



RELIEVE. Revista Electrónica de
Investigación y Evaluación Educativa

E-ISSN: 1134-4032

relieve@uv.es

Universitat de València
España

Martínez-Abad, Fernando; Rodríguez-Conde, María José
Comportamiento de las correlaciones producto-momento y tetracórica-policórica en
escalas ordinales: un estudio de simulación
RELIEVE. Revista Electrónica de Investigación y Evaluación Educativa, vol. 23, núm. 2,
julio-diciembre, 2017, pp. 1-21
Universitat de València
Valencia, España

Disponible en: <http://www.redalyc.org/articulo.oa?id=91655440003>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica
Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Comportamiento de las correlaciones producto-momento y tetracórica-policórica en escalas ordinales: un estudio de simulación

Behaviour of product-moment and tetrachoric-polychoric correlations in ordinal scales: a simulation study

Martínez-Abad, Fernando & Rodríguez-Conde, María José
Universidad de Salamanca (España)

Abstract

The statistical multivariate analysis of Likert response scales, given their widespread use, is a controversial issue in the scientific community, mainly from the specification of the problem of measurement. This work aims to study various conditions of these ordinal scales affect the calculation of the product-moment and tetrachoric-polychoric correlation coefficients. For this purpose, a simulation study was applied in which 90 databases with 10 items each were generated. In the estimation of the databases, the following variables were controlled: number of response categories, symmetrical or asymmetric distributions of data, sample size and level of relationship between items. Thus, 90 matrices (10x10) were obtained which included the difference between the product-moment and tetrachoric-polychoric correlations. The graphical and variance analysis show how the product-moment correlation coefficient significantly underestimates the relationship between variables mainly when the number of response categories of the ordinal scale is small and the relationship between the variables is large. On the other hand, the statistical estimation of both coefficients is very similar when the starting relationship between pairs of variables is small and/or when the number of response options of the variables is greater than 5. The study concludes by making a recommendation to the applied researcher on the most appropriate correlation coefficient depending on the type of data available. Finally, the results are discussed from the previous studies, which reach some similar conclusions.

**Reception
Date**
2017
January 6

**Approval
Date**
2017
December 17

**Publication
Date:**
2017
December 19

Keywords: Simulation, Multivariate analysis, correlation analysis, Product-moment correlation, Tetrachoric correlation, Polychoric correlation, Measurement, Attitude scale

Resumen

El análisis estadístico multivariante de escalas de respuesta tipo Likert, dado su empleo generalizado, resulta un tema controvertido en la comunidad científica, principalmente desde la especificación del problema de la medida. Este trabajo tiene como objeto estudiar cómo afectan diversas condiciones de estas escalas ordinales al cálculo de los coeficientes de correlación producto-momento y tetracórica-policórica. Para ello, se aplica un estudio de simulación en el que se generaron 90 bases de datos con 10 ítems cada una, controlando las siguientes variables: número de categorías de respuesta, distribución simétrica o asimétrica de los datos, tamaño de la muestra y nivel de relación entre los ítems. Así, se obtuvieron 90 matrices (10x10) que incluyeron la diferencia entre la correlación producto-momento y tetracórica-policórica. El análisis gráfico y de varianza muestran cómo la estimación producto-momento infravalora, en contraste con la estimación tetracórica-politórica, la relación entre las variables principalmente cuando el número de opciones de respuesta de la escala ordinal es pequeño y la relación entre las variables grande. Por su parte, las estimaciones de ambos coeficientes son muy similares cuando la relación de partida entre las parejas de variables es pequeña y/o cuando el número de opciones de respuesta de las variables es mayor que 5. El trabajo concluye realizando una serie de recomendaciones generales al investigador aplicado sobre el coeficiente de correlación que se considera más pertinente, en base a las evidencias recabadas, en función del tipo de datos disponibles, y se discuten los resultados con estudios previos que alcanzan algunas conclusiones similares.

**Fecha de
recepción**
2017
Enero 06

**Fecha de
aprobación**
2017
Diciembre 17

**Fecha de
publicación**
2017
Diciembre 19

Palabras clave: Simulación, Análisis multivariado, Análisis de correlación, Correlación producto-momento, Correlación tetracórica, Correlación policórica, Medida, Escala de actitud

Corresponding author / Autor de contacto

Fernando Martínez Abad. Facultad de Educación. Universidad de Salamanca (España). fma@usal.es

La medida formal de rasgos de personalidad, atributos, actitudes, disposiciones, etc. en disciplinas relacionadas con la Psicología, las Ciencias Sociales o incluso las Ciencias de la Salud, ha alcanzado un alto grado de desarrollo (Morales Vallejo, 2000; Morales Vallejo, Urosa, & Blanco, 2003). Para asegurar estimaciones más precisas se suceden y desarrollan, principalmente desde inicios del siglo XX, un conjunto de estadísticos de naturaleza multivariante hasta adquirir en la actualidad niveles de complejidad muy elevados (Abad, 2011; Freiberg Hoffmann, Stover, de la Iglesia, & Fernández Liporace, 2013; Lévy Mangin, 2006; López González, 2012; Pearson, 1900, 1910), gracias en gran parte a la capacidad de cálculo de los procesadores modernos y a la generalización de paquetes estadísticos especializados.

Se pone en evidencia al respecto el empleo generalizado de escalas de medida de naturaleza no métrica para la estimación de constructos (Ferreira & Backhoff-Escudero, 2016; González-González, Álvarez-Castillo, & Fernández-Caminero, 2015; Olmos Migueláñez, Martínez Abad, Torrecilla Sánchez, & Mena Marcos, 2014; Pearse, 2011; Preston & Colman, 2000; Shafel, Brooke, & Gillmor, 2012), manteniéndose *el problema de la medida* (Marcus-Roberts & Roberts, 1987; Muñiz, 1998; Stevens, 1946) aún muy presente en la investigación aplicada: “Although it is common practice to use Pearson’s r to estimate the association between two sets of ordinal data, the validity of the results from such analysis is quite questionable” (Choi, Peters, & Mueller, 2010, p. 465).

A pesar de las numerosas técnicas estadísticas disponibles que respetan la naturaleza de las escalas de medida, lo cierto es que no acaban de sustituir a las empleadas tradicionalmente. Se pueden enumerar varias razones que lo explican (López González, 2012):

- Creencia generalizada de que las técnicas alternativas (en la mayor parte de los casos no paramétricas) poseen una menor

potencia estadística que las tradicionales. Si bien esto es cierto (Corder & Foreman, 2009; Siegel, 1970), la pérdida es mínima, más aun teniendo en cuenta los tamaños de muestra empleados habitualmente en los estudios actuales.

- Reducción en la oferta de prestaciones y versatilidad con respecto a las técnicas clásicas.
- Desconocimiento por parte de los investigadores (por habituación a las técnicas clásicas) o ausencia de las técnicas alternativas en la oferta de los principales paquetes estadísticos, de empleo generalizado.

Como cuestión añadida a estos factores, parte de la comunidad científica de Ciencias Sociales y de la Salud mantiene el convencimiento de que el empleo de las técnicas clásicas supone un sesgo mínimo, incidiendo en que la pérdida de prestaciones sufrida con el cambio no se suple por la pequeña ganancia en la reducción del error de estimación (Morales Vallejo, 2000; Nunnally, 2010).

Así, en un punto intermedio entre el empleo de las técnicas alternativas y el mantenimiento de los procedimientos clásicos se encuentra el cálculo la matriz de correlaciones *tetracóricas* (entre variables dicotómicas) o *policóricas* (entre variables ordinales politómicas) para la estimación insesgada de la relación entre las parejas de ítems de una escala con medida no numérica (Choi et al., 2010; Freiberg Hoffmann et al., 2013; Gilley & Uhlig, 1993; Holgado-Tello, Chacón-Moscoso, Barbero-García, & Vila-Abad, 2008; Lara, 2014; Morata-Ramírez & Holgado-Tello, 2013; Panter, Swygert, Grant Dahlstrom, & Tanaka, 1997; Saris & Coenders, 1995). De este modo, es posible aprovechar el potencial de las técnicas clásicas sin cometer un sesgo importante en la estimación de las relaciones entre las variables de naturaleza categórica. Al respecto cabe destacar que, en el marco del presente estudio, en base al estado de la cuestión, se entiende el estimador tetracórico-policórico como el coeficiente de correlación más apropiado en los casos en los que se

aplican técnicas estadísticas multivariantes con variables obtenidas a partir de una escala de respuesta tipo Likert, en la que se puede entender que subyace una escala continua latente a las categorías de la propia escala observada.

Este coeficiente de correlación, definido teóricamente en primer lugar por el propio

Pearson (1900), recurre a la obtención de las variables continuas subyacentes (latentes) que existen entre cada par de ítems ordinales (x , y), para lo cual aprovecha las propiedades de la función de densidad de la distribución normal bivalente:

$$\Phi(a_i, b_j, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{a_i} \int_{-\infty}^{b_j} \exp \frac{1}{(2(1-\rho^2))} (x^2 - 2\rho xy + y^2) dx dy$$

Figura 1. Función de densidad de la distribución normal bivalente

en donde debemos estimar mediante procedimientos de máxima verosimilitud los parámetros de los umbrales a_i y b_j , y el coeficiente de correlación entre las variables continuas latentes ρ (Holgado-Tello et al., 2008). Este coeficiente ρ , estimado a partir de la función de densidad de la distribución normal bivalente, es el que denominamos *coeficiente de correlación tetracórico o policórico*. Es importante destacar, por tanto, a pesar de que multitud de estudios demuestran que la falta de normalidad en las variables originales no afecta de manera significativa al cálculo de este coeficiente (Freiberg Hoffmann et al., 2013; Morata-Ramírez & Holgado-Tello, 2013), que su computación trae aparejado este supuesto básico. Y es sabido que en buena parte de las escalas aplicadas en Ciencias Sociales, principalmente los estudios de encuesta mediante escalas de respuesta tipo Likert (Serrano Angulo, Cebrián Robles & Serrano Puerto, 2015), la normalidad univariante y multivariante se viola de manera generalizada y extrema.

