Revista ELectrónica de Investigación y EValuación Educativa

RELIEVE

ISSN: 1134-4032

e-Journal of Educational Research, Assessment and Evaluation

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

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

Martínez-Abad, Fernando & Rodríguez-Conde, María José
University of Salamanca (Spain)

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 tetracórico-policórico. 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. Faculty of Education. University of Salamanca (Spain). fma@usal.es

The formal measurement of personality traits, attributes, attitudes, aptitudes, etc., in disciplines related to Psychology, Social Sciences or even Health Sciences is highly developed (Morales Vallejo, 2000; Morales Vallejo, Urosa, & Blanco, 2003). To ensure more accurate estimates, a series of multivariate statistics were developed, primarily from the early 20th century, to achieve very high levels of complexity today (Abad, 2011; Freiberg Hoffmann, Stover, de la Iglesia, & Fernández Liporace, 2013; Lévy Mangin, 2006; López González, 2012; Pearson, 1900, 1910), thanks in part to the calculation capacity of modern processors and the generalisation of specialised statistical packages.

The widespread use of non-metric measurement scales to estimate constructs is evidenced in this respect (Ferreyra & 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), and *the problem of measurement* (Marcus-Roberts & Roberts, 1987; Muñiz, 1998; Stevens, 1946) is still very much present in applied research: “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).

Despite the numerous statistical techniques available that respect the nature of measurement scales, the truth is that they have not replaced traditionally-used techniques. There are various reasons to explain this (López González, 2012):

- The widespread belief that alternative techniques (mostly non-parametric cases) have lower statistical power than traditional techniques. While this is true (Corder & Foreman, 2009; Siegel, 1970), the loss is minimal, even more so considering the sample sizes commonly used in today’s studies.
- Reduced benefits and versatility compared to classic techniques.
- Researchers’ lack of knowledge (as they are accustomed to classic techniques) or a lack of alternative techniques offered in the main, widely-used statistical packages.

An added issue to these factors is that part of the Social and Health Sciences scientific community is still convinced that using classic techniques entails minimum bias, stressing that the loss of benefits suffered with the change is not offset by the small gain in reduced estimation error (Morales Vallejo, 2000; Nunnally, 2010).

Therefore, mid-way between using alternative techniques and maintaining classic procedures is *tetrachoric* (between dichotomous variables) or *polychoric* (between polytomous ordinal variables) correlation matrix calculation for an unbiased estimate of the relationship between pairs of items on a scale with non-numerical measurements (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). Thus, we can use the potential of classic techniques without significant bias in the estimation of relationships between categorical variables. Note that, under this study and based on the literature review, the tetrachoric-polychoric estimator is understood as the most appropriate correlation coefficient in cases where multivariate statistical techniques with variables obtained from a Likert response scale are applied, in which it can be understood that a latent continuous scale underlies the categories of the scale observed.

This correlation coefficient, first defined theoretically by Pearson (1900), obtains the underlying continuous variables (latent) between each pair of ordinals (x, y), for which it uses the properties of the density function of normal bivariate distribution:

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

Density function of normal bivariate distribution

where we must use maximum likelihood procedures to estimate the parameters of thresholds a_i and b_j, and the correlation coefficient between the latent continuous variables ρ (Holgado–Tello et al., 2008). This coefficient ρ, estimated based on the density function of normal bivariate distribution, is what is known as the *tetrachoric or polychoric correlation coefficient*. Therefore, it is important to note that, although many studies prove that the lack of normality in original variables does not significantly affect how this coefficient is calculated (Freiberg Hoffmann et al., 2013; Morata-Ramírez & Holgado-Tello, 2013), this basic assumption must be considered in the calculation. It is well-known that in many of the scales applied in Social Sciences, primarily survey studies using Likert response scales (Serrano Angulo, Cebrián Robles & Serrano Puerto, 2015), univariate and multivariate normality is generally breached.

Although the complexity of calculating the tetrachoric-polychoric correlation is evident, the processing capacity of modern computers makes it possible to use this procedure to calculate correlation matrices even in cases involving ordinal variables with multiple response levels and large sample sizes. Thus, there has been extensive scientific output on the study of different estimates offered by the product-moment correlation coefficient and the tetrachoric-polychoric correlation coefficient in non-quantitative variables in recent years.

Most of these studies focus on verifying how general reliability and validity indexes behave in exploratory and confirmatory factor analysis based on both correlation matrices (Burga León, 2012; Freiberg Hoffmann et al., 2013; Gilley & Uhlig, 1993; Holgado–Tello et al., 2008; Muthen & Kaplan, 1992; Panter et al., 1997; Richaud, 2005), concentrating almost exclusively on the number of response categories of the ordinal variable (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). The majority of these studies seem to agree that, to avoid the bias associated with using the product-moment correlation coefficient, the original variables must have at least five response levels (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).

