sentido que puedo ser un poco torpe con las manos o mi cuerpo.
24. Nunca he dejado que alguien fuera castigado por cosas que había hecho yo.
25. Nunca me enfado cuando me piden que devuelva algún favor que me han hecho.
26. Nunca me enojo cuando la gente expresa ideas muy distintas de las mías.
27. Nunca emprendo un viaje largo sin revisar el auto (moto, bicicleta, etc.).
28. En algunas ocasiones me he sentido bastante celoso de la buena suerte de los demás.
29. No he tenido nunca la necesidad de decirle a alguien que me dejara en paz.
30. A veces me irrita la gente que me pide favores.
31. Nunca he considerado que me han castigado sin motivo.
32. A veces pienso que cuando la gente tiene mala suerte es porque se lo merece.
33. Nunca he dicho nada que pudiera herir los sentimientos de alguien.

ANEXO 3

Sintaxis del programa LISREL utilizada para el cálculo de RMSEA

```
SY='C:\Users\Andres\Dropbox\Universidades\UL\IDIC\Desebilidad social\Bases e  
instrumentos\factor\data_DS.PSF'SE 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24  
25 26 27 28 SE 29 30 31 32 33 FA NF=1OU MA=CM XM
```

Sintaxis del programa SPSS utilizada para estimar la confiabilidad de las puntuaciones y discriminación de los ítems

```
RELIABILITY /VARIABLES= cdo24_03 cdo24_04 cdo24_05 cdo24_06 cdo24_07 cdo24_08  
cdo24_09 cdo24_10 cdo24_11 cdo24_12 cdo24_13 cdo24_14 cdo24_15cdo24_19 cdo24_21  
cdo24_22 cdo24_23 cdo24_25 cdo24_26 cdo24_27cdo24_30 cdo24_32 cdo24_33 /SCALE('DS  
Marlowe-Crowne') ALL /MODEL=ALPHA/STATISTICS=DESCRIPTIVE /SUMMARY=TOTAL  
MEANS.
```

Sintaxis del programa SPSS utilizada para obtener los estadísticos descriptivos en la muestra

```
CTABLES  
/VLABELS VARIABLES=gestion area total_ds DISPLAY=LABEL  
/TABLE gestion [C] + area [C] BY total_ds [S][COUNT F40.0, MEAN F40.2, SEMEAN F40.2,  
STDDEV F40.2]  
/CATEGORIES VARIABLES=gestion ORDER=A KEY=VALUE EMPTY=INCLUDE TOTAL=YES  
POSITION=BEFORE  
/CATEGORIES VARIABLES=area ORDER=A KEY=VALUE EMPTY=INCLUDE.
```

Sintaxis del programa SPSS utilizada para aplicar el modelo de análisis factorial de varianza

```
UNIANOVA total_ds BY gestion area  
/METHOD=SSTYPE(3)  
/INTERCEPT=INCLUDE  
/SAVE=ZRESID  
/EMMEANS=TABLES(area*gestion)  
/PRINT=ETASQ  
/CRITERIA=ALPHA(.05)  
/DESIGN=gestion area area*gestion.
```