Aunque es evidente la complejidad que supone el cálculo de la correlación tetracórica-policórica, la capacidad de procesamiento de los ordenadores modernos posibilita su empleo en el cálculo de matrices de correlaciones, incluso en los casos en los que se dispone de variables ordinales con múltiples niveles de respuesta y tamaños de muestra grandes. Así, la producción científica

alrededor del estudio de las diferentes estimaciones que realizan el coeficiente de correlación producto-momento y el coeficiente de correlación tetracórico-policórico en variables no cuantitativas ha sido amplia en los últimos años.

La mayor parte de estos estudios se centran en comprobar cómo se comportan los índices generales de fiabilidad y validez en el análisis factorial exploratorio y confirmatorio a partir de ambas matrices de correlaciones (Burga León, 2012; Freiberg Hoffmann et al., 2013; Gilley & Uhlig, 1993; Holgado-Tello et al., 2008; Muthén & Kaplan, 1992; Panter et al., 1997; Richaud, 2005), centrando su atención de manera prácticamente exclusiva sobre el número de categorías de respuesta de la variable ordinal (Bandalos & Enders, 1996; Birkett, 1986; Chan, 1991; Cicchetti, Shoinralter, & Tyrer, 1985; Cox, 1980; García Cueto, Muñiz Fernández, & Hernández Baeza, 2000; Lozano, García-Cueto, & Muñiz, 2008; Matell & Jacoby, 1972; Maydeu-Olivares, Kramp, García-Forero, Gallardo-Pujol, & Coffman, 2009; Oliden & Zumbo, 2008; Preston & Colman, 2000; Shafel et al., 2012; Weijters, Cabooter, & Schillewaert, 2010; Weng, 2004). La mayor parte de estos estudios parecen coincidir en que, para evitar los sesgos asociados al empleo del coeficiente de correlación producto-momento, es necesario que las variables originales posean, al menos, 5 niveles de respuesta (Choi et al., 2010; García

Cueto et al., 2000; Holgado-Tello et al., 2008; Lozano et al., 2008; Oliden & Zumbo, 2008; Preston & Colman, 2000; Weijters et al., 2010; Weng, 2004).

No se localizan, no obstante, estudios que analicen directamente las diferencias entre las matrices de correlaciones producto-momento y tetracórica-policórica resultantes de una serie de ítems ordinales. Por otro lado, se publican sólo de manera aislada trabajos que tengan en cuenta otros factores que afectan a la estimación de ambos coeficientes de correlación, como el nivel de relación entre las parejas de variables (Lozano et al., 2008), los niveles de asimetría de las distribuciones marginales de las variables ordinales o el tamaño de la muestra disponible en el conjunto de datos.

A pesar de todo ello, se constata que la información disponible para el investigador aplicado sobre las condiciones y beneficios del empleo del coeficiente de correlación tetracórico-policórico no es clara, es insuficiente, e incluso en ocasiones es contradictoria. Mientras que en algunos foros se defiende apasionadamente la necesidad de la estimación tetracórica-policórica en el caso de cualquier escala dicotómica u ordinal (López González, 2012), otros expertos señalan que a partir de una cierta cantidad de niveles de respuesta en la escala ordinal es posible (e incluso se recomienda por cuestiones de computación) el cálculo de la matriz de correlaciones producto-momento, aportando evidencias parciales, poco generalizables o sesgadas que fundamenten esta recomendación (Burga León, 2012; Bandalos & Enders, 1996; Birkett, 1986; Chan, 1991; Cicchetti, Shoinralter, & Tyrer, 1985; Cox, 1980; García Cueto, Muñiz Fernández, & Hernández Baeza, 2000; Freiberg Hoffmann et al., 2013; Gilley & Uhlig, 1993; Holgado-Tello et al., 2008; Lozano, García-Cueto, & Muñiz, 2008; Matell & Jacoby, 1972; Maydeu-Olivares, Kramp, García-Forero, Gallardo-Pujol, & Coffman, 2009; Muthén & Kaplan, 1992; Oliden & Zumbo, 2008; Panter et al., 1997;

Preston & Colman, 2000; Richaud, 2005; Shafel et al., 2012; Weijters, Cabooter, & Schillewaert, 2010; Weng, 2004).

En concreto, entendiendo la estimación tetracórica-policórica como la estimación más ajustada en estos casos, el objetivo de este estudio no es otro que evaluar, mediante un procedimiento sistemático de simulación, cómo afecta a la estimación producto-momento en variables ordinales, en comparación con el coeficiente de correlación policórico-tetracórico, diversas condiciones relacionadas con: el número de categorías de respuesta, la asimetría, el tamaño de la muestra y la intensidad de la relación entre dichas variables. Así, bajo la premisa fundamental de que el coeficiente de correlación tetracórico-policórico realiza una estimación insesgada de la relación entre variables de naturaleza categórica (en este caso ordinal) que provienen originalmente de una variable latente continua, se espera, a partir de los resultados de esta investigación, aportar recomendaciones concretas al investigador de las Ciencias Sociales y de la Salud basadas principalmente en la dirección e intensidad de estas diferencias. Se pretende establecer, por tanto, en base a las evidencias empíricas recabadas, bajo qué condiciones se mantiene la recomendación general del empleo del coeficiente de correlación tetracórico-policórico para estimar la relación entre las variables de una escala, y en qué casos, al no localizar diferencias significativas entre ambas estimaciones, puede resultar más simple, y por ende apropiada, la solución aportada por la estimación producto-momento.

Método

En cuanto al diseño de la investigación, el estudio se implementó a partir de la simulación de datos ordinales. En total, se simularon 90 bases de datos, en función de las diferentes condiciones aplicadas, cada una de las cuales estuvo formada por 10 variables o ítems, considerando en todos los casos que procedían de una escala unidimensional. Las condiciones experimentales impuestas a los

datos fueron el número de categorías de respuesta de los ítems (2, 3, 4, 5 o 7), la forma de la distribución de los ítems (simétrica o asimétrica negativa), el tamaño de muestra (pequeño, mediano o grande) y el nivel de relación entre los ítems (bajo, medio o alto).

Los estudios de simulación de datos son considerados como un procedimiento de investigación formal generalizado y de gran utilidad para la comunidad científica. En concreto, se puede consultar una extensa bibliografía relacionada con el estudio del comportamiento de técnicas y estadísticos multivariantes a partir de datos simulados, alcanzando gran difusión en los últimos años (Cain, Zhang & Yuan, en prensa; Myers, Ahn, Lu, Celimli & Zopluoglu, 2017; Ulitzsch, Schultze & Eid, 2017).

Se plantea como hipótesis principal de la investigación que *la existencia de diversas condiciones relacionadas con el número de categorías de respuesta, la simetría de los datos, el tamaño de la muestra y el nivel de relación entre los ítems ejercerá diferentes efectos significativos sobre el comportamiento del coeficiente de correlación de Pearson, en comparación con el coeficiente de correlación tetracórico-policórico.*

Como añadido, y partiendo de la base de los aportes realizados por las investigaciones previas estudiadas, se plantean las siguientes hipótesis derivadas:

- H1: El valor de la estimación del coeficiente de correlación producto-momento será más bajo que el de la estimación tetracórica-policórica, en términos absolutos, cuando el número de niveles de respuesta en las variables sea bajo. Ambas estimaciones serán similares en los niveles de respuesta altos, y la distancia entre ambas estimaciones aumentará a medida que disminuya el número niveles de respuesta en las variables.
- H2: No se localizará una tendencia clara que permita caracterizar las diferencias en

las estimaciones tetracórica-policórica y producto-momento en función de los distintos niveles de asimetría estudiados en las variables.

- H3: Los 3 niveles de tamaños de muestra estudiados no afectarán de manera diferente a las tendencias de las estimaciones tetracórica-policórica y producto-momento.
- H4: La relación encontrada por el estimador producto-momento será de menor intensidad a la devuelta por el estimador tetracórico-policórico cuando la relación real entre las variables sea alta. En concreto, mientras que en los niveles de relación bajos ambas estimaciones serán similares, a medida que se incremente el nivel de relación entre las variables la distancia entre ambas estimaciones tenderá a aumentar.

Atendiendo a las variables del estudio, se identificó en primer lugar como variable dependiente la diferencia de correlaciones (producto-momento - tetracórica-policórica) entre cada una de las parejas de las 10 variables que componen cada base de datos simulada. Por otro lado, como variables independientes del estudio, se señalan las diferentes condiciones aplicadas en las simulaciones:

- *Número de categorías de respuesta de las variables ordinales:* Se simulan datos con 2, 3, 4, 5 y 7 niveles de respuesta. Se seleccionan estos niveles y no otros al ser los de más habitual empleo en los estudios con escalas de respuesta tipo Likert en Ciencias Sociales. Por otro lado, los estudios previos analizados en este trabajo indican que en estos niveles de respuesta es en los que se marcan las mayores diferencias entre ambos coeficientes de correlación, reduciéndose las diferencias a partir de los 7 niveles de respuesta a la mínima expresión.
- *Niveles de simetría de las variables:* Dadas las cuestiones señaladas en el estudio del estado de la cuestión al respecto de la normalidad y simetría, se simulan datos

con distribuciones normales (simetría) y datos con asimetría negativa acentuada.

- *Tamaño de la muestra*: Dada la variedad de tamaños de muestra existente en los estudios desarrollados en el ámbito de las Ciencias Sociales y Ciencias de la Salud, cabe estudiar los efectos que ejercen 3 tamaños de muestra diferentes: muestras pequeñas ($n=30$) adoptando el límite establecido habitualmente por la estadística tradicional, muestras de tamaño medio ($n=200$) y muestras grandes ($n=1000$).
- *Nivel de relación entre las variables*: Se tiene en cuenta que la relación entre variables en los estudios de encuesta suele alcanzar valores moderados, por lo que se entiende el nivel de relación baja como $r \approx .2$, el nivel de relación media como $r \approx .5$ y el nivel de relación alto como de $r \approx .7$ (valores más altos son difíciles de alcanzar en Ciencias Sociales y pueden ir asociados a problemas de colinealidad y multicolinealidad). Cabe destacar en este aspecto que los niveles de relación entre los 10 ítems que integra cada una de las situaciones experimentales no es exactamente el valor propuesto, sino que el paquete empleado para simular datos incluye como criterio que la correlación entre cada pareja de ítems sea cercana al valor indicado, con un nivel de error bajo. Así, los errores en las estimaciones de la correlación producto-momento (indicador empleado en la simulación de datos para indicar la relación entre las variables) están en todos los casos entre aproximadamente ± 0.1 puntos alrededor del valor de la relación forzado en la simulación.

De entre las alternativas existentes para llevar a cabo la simulación de los datos, se optó por el empleo del software R, en concreto del paquete GenOrd¹. Este paquete proporciona la flexibilidad, fiabilidad y

herramientas necesarias para implementar las bases de datos simuladas bajo las condiciones señaladas más arriba.

Finalmente, en lo que respecta al análisis de los datos, fue necesario el empleo de varios paquetes estadísticos. Las matrices de correlaciones y de diferencias de correlaciones fueron calculadas generando una breve aplicación en R (para la estimación de la correlación tetracórica-policórica se empleó el paquete Polycor²). A partir de esa información, se generó una base de datos con información de las 4050 diferencias de correlaciones (45 indicadores por cada una de las 90 situaciones experimentales) y de las diferentes condiciones aplicadas en cada situación.

Para el análisis de los efectos de las diferentes condiciones experimentales sobre las diferencias de correlaciones se aplicó la técnica del Análisis de Varianza (ANOVA), considerando cada condición experimental como un factor del modelo, bajo un nivel de significación del 5%. Como medida complementaria a los contrastes de hipótesis implementados se calcularon los tamaños del efecto tanto de los factores del modelo ANOVA (estadístico η^2), como de los contrastes de las pruebas post-hoc (estadístico d de Cohen). Por último, también se generaron gráficos de las diferencias de correlaciones. Todos estos cálculos se realizaron a partir del empleo del software SPSS V.23 y Microsoft Excel 2010 (ambos bajo licencia campus USAL).

Resultados

En primer lugar, dado que se trataba de una variable que no afectaba directamente a la configuración y características de las variables ordinales simuladas, se tomó la decisión de emplear los 3 niveles de tamaño de muestra (pequeño, medio y alto) como agrupaciones para la segmentación de la base

¹ Simulation of Discrete Random Variables with Given Correlation Matrix and Marginal Distributions. La documentación relativa a este paquete puede ser consultada en: <https://cran.r-project.org/web/packages/GenOrd/GenOrd.pdf>

² Polychoric and Polyserial Correlations, 25/03/2010, John Fox. Documentación: <https://cran.r-project.org/web/packages/polycor/polycor.pdf>

de datos. Como consecuencia de esta decisión, se generaron 3 modelos de ANOVA independientes en los que se integraron como factores fijos el número de categorías de respuesta, los niveles de asimetría y el nivel de relación entre las variables. Los modelos finales integraron igualmente los efectos de interacción de todas las combinaciones entre estos factores. Igualmente, en base a esta cuestión, se presentan los resultados descriptivos para las bases de datos simuladas de tamaños de muestra pequeño, mediano y grande. Así, la tabla 1 muestra la diferencia de

correlaciones en los distintos niveles de respuesta simulados. Se observa cómo en todos los casos la correlación producto-momento estima relaciones más bajas entre las variables con respecto a la correlación tetracórica-policórica. Esta tendencia se va moderando a medida que el número de opciones de respuesta es más elevado. Los diferentes tamaños de muestra no parecen ejercer un efecto claro sobre las diferencias de correlaciones, aunque sí que se observa cómo la variabilidad se reduce claramente a medida que el tamaño de muestra aumenta.

Tabla 1. Diferencia de correlaciones media en función de niveles de respuesta.

	O2		O3		O4		O5		O7	
	\bar{X}	S_x	\bar{X}	S_x	\bar{X}	S_x	\bar{X}	S_x	\bar{X}	S_x
n=30	-0.165	0.062	-0.122	0.059	-0.091	0.051	-0.064	0.063	-0.044	0.056
n=200	-0.182	0.047	-0.119	0.042	-0.084	0.035	-0.054	0.028	-0.046	0.025
n=1000	-0.179	0.044	-0.120	0.041	-0.084	0.034	-0.059	0.023	-0.042	0.017

En lo que respecta al efecto de los diferentes niveles de relación entre las variables sobre la diferencia de correlaciones media (tabla 2), se observa cómo ésta aumenta a medida que el nivel de relación entre los ítems es más importante. Igualmente, se vuelve a localizar en todos los casos que la correlación

producto-momento estima relaciones de menor intensidad que la correlación tetracórica-policórica, y que los diferentes tamaños de muestra no alcanzan una tendencia clara en cuanto a sus efectos medios.

Tabla 2. Diferencia de correlaciones media en función de niveles de relación.

	$r_{xy} \approx .2$		$r_{xy} \approx .5$		$r_{xy} \approx .7$	
	\bar{X}	S_x	\bar{X}	S_x	\bar{X}	S_x
n=30	-0.047	0.070	-0.119	0.060	-0.126	0.058
n=200	-0.060	0.046	-0.112	0.067	-0.119	0.051
n=1000	-0.057	0.038	-0.113	0.063	-0.121	0.051

Finalmente, se muestra el efecto asociado al tipo de distribución (simétrica-asimétrica) de los datos ordinales. En este caso la tendencia no es tan clara, aunque sí se observa que en muestras pequeñas y grandes las distribuciones asimétricas tienen una ligera tendencia a aumentar la diferencia de correlaciones. Cabe recordar que la normalidad de las variables es un supuesto

previo para el cálculo del coeficiente de correlación tetracórico-policórico, pero que la violación de este supuesto no posee efectos importantes sobre el sesgo cometido en el cálculo de este coeficiente (Freiberg Hoffmann et al., 2013; Holgado-Tello et al., 2008; Morata-Ramírez & Holgado-Tello, 2013).

Tabla 3. Diferencia de correlaciones en función de niveles de relación.

	Distribución simétrica		Distribución asimétrica	
	\bar{X}	S_x	\bar{X}	S_x
n=30	-0.099	0.073	-0.095	0.071
n=200	-0.091	0.064	-0.103	0.059
n=1000	-0.093	0.061	-0.101	0.057

En cuanto a los análisis inferenciales desarrollados, la tabla 4 muestra la información global sobre el ajuste de los 3 modelos con y sin los efectos de las interacciones. En todos los casos resultan modelos altamente significativos con unos porcentajes de varianza explicada (R^2 , coeficiente de determinación) muy importantes, que van creciendo a medida que

los tamaños de las muestras simuladas son más grandes. Estos resultados aportan un indicador de validez interna y consistencia de la perspectiva empleada en el estudio, confirmando que los factores seleccionados son esenciales en la configuración de las diferentes estimaciones de los indicadores producto-momento y tetracórico-policórico.

Tabla 4. Significación y varianza explicada modelos ANOVA efectos fijos y completos.

	F	p.	R²
Muestra pequeña (n=30). Efectos fijos	276.11	<.001	.588
Muestra pequeña (n=30). Modelo completo	88.10	<.001	.652
Muestra media (n=200). Efectos fijos	1092.72	<.001	.850
Muestra media (n=200). Modelo completo	600.70	<.001	.928
Muestra grande (n=1000). Efectos fijos	2148.91	<.001	.918
Muestra grande (n=1000). Modelo completo	3320.58	<.001	.986

El coeficiente de determinación resulta sensiblemente superior en los modelos completos, por lo que se plantea la posibilidad de interacciones significativas entre los factores. Así, se considera apropiado incorporar estas interacciones en los 3 modelos finales. El primer modelo, correspondiente a las bases de datos simuladas con un tamaño de muestra de 30

sujetos, se muestra en la tabla 5. En este caso, el factor simetría resulta no significativo, con un tamaño del efecto (estadístico eta cuadrado) muy pequeño. Por otro lado, la mayor parte de las interacciones resultan significativas, con tamaños del efecto bajos y moderados en las interacciones significativas y altos en los factores fijos significativos (Hopkins, 2000).

Tabla 5. ANOVA para la diferencia de correlaciones. Tamaño muestra pequeño (n=30)

	S.C.	G.L.	M.C.	F	p.	η^2
Intersección	12.753	1	12.753	6969.033	<.001	.841
OPCIONES	2.464	4	0.616	336.650	<.001	.505
SIMETRÍA	0.005	1	0.005	2.743	.098	.002
RELACIÓN	1.715	2	0.858	468.663	<.001	.415
OPC * SIM	0.059	4	0.015	8.107	<.001	.024
OPC * REL	0.303	8	0.038	20.712	<.001	.112
SIM * REL	0.003	2	0.001	0.718	<.001	.001
OPC * SIM * REL	0.125	8	0.016	8.522	<.001	.049
Error	2.416	1320	0.002			
Total	19.843	1350				

Dado que se obtienen interacciones altamente significativas en el modelo, la interpretación de los factores principales podría resultar problemática. Así, el gráfico 1 muestra las medias marginales de las diferencias de correlaciones del número de opciones de respuesta en función de los factores simetría y relación. A pesar de observarse algunas interacciones en los diagramas, se evidencia

que existe una tendencia clara al acercamiento entre las estimaciones de ambos coeficientes de correlación a medida que el número de opciones de respuesta aumenta y la relación entre las variables es menos intensa. El acercamiento entre ambas estimaciones de los coeficientes de correlación es importante para 2, 3, 4 y 5 niveles de respuesta, y la tendencia se va estabilizando entre 5 y 7 niveles.

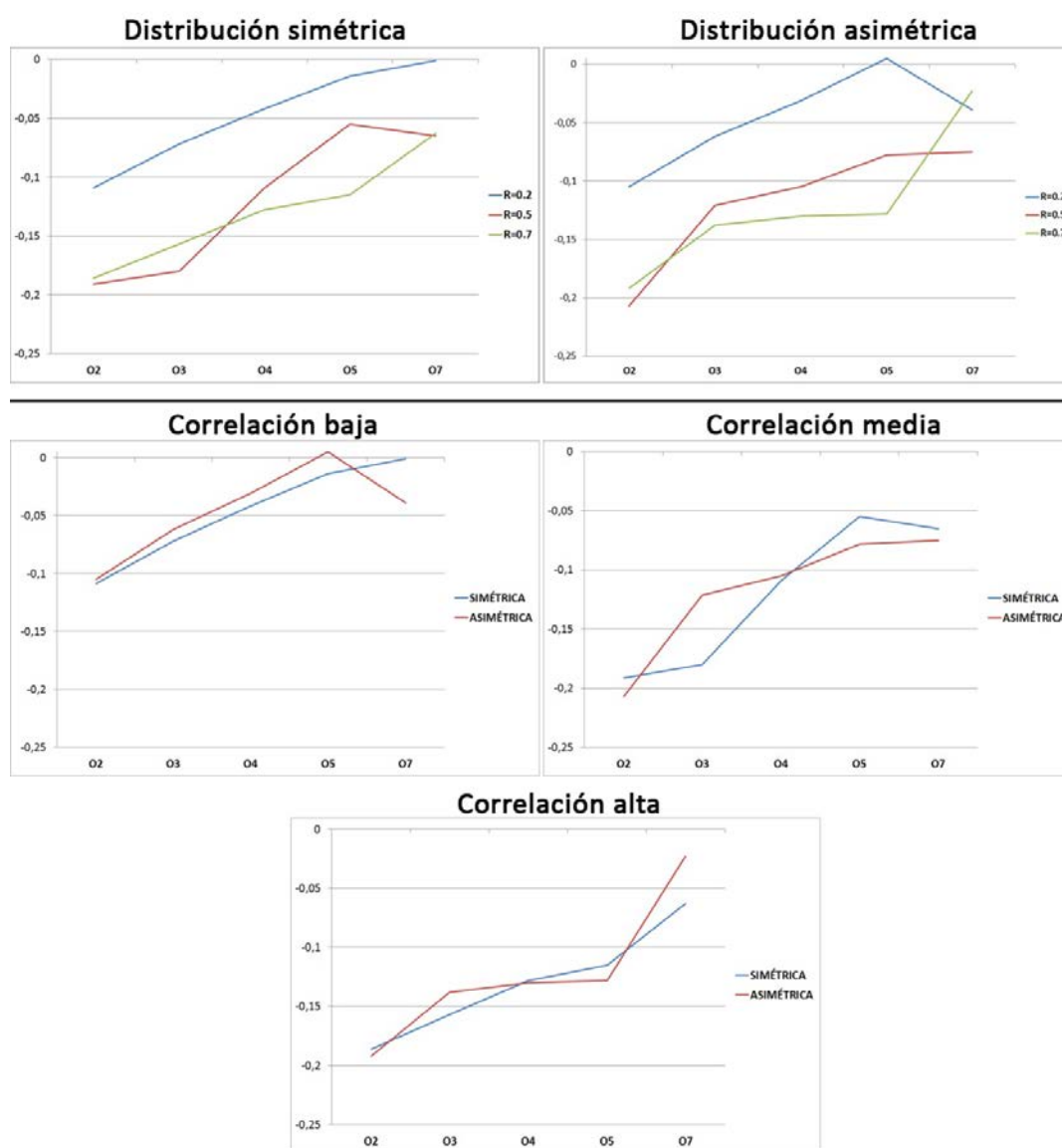


Gráfico 1. Gráfico de medias para factores con mayor interacción (n=30)

El modelo aplicado para las muestras de 200 sujetos (tabla 6) muestra efectos significativos en todos los factores e interacciones, con tamaños del efecto que van de moderados (simetría) a muy altos (opciones y relación).

Parece que para este tamaño de muestra las estimaciones de los coeficientes de correlación se estabilizan y los errores se minimizan. Sigue existiendo un efecto importante de interacción.

Tabla 6. ANOVA para la diferencia de correlaciones. Tamaño muestra medio (n=200)

	S.C.	G.L.	M.C.	F	p.	η^2
Intersección	12.699	1	12.699	46911.587	<.001	.973
OPCIONES	3.319	4	0.830	3065.352	<.001	.903
SIMETRÍA	0.044	1	0.044	161.860	<.001	.109
RELACIÓN	0.953	2	0.476	1759.877	<.001	.727
OPC * SIM	0.072	4	0.018	66.106	<.001	.167
OPC * REL	0.281	8	0.035	129.688	<.001	.440
SIM * REL	0.022	2	0.011	40.969	<.001	.058
OPC * SIM * REL	0.025	8	0.003	11.663	<.001	.066
Error	0.357	1320	<0.001			
Total	17.772	1350				

El gráfico 2 muestra como esta interacción es menos acusada que en el caso de los

tamaños de muestra pequeños, y las tendencias son más estables.

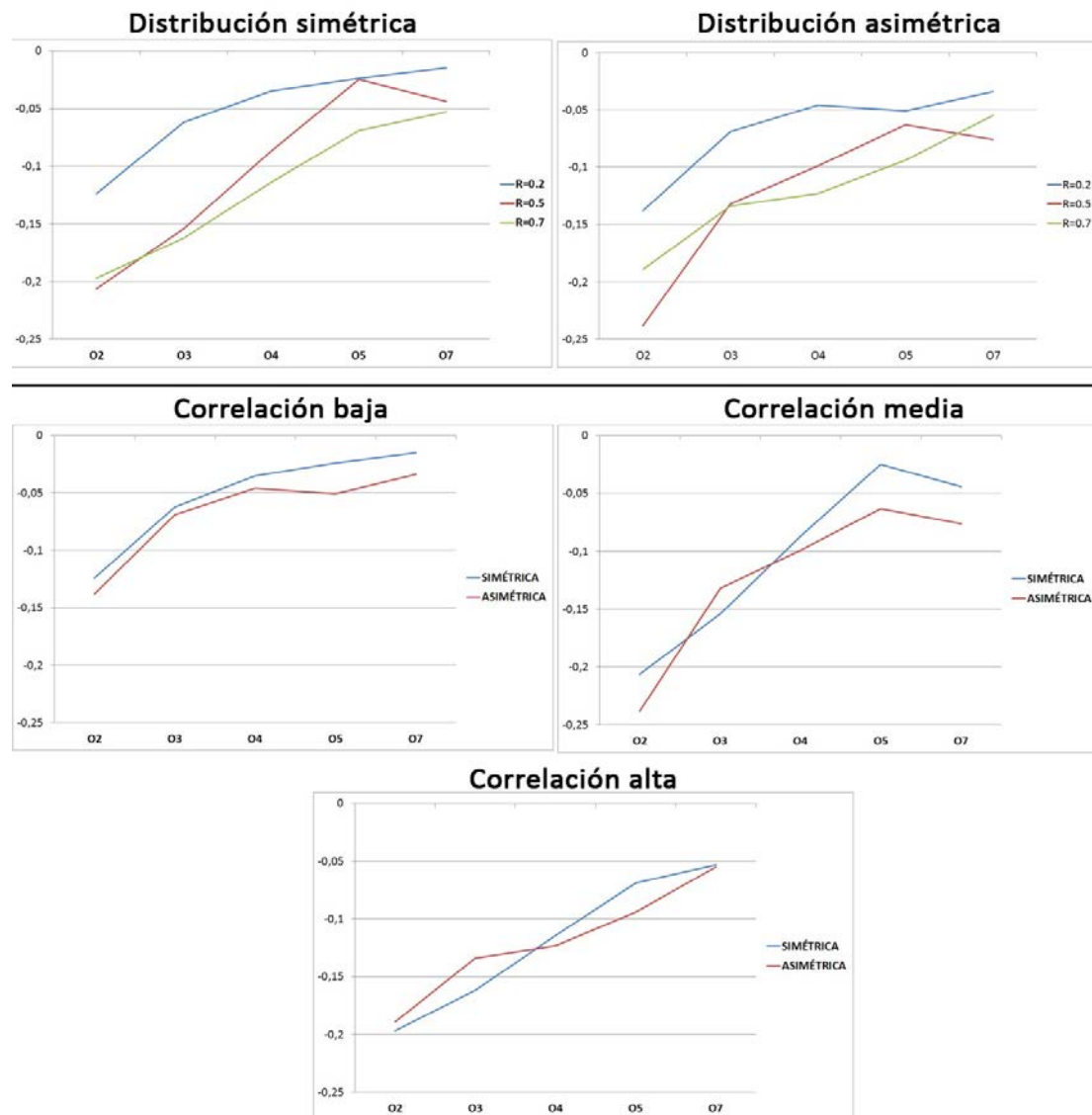


Gráfico 2. Gráfico de medias para factores con mayor interacción (n=200)

El último modelo, que incorpora los datos simulados de las muestras de tamaño grande ($n=1000$), se muestra en la tabla 7. Finalmente, se confirman las tendencias anteriormente observadas. Mientras que el número de opciones de respuesta, seguido de cerca por el nivel de relación entre las variables, es el factor con unos efectos más importantes, la simetría-asimetría de las

distribuciones marginales de las variables simuladas alcanza efectos significativos con tamaños del efecto mucho más limitados. Por su parte, las interacciones siguen resultando significativas, y es especialmente importante el efecto de la interacción entre el número de opciones de respuesta y la relación entre las variables.

Tabla 7. ANOVA para la diferencia de correlaciones. Tamaño muestra grande ($n=1000$)

	S.C.	G.L.	M.C.	F	p.	η^2
Intersección	12.655	1	12.655	262510.429	<.001	.995
OPCIONES	3.207	4	0.802	16631.085	<.001	.981
SIMETRÍA	0.023	1	0.023	469.461	<.001	.262
RELACIÓN	1.091	2	0.545	11313.772	<.001	.945
OPC * SIM	0.066	4	0.017	343.593	<.001	.510
OPC * REL	0.217	8	0.027	562.652	<.001	.773
SIM * REL	0.024	2	0.012	250.141	<.001	.275
OPC * SIM * REL	0.014	8	0.002	37.457	<.001	.185
Error	0.064	1320	<0.001			
Total	17.361	1350				

Dado que se mantienen los citados efectos de interacción significativos, se incluye el gráfico 3, que pormenoriza estas cuestiones en el caso de las muestras grandes. La tendencia continua estabilizándose, mostrando interacciones más ligeras. Esta cuestión posibilita la interpretación directa de los factores principales, a excepción de la simetría, cuya tendencia, aunque de manera más moderada, sigue mostrando síntomas de interacción importantes con el número de opciones de respuesta.

En todo caso, estos gráficos y técnicas de ANOVA de varios factores aplicados confirman que el coeficiente de correlación producto-momento estima que, en términos absolutos, la relación entre las variables ordinales es más baja (en relación con la estimación tetracórica-policórica) cuando éstas poseen pocas opciones de respuesta en

la escala y cuando la relación entre las variables es más intensa. Por otro lado, parece que cuando el número de opciones de respuesta es 5 o más, las diferencias entre ambos coeficientes de correlación se minimizan, llegando a acercarse a 0. Lo contrario ocurre con el nivel de relación entre las variables. Parece que, mientras que existen muy pocas diferencias entre las estimaciones de ambas correlaciones cuando la relación entre las variables es baja ($r=.2$), este valor aumenta mucho cuando la relación llega hasta .5, estancándose en este nivel y manteniéndose similar cuando la relación es de .7. De hecho, en las pruebas post-hoc, mostradas a continuación, la única pareja en la que las diferencias no resultan altamente significativas es la conformada por las bases de datos simuladas de tamaños de muestra pequeños con relaciones entre las variables medias y altas.

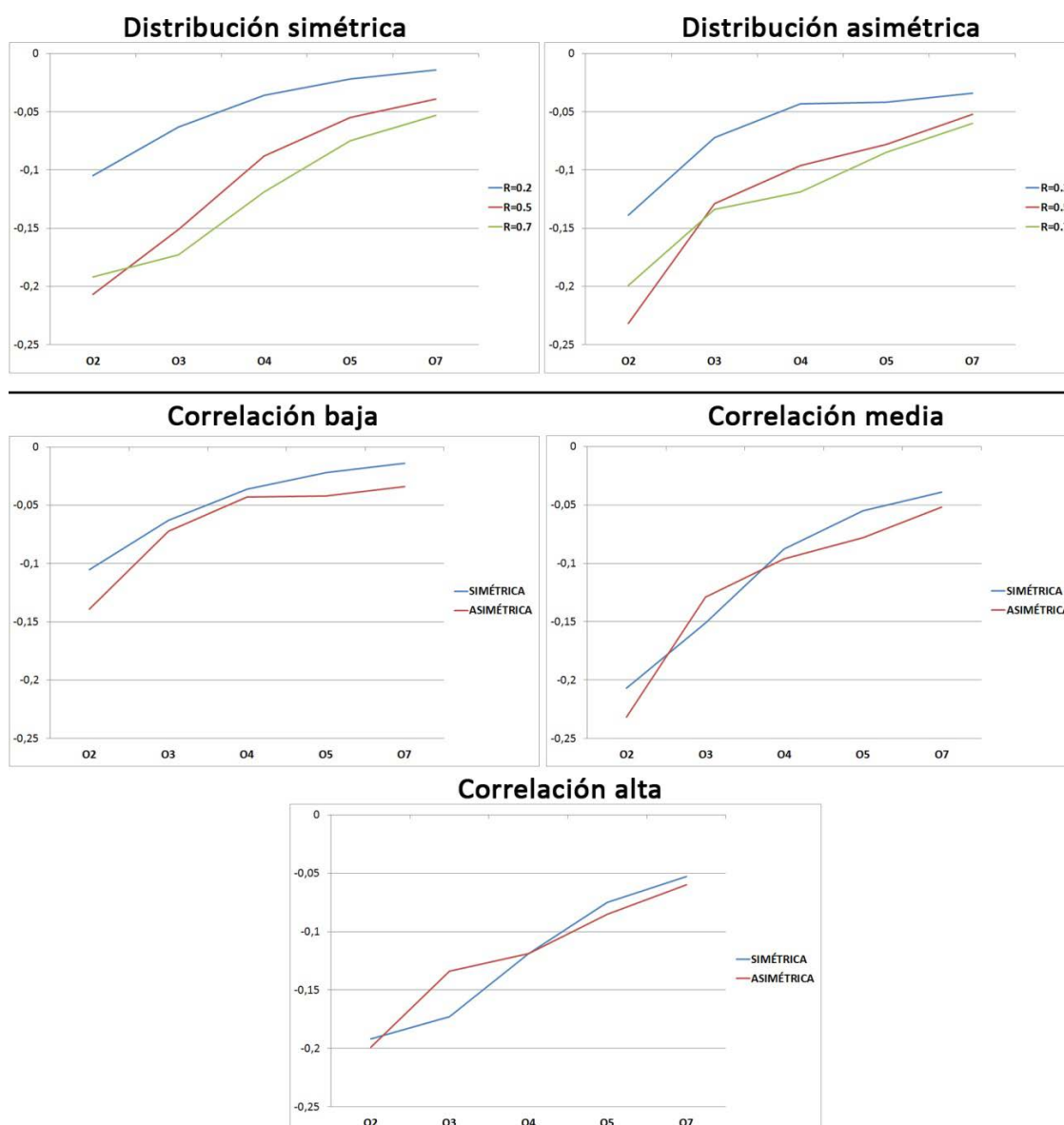


Gráfico 3. Gráfico de medias para factores con mayor interacción (n=1000)

En lo que respecta a las pruebas post-hoc, como ya se ha anticipado más arriba, todas las parejas contrastadas resultan significativas. Para un análisis más detallado de estas significaciones, se ha procedido al cálculo del estadístico del tamaño del efecto d de Cohen (Cohen, 1969). Así, la tabla 8 muestra los tamaños del efecto para las pruebas post-hoc entre todas las parejas posibles del factor número de categorías de respuesta de la variable ordinal. Se confirma, como ya se había anticipado en el análisis gráfico anterior, que los tamaños del efecto se hacen más importantes cuando se comparan parejas

distantes en cuanto al número de opciones de respuesta, y que, cuando se comparan parejas consecutivas con un número de opciones de respuesta más elevado, el tamaño del efecto se reduce. Por ejemplo, mientras que cuando se comparan las muestras con 2 y 3 opciones de respuesta se obtienen unos tamaños del efecto superiores a 1 en los 3 niveles de tamaño de muestra, cuando se comparan las muestras con 5 y 7 opciones de respuesta 2 de estos valores no superan alcanzan un valor del tamaño del efecto superior a .5, inferior a los valores que podríamos interpretar como de efectos moderados (Hopkins, 2000).

Tabla 8. Tamaños del efecto pruebas post-hoc (d de Cohen). Opciones de respuesta

		O2	O3	O4	O5	O7
O2	n=30	-	1.01	1.74	2.36	2.82
	n=200	-	3.84	5.96	7.76	8.26
	n=1000	-	8.43	13.72	17.23	19.71
O3	n=30		-	0.73	1.34	1.81
	n=200		-	2.12	3.93	4.43
	n=1000		-	5.29	8.80	11.29
O4	n=30			-	0.62	1.09
	n=200			-	1.81	2.30
	n=1000			-	3.51	5.99
O5	n=30				-	0.47
	n=200				-	0.50
	n=1000				-	2.49

Igualmente, la tabla 9 muestra los tamaños del efecto de las pruebas post-hoc en el caso del factor relación entre las variables. Nuevamente, se observa cómo los tamaños del efecto aumentan sistemáticamente a medida que se comparan niveles de relación

más alejados y cuando se compara la relación baja con la moderada. De hecho, cuando se comparan los niveles de relación media y alta, para 2 de los 3 tamaños de muestra los efectos resultan bajos.

Tabla 9. Tamaños del efecto pruebas post-hoc (d de Cohen). Relación entre las variables

		$r_{xy} \approx .2$	$r_{xy} \approx .5$	$r_{xy} \approx .7$
$r_{xy} \approx .2$	n=30	-	1.67	1.85
	n=200	-	3.21	3.61
	n=1000	-	8.03	9.22
$r_{xy} \approx .5$	n=30		-	0.18
	n=200		-	0.40
	n=1000		-	1.19

A nivel general, por tanto, las evidencias muestran las siguientes cuestiones:

- Los niveles de simetría-asimetría no parecen ejercer efectos grandes sobre las diferencias entre la estimación tetracórica-policórica y la estimación producto-momento, y las diferencias localizadas no tienen una tendencia clara. Debe destacarse, además, de cara a la posterior propuesta que el sesgo asociado al factor de simetría-asimetría puede deberse tanto a efectos de interacción entre este y el resto de factores en la estimación del coeficiente de correlación producto-momento, como a la violación del supuesto de normalidad en la estimación tetracórica-policórica.
- Mientras que el número de niveles de respuesta de la variable ordinal parece alcanzar los efectos más altos, no se localizan tamaños del efecto grandes en las pruebas post-hoc cuando se compara la situación de 5 niveles de respuesta con 7 niveles.
- El nivel de relación de partida entre las variables parece alcanzar igualmente efectos altamente significativos, con tamaños del efecto altos (valor del eta cuadrado), en los modelos planteados, aunque los tamaños del efecto de las pruebas post-hoc indican que no existen grandes diferencias entre los niveles de relación medios y altos.

Una vez desarrollado este conjunto de análisis, estamos en condiciones de llevar a cabo una propuesta razonada y bien delimitada en torno a los escenarios en los que, dada la importante infravaloración que el coeficiente de correlación producto-momento realiza de la verdadera relación entre las variables (entendiendo el coeficiente tetracórico-policórico como el estimador insesgado), puede ser recomendable el empleo de la estimación tetracórica-policórica. Por el contrario, se determinan algunos otros casos en los que, dado que no existen grandes diferencias entre ambos coeficientes de correlación (fundamentadas en tamaños del efecto moderados o bajos), y en base al principio de parsimonia que debe estar presente en todo análisis estadístico, se

considera más recomendable el empleo del coeficiente producto-momento. Al respecto, no olvidemos que en este caso se puede asumir la perspectiva de algunos autores (Morales Vallejo, 2000; Nunnally, 2010), que destacan que si el empleo de las técnicas clásicas suponen un sesgo mínimo, dada la pérdida de prestaciones asociada a la aplicación del estimador tetracórico-policórico, no se justifica su uso. En este contexto, la tabla 10 resume la propuesta realizada. Dado el hecho de que se considera necesario recabar información más profunda al respecto a partir de diversas fuentes de información para llevar a cabo una propuesta más fundamentada y concreta, se lleva a cabo una recomendación general conservadora.

Tabla 10. Coeficiente de correlación recomendado en función de la matriz de datos

		n=30	n=200	n=1000
O2	$r_{xy} \approx .2$	N*	T-P**	T-P
	$r_{xy} \approx .5$	N	T-P	T-P
	$r_{xy} \approx .7$	N	T-P	T-P
O3	$r_{xy} \approx .2$	N	T-P	T-P
	$r_{xy} \approx .5$	N	T-P	T-P
	$r_{xy} \approx .7$	N	T-P	T-P
O4	$r_{xy} \approx .2$	N	P-M	P-M
	$r_{xy} \approx .5$	N	T-P	T-P
	$r_{xy} \approx .7$	N	T-P	T-P
O5	$r_{xy} \approx .2$	P-M***	P-M	P-M
	$r_{xy} \approx .5$	N	T-P	T-P
	$r_{xy} \approx .7$	N	T-P	T-P
O7	$r_{xy} \approx .2$	P-M	P-M	P-M
	$r_{xy} \approx .5$	N	P-M	P-M
	$r_{xy} \approx .7$	N	P-M	P-M

* N=No se recomienda el empleo de ningún tipo de técnica multivariante

** T-P=Coeficiente de correlación tetracórico-policórico

*** P-M=Coeficiente de correlación producto-momento

Los criterios seguidos para el establecimiento de la propuesta han sido los siguientes:

1. Dado que los análisis han mostrado que la existencia o falta de simetría no ejerce un efecto diferenciador en ambas estimaciones, no se contempla esta variable en la propuesta.

2. En los casos en los que se dispone de una muestra de pequeño tamaño ($n=30$), debemos tener en cuenta que no se recomienda en ningún caso la estimación tetracórica-policórica con tamaños de muestra inferiores a 150 sujetos (Freiberg Hoffmann et al., 2013), debido a la alta inestabilidad que esta estimación puede conllevar. Por otro lado, en cuanto al

empleo de la estimación producto-momento, dadas las altas diferencias localizadas entre ambas estimaciones en los casos de menos de 5 niveles de respuesta, y en los casos de relaciones medias y altas entre las variables, no se recomienda su empleo en estos casos. Así, sólo se recomienda el empleo de este estimador con muestras pequeñas en los casos en los que la relación entre las variables sea baja y los niveles de respuesta sean no inferiores a 5.

3. En los casos en los que los tamaños de muestra sean de medios a grandes, se establece una recomendación exactamente igual, basándonos en los indicadores obtenidos en los análisis previos. Dado que las diferencias en las estimaciones producto-momento y tetracórica-policórica, como muestran los gráficos y los modelos aplicados, son muy grandes cuando las variables originales tienen menos de 4 niveles de respuesta, se entiende que el coeficiente de correlación tetracórico-policórico, debe ser empleado en todos los casos cuando se da esta situación. Sin embargo, dado que el análisis gráfico muestra que la diferencia de correlaciones es mínima cuando la relación entre las variables es baja y disponemos de 4 o 5 niveles de respuesta en las mismas, se propone como plausible el empleo de la correlación producto-momento en estos casos.

4. Finalmente, dado que las evidencias muestran que la diferencia de correlaciones es mínima en todos los casos en los que se dispone de 7 niveles de respuesta en las variables (se alcanzan diferencias de correlaciones cercanas o inferiores a .05 puntos), se propone como plausible el empleo de la correlación producto-momento cuando nos encontramos con escalas de este tipo (o con más niveles de respuesta).

Discusión

El manejo de instrumentos para la medición de rasgos, características, actitudes, etc., es

una práctica habitual en el ámbito de las Ciencias Sociales y de la Salud (Abad, 2011; Morales Vallejo, 2000), integrándose de manera generalizada escalas de medida de naturaleza no cuantitativa como, principalmente, las escalas de respuesta tipo Likert (García Cueto et al., 2000; Pearse, 2011; Preston & Colman, 2000; Shafel et al., 2012). Así, el empleo generalizado de técnicas estadísticas propias de escalas de intervalo o razón en estas variables no métricas está en entredicho, al menos desde un punto de vista matemático (Marcus-Roberts & Roberts, 1987; Stevens, 1946). No obstante, mientras que es difícil localizar criterios claros y compartidos sobre las técnicas multivariantes más apropiadas en función de las escalas e ítems disponibles, ni siquiera existe un consenso en la comunidad científica sobre la conveniencia de la sustitución de las técnicas clásicas.

Se evidencia que, en lo que se refiere al estudio de las propiedades de la estimación de la correlación tetracórica-policórica en relación a la estimación de la correlación producto-momento, el estado de la cuestión no tiene en cuenta, generalmente, las diferentes características que puede poseer el conjunto de variables ordinales que conforman la escala (Lozano et al., 2008; Maydeu-Olivares et al., 2009; Oliden & Zumbo, 2008). Por otro lado, la mayor parte de los estudios analizan directamente los efectos de las diferentes estimaciones de estos dos coeficientes sobre los resultados de los análisis factoriales en las escalas (Burga León, 2012; Freiberg Hoffmann et al., 2013; Gilley & Uhlig, 1993; Holgado-Tello et al., 2008; Muthén & Kaplan, 1992; Panter et al., 1997; Richaud, 2005). Así, no se localizan trabajos que analicen directamente las diferencias entre estos dos estimadores, contemplando los factores principales que pueden sesgar las estimaciones.

En este contexto, el trabajo presentado trata de aportar información útil para el investigador aplicado en relación al sesgo cometido en la estimación del coeficiente de

correlación producto-momento, en función de diversas características clave de las variables ordinales de las que proceden las escalas de respuesta tipo Likert, entendiendo que el estimador tetracórico-policórico ofrece un índice insesgado en los casos en los que se puede entender que subyace a la escala ordinal observada una escala continua latente. Así, se puede afirmar que se han satisfecho los objetivos e hipótesis planteados inicialmente: por un lado, se ha estudiado y aplicado un procedimiento para la simulación de datos ordinales controlando multitud de factores de caracterización de las variables ordinales que, en base al estado de la cuestión, afectan a la estimación producto-momento; por otro, se han aplicado un conjunto de técnicas de análisis de datos simples y claras en el nivel básico de la diferencia de correlaciones, lo cual ha permitido una interpretación amplia de los resultados; y finalmente, se ha llevado a cabo una propuesta concreta y operativa en la que, partiendo de la estimación tetracórica-policórica como la más apropiada de partida para calcular la relación entre variables provenientes de escalas de respuesta tipo Likert, se indican los casos en los que las diferencias de correlaciones entre los dos estimadores se consideran tan insignificantes que, en base al principio de parsimonia y la perspectiva de algunos autores de referencia (Morales Vallejo, 2000; Nunnally, 2010), se sugiere el empleo de la correlación producto-momento.

Destaca en cuanto a las evidencias obtenidas la robustez de los modelos de análisis de varianza elaborados, que han alcanzado niveles de bondad de ajuste superiores al 90%, e incluso muy cercanos al 100%, cuando los tamaños de muestra han sido suficientemente grandes. Los resultados indican que, mientras que el número de categorías de respuesta de las variables y el nivel de relación entre las mismas son los factores más importantes, los niveles de asimetría tienen una influencia limitada y poco definida. Por otro lado, el tamaño de la muestra implica únicamente diferentes niveles

de variabilidad en las estimaciones. Cabe añadir al respecto que, a pesar de que estos niveles de bondad de ajuste son poco frecuentes, incluso con muestras grandes, pueden estar influenciados en buena medida por los algoritmos incorporados en el paquete estadístico encargado de implementar el procedimiento de simulación de los datos, cuestión que podría ser considerada objeto de análisis en futuros estudios.

Más en concreto, las evidencias confirman que en las variables con más de 5 niveles de respuesta, a excepción de que la relación entre los ítems sea muy alta (hecho poco habitual en este tipo de escalas), el empleo del coeficiente de correlación producto-momento no infravalora de manera reseñable la relación entre las variables en relación con la estimación tetracórica-policórica. Estos resultados concuerdan con los obtenidos en los estudios previos analizados, que identifican un punto crítico en variables a partir de 5 niveles de respuesta (Choi et al., 2010; García Cueto et al., 2000; Holgado-Tello et al., 2008; Lozano et al., 2008; Oliden & Zumbo, 2008; Preston & Colman, 2000; Weijters et al., 2010; Weng, 2004). Por otro lado, parece de obligado empleo el coeficiente de correlación tetracórico-policórico en el caso de variables con 5 o menos niveles de respuesta, a no ser que la relación entre las variables sea muy baja. Cabe recordar al respecto que, a medida que la relación entre las variables se amplía, también lo hacen las diferencias entre las dos estimaciones de las correlaciones. Así, en el caso de las variables con 5 opciones de respuesta, se puede señalar que si la relación es de baja a media-baja, se puede considerar como apropiado el empleo del coeficiente de correlación producto-momento, cuestión que en el resto de circunstancias puede resultar comprometida.

En lo que respecta a los niveles de asimetría-simetría de las variables, cabe destacar que, si bien las distribuciones asimétricas suponen un ligero aumento de las diferencias entre ambos coeficientes de correlación, no queda claro si este aumento es debido a la infraestimación

del coeficiente de correlación producto-momento sobre este tipo de variables o al pequeño sesgo producido en la estimación del coeficiente de correlación tetracórico-policórico cuando no existe normalidad univariante-multivariante (Freiberg Hoffmann et al., 2013; Holgado-Tello et al., 2008; Jöreskog, 1994; Morata-Ramírez & Holgado-Tello, 2013; Olsson, 1979). Es por esto, que la propuesta integrada en este trabajo no contempla una diferenciación entre los niveles de simetría-asimetría de las variables de referencia.

Para finalizar, debemos destacar en este punto las fortalezas y puntos débiles del presente estudio. Como aspectos positivos que aportan solidez al trabajo cabe destacar, como ya se ha señalado, tanto la multitud de factores tenidos en cuenta a la hora de llevar a cabo la simulación como la sencillez en el análisis estadístico de las diferencias. Esto ha facilitado un alto grado de profundidad y simplificación en la interpretación de los resultados, aumentando la validez interna del proceso. No obstante, la selección de múltiples factores y no haber incluido técnicas más complejas y globales, como por ejemplo un contraste de hipótesis para comparar directamente las parejas de matrices de correlaciones, también puede ser considerado como una debilidad del trabajo. En este sentido, la no inclusión de técnicas más globales y la amplitud de los factores analizados son cuestiones que pueden estar atentando, bajo nuestro punto de vista, sobre la propia validez externa de los resultados obtenidos. De hecho, cabe remarcar que la propuesta realizada, a pesar de ser básicamente coincidente con el estado de la cuestión previo y fundamentarse en las evidencias recogidas, se puede considerar parcial y en cierto modo insuficiente, ya que está basada únicamente en la diferencia de correlaciones obtenida entre ambos estimadores, sin tener en cuenta otras cuestiones estadístico-teóricas importantes. Finalmente, el hecho de que este estudio se quede en el nivel de simulación de datos supone una dificultad añadida sobre la validez

externa de los resultados, ya que implica obviar algunos sesgos y casuísticas propias de la suministración de escalas de medida a sujetos reales.

Así, dados los puntos débiles analizados, se abren nuevas posibilidades para estudios futuros, como la replicación de estudios similares a partir de muestras reales de sujetos, saliendo de la situación de laboratorio aquí planteada. Igualmente, suscita un alto interés la profundización en el estudio del sesgo y de las técnicas estadísticas más apropiadas en los casos en los que las variables poseen altos grados de asimetría. Por último, no se debe olvidar que este trabajo se ha realizado bajo el supuesto de disponer de escalas unidimensionales, por lo que cabe, asimismo, replicar el estudio con la simulación o aplicación de escalas de naturaleza multidimensional.







Referencias

- Abad, F. J. (2011). *Medición en ciencias sociales y de la salud*. Madrid: Síntesis.
- Bandalos, D. L., & Enders, C. K. (1996). The Effects of Nonnormality and Number of Response Categories on Reliability. *Applied Measurement in Education*, 9(2), 151-160. doi: https://doi.org/10.1207/s15324818ame0902_4
- Birkett, N. J. (1986). Selecting the number of response categories for a Likert-type scale. En *Proceedings of the American Statistical Association* (pp. 488-492). USA.
- Burga León, A. (2012). La unidimensionalidad de un instrumento de medición: perspectiva factorial. *Revista de Psicología*, 24(1), 53-80.
- Cain, M.K., Zhang, Z. & Yuan, K.H. (en prensa). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods*. doi: <https://doi.org/10.3758/s13428-016-0814-1>
- Chan, J. C. (1991). Response-Order Effects in Likert-Type Scales. *Educational and*

- Psychological Measurement*, 51(3), 531-540. doi: <https://doi.org/10.1177/0013164491513002>
- Choi, J., Peters, M., & Mueller, R. O. (2010). Correlational analysis of ordinal data: from Pearson's r to Bayesian polychoric correlation. *Asia Pacific education review*, 11(4), 459-466. doi: <https://doi.org/10.1007/s12564-010-9096-y>
- Cicchetti, D. V., Shoinralter, D., & Tyrer, P. J. (1985). The Effect of Number of Rating Scale Categories on Levels of Interrater Reliability: A Monte Carlo Investigation. *Applied Psychological Measurement*, 9(1), 31-36. doi: <https://doi.org/10.1177/014662168500900103>
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Corder, G. W., & Foreman, D. I. (2009). *Nonparametric statistics for non-statisticians: a step-by-step approach*. San Francisco: Wiley. doi: <https://doi.org/10.1002/9781118165881>
- Cox, E. P. (1980). The Optimal Number of Response Alternatives for a Scale: A Review. *Journal of Marketing Research*, 17(4), 407-422. doi: <https://doi.org/10.2307/3150495>
- Ferreira, M. F., & Backhoff-Escudero, E. (2016). Validez del Generador Automático de Ítems del Examen de Competencias Básicas (Excoba). *RELIEVE*, 22(1). doi: <https://doi.org/10.7203/relieve.22.1.8048>
- Freiberg Hoffmann, A., Stover, J. B., de la Iglesia, G., & Fernández Liporace, M. (2013). Correlaciones policóricas y tetracóricas en estudios factoriales exploratorios y confirmatorios. *Ciencias Psicológicas*, 7(2), 151-164.
- García Cueto, E., Muñiz Fernández, J., & Hernández Baeza, A. (2000). Comportamiento del modelo de respuesta graduada en función del número de categorías de la escala. *Psicothema*, 12(2), 288-291.
- Gilley, W. F., & Uhlig, G. E. (1993). Factor Analysis and Ordinal Data. *Education*, 114(2), 258.
- González-González, H., Álvarez-Castillo, J.-L., & Fernández-Caminero, G. (2015). Desarrollo y validación de una escala de medida de la empatía intercultural. *RELIEVE*, 21(2). doi: <https://doi.org/10.7203/relieve.21.2.7841>
- Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2008). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, 44(1), 153-166. doi: <https://doi.org/10.1007/s11135-008-9190-y>
- Hopkins, W. G. (2000). A new view of statistics. Recuperado a partir de <http://www.sportsci.org/resource/stats/>
- Lara, S. A. D. (2014). ¿Matrices Policóricas/Tetracóricas o Matrices Pearson? Un estudio metodológico. Recuperado 16 de junio de 2016, a partir de <http://www.redalyc.org/articulo.oa?id=333430869006>
- Lévy Mangin, J.-P. (2006). *Modelización con estructuras de covarianzas en ciencias sociales: temas esenciales, avanzados y aportaciones especiales*. España: Netbiblo.
- López González, E. (2012). Sugerencias para el análisis de escalas con métrica delicada. *Revista Iberoamericana de Evaluación Educativa*, 5(1), 84-105.
- Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the Number of Response Categories on the Reliability and Validity of Rating Scales. *Methodology*, 4(2), 73-79. doi: <https://doi.org/10.1027/1614-2241.4.2.73>
- Marcus-Roberts, H. M., & Roberts, F. S. (1987). Meaningless Statistics. *Journal of Educational and Behavioral Statistics*, 12(4), 383-394. doi: <https://doi.org/10.3102/10769986012004383>
- Muñiz, J. (1998). La medición de lo psicológico. *Psicothema*, 10(1), 1-21.

- Matell, M. S., & Jacoby, J. (1972). Is there an optimal number of alternatives for Likert-scale items? Effects of testing time and scale properties. *Journal of Applied Psychology*, 56(6), 506-509. doi: <https://doi.org/10.1037/h0033601>
- Maydeu-Olivares, A., Kramp, U., García-Forero, C., Gallardo-Pujol, D., & Coffman, D. (2009). The effect of varying the number of response alternatives in rating scales: Experimental evidence from intra-individual effects. *Behavior Research Methods*, 41(2), 295-308. doi: <https://doi.org/10.3758/BRM.41.2.295>
- Morales Vallejo, P. (2000). *Medición de actitudes en psicología y educación: construcción de escalas y problemas metodológicos*. Madrid: Universidad Pontificia Comillas.
- Morales Vallejo, P., Urosa, S., & Blanco, A. (2003). *Construcción de escalas de actitudes tipo likert: una guía práctica*. Madrid: La Muralla.
- Morata-Ramírez, M. de los Á., & Holgado-Tello, F. P. (2013). Construct Validity of Likert Scales through Confirmatory Factor Analysis: A Simulation Study Comparing Different Methods of Estimation Based on Pearson and Polychoric Correlations. *International Journal of Social Science Studies*, 1(1), 54-61. doi: <https://doi.org/10.11114/ijsss.v1i1.27>
- Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, 45(1), 171-189. doi: <https://doi.org/10.1111/j.2044-8317.1992.tb00975.x>
- Myers, N.D., Ahn, S., Lu, M., Celimli, S., & Zopluoglu, C. (2017). Reordering and reflecting factors for simulation studies with exploratory factor analysis. *Structural Equation Modeling*, 24(1), 112-128. doi: <https://doi.org/10.1080/10705511.2016.1230721>
- Nunnally, S. W. (2010). *Construction Methods and Management* (8 edition). Upper Saddle River, N.J: Pearson.
- Oliden, P. E., & Zumbo, B. D. (2008). Coeficientes de fiabilidad para escalas de respuesta categórica ordenada. *Psicothema*, 20(4), 896-901.
- Olmos Migueláñez, S., Martínez Abad, F., Torrecilla Sánchez, E. M., & Mena Marcos, J. (2014). Análisis psicométrico de una escala de percepción sobre la utilidad de Moodle en la universidad. *RELIEVE*, 20(2), art. 1. doi: <https://doi.org/10.7203/relieve.20.2.4221>
- Panther, A. T., Swygert, K. A., Grant Dahlstrom, W., & Tanaka, J. S. (1997). Factor analytic approaches to personality item-level data. *Journal of Personality Assessment*, 68(3), 561-589. doi: https://doi.org/10.1207/s15327752jpa6803_6
- Pearse, N. (2011). Deciding on the Scale Granularity of Response Categories of Likert type Scales: The Case of a 21-Point Scale. *Electronic Journal of Business Research Methods*, 9(2), 159-171.
- Pearson, K. (1900). Mathematical Contributions to the Theory of Evolution. VII. On the Correlation of Characters not Quantitatively Measurable. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 195, 1-405. doi: <https://doi.org/10.1098/rsta.1900.0022>
- Pearson, K. (1910). On a New Method of Determining Correlation, when One Variable is Given by Alternative and the Other by Multiple Categories. *Biometrika*, 7(3), 248-257. doi: <https://doi.org/10.2307/2345385>
- Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), 1-15. doi: [https://doi.org/10.1016/S0001-6918\(99\)00050-5](https://doi.org/10.1016/S0001-6918(99)00050-5)

- Richaud, M. C. (2005). Desarrollos del análisis factorial para el estudio de ítem dicotómicos y ordinales. *Interdisciplinaria*, 22(2), 237-251.
- Saris, W. E., & Coenders, G. (1995). Categorization and measurement quality. The choice between Pearson and Polychoric correlations. En W. Saris & A. Munnich, *The multitrait-multimethod approach to evaluate measurement instruments* (pp. 125-145). Budapest: Eotvos University Press. Recuperado a partir de <http://dare.uva.nl/record/1/113406>
- Serrano Angulo, J., Cebrián Robles, D. & Serrano Puerto, J. (2015). Control de calidad de datos obtenidos de cuestionarios en escalas Lickert. En AIDIPE, *Investigar con y para la sociedad* (pp. 167-176). Cádiz: Bubok Publishing S.L.
- Shafel, J., Brooke, L. N., & Gillmor, S. C. (2012). Effects of the Number of Response Categories on Rating Scales (pp. 1-24). Presentado en Anual conference of the American Educational Research Association, Vancouver.
- Siegel, S. (1970). *Estadística no paramétrica: aplicada a las ciencias de la conducta*. México: Trillas.
- Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, 103(2684), 677-680. doi: <https://doi.org/10.1126/science.103.2684.677>
- Ulitzsch, E., Holtmann, J., Schultze, M., & Eid, M. (2017). Comparing Multilevel and Classical Confirmatory Factor Analysis Parameterizations of Multirater Data: A Monte Carlo Simulation Study. *Structural Equation Modeling*, 24(1), 80-103. doi: <https://doi.org/10.1080/10705511.2016.1251846>
- Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27(3), 236-247. doi: <https://doi.org/10.1016/j.ijresmar.2010.02.004>
- Weng, L.-J. (2004). Impact of the Number of Response Categories and Anchor Labels on Coefficient Alpha and Test-Retest Reliability. *Educational and Psychological Measurement*, 64(6), 956-972. doi: <https://doi.org/10.1177/0013164404268674>

Authors / Autores	To know more / Saber más
<p>Martínez-Abad, Fernando (fma@usal.es).</p> <p>Doctor en Ciencias de la Educación con premio extraordinario por la Universidad de Salamanca, es actualmente Profesor Ayudante Doctor. Sus principales intereses de investigación son los métodos cuantitativos en Ciencias de la Educación y la evaluación del rendimiento académico y de competencias, habiendo participado en los últimos años en numerosos proyectos de investigación y publicaciones de impacto nacional e internacional. Su dirección postal es: Instituto Universitario de Ciencias de la Educación. Paseo de Canalejas, 169. 37008 Salamanca (España)</p>	<p> 0000-0002-1783-8198</p> <p> </p>
<p>Rodríguez-Conde, María José (mjrconde@usal.es).</p> <p>Catedrática de Universidad de la Universidad de Salamanca y directora de su Instituto Universitario de Ciencias de la Educación. Su principal línea de investigación se centra en la evaluación y calidad educativa. Su dirección postal es: Instituto Universitario de Ciencias de la Educación. Paseo de Canalejas, 169. 37008 Salamanca (España)</p>	<p> 0000-0002-2509-1901</p> <p> </p>



Revista Electrónica de **I**nvestigación y **E**valuación **E**ducativa
E-Journal of Educational Research, Assessment and Evaluation

[ISSN: 1134-4032]

© Copyright, RELIEVE. Reproduction and distribution of this articles it is authorized if the content is no modified and their origin is indicated (RELIEVE Journal, volume, number and electronic address of the document).

© Copyright, RELIEVE. Se autoriza la reproducción y distribución de este artículo siempre que no se modifique el contenido y se indique su origen (RELIEVE, volumen, número y dirección electrónica del documento).