However, no studies have been found that directly analyse the differences between the product-moment and tetrachoric-polychoric correlation matrices resulting from a series of ordinal items. Meanwhile, few papers are published that take into account other factors affecting the estimation of both correlation coefficients, such as level of relationship between pairs of variables (Lozano et al., 2008), asymmetry levels of the marginal distributions of ordinal variables, or the sample size available in the data set.

Despite the above, information available to researchers on the conditions and benefits of using the tetrachoric-polychoric correlation coefficient is unclear, insufficient and sometimes even contradictory. While some forums passionately defend the need for tetrachoric-polychoric estimation in the case of any dichotomous or ordinal scale (López González, 2012), other experts note that as of a certain quantity of response levels in the ordinal scale it is possible (and even recommended for computational reasons) to calculate the product-moment correlation matrix, providing partial, non-generalisable or biased evidence to support this recommendation (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; Muthen & Kaplan, 1992; Oliden & Zumbo, 2008; Panter et al., 1997; Preston & Colman, 2000; Richaud, 2005; Shafel et al., 2012; Weijters, Cabooter, & Schillewaert, 2010; Weng, 2004).

Specifically, understanding tetrachoric-polychoric estimation as the most suitable estimation in these cases, the purpose of this study is to use a systematic simulation procedure to assess how various conditions affect product-moment estimation in ordinal variables compared to the polychoric-tetrachoric correlation coefficient. These conditions are related to: number of response categories, asymmetry, sample size and intensity of the relationship between these variables. Under the fundamental premise that the tetrachoric-polychoric correlation coefficient offers an unbiased estimation of the relationship between categorical variables (in this case ordinal) that originally come from a latent continuous variable, based on the results of this research we expect to provide specific recommendations for Social and Health Science researchers based primarily on the direction and intensity of these differences. Therefore, based on the empirical evidence gathered, the aim is to establish under which conditions the general recommendation of using the tetrachoric-polychoric correlation coefficient to estimate the relationship between variables on a scale is maintained, and in which cases it might be simpler and therefore more appropriate to use the solution provided by the product-moment estimate as there are no significant differences between both estimates.

Method

In terms of research design, the study is based on the simulation of ordinal data. A total of 90 databases were simulated depending on the different conditions applied. Each database included 10 variables or items, considering that they came from a one-dimensional scale in all cases. The experimental conditions imposed on the data were the number of item response categories (2, 3, 4, 5 or 7), distribution of items (symmetric or negative asymmetric), sample size (small, medium or large) and the relationship between items (low, medium or high).

Data simulation studies are considered to be a widespread, highly useful formal research procedure for the scientific community. There is an extensive bibliography related to studying the behaviour of multivariate techniques and statistics using simulated data, which has become widespread in recent years (Cain, Zhang & Yuan, in press; Myers, Ahn, Lu, Celimli & Zopluoglu, 2017; Ulitzsch, Schultze & Eid, 2017).

The main hypothesis of the research project suggests *the existence of diverse conditions related to the number of response categories, data symmetry, sample size and relationship between the items will have different significant effects on the behaviour of the Pearson correlation coefficient compared to the tetrachoric-polychoric correlation coefficient.*

Based on the contributions of previous research projects studies, the following derived hypotheses are also proposed:

- H1: The value of the product-moment correlation coefficient estimate will be lower than the tetrachoric-polychoric estimate, in absolute terms, when the number of response levels in variables is low. Both estimates will be similar in high response levels, and the distance between the estimates will increase as the number of response levels in variables decreases.

- H2: There will be no clear trend that enables us to characterise the differences in tetrachoric-polychoric and product-moment estimates based on the different levels of asymmetry studied in the variables.
- H3: The three sample size levels studied will not affect tetrachoric-polychoric and product-moment estimate trends differently.
- H4: The relationship found by the product-moment estimator will be less intense than that of the tetrachoric-polychoric estimator when the real relationship between variables is high. Specifically, while in low relationship levels both estimates will be similar, as the relationship between variables rises the distance between both estimates will tend to increase.

Based on the **variables** used in the study, the different correlations (product-moment - tetrachoric-polychoric) between each pair of the 10 variables included in each simulated database was identified as the dependent variable. The independent variables in the study are the different conditions applied in the simulations:

- *Number of ordinal variable response categories*: Data were simulated with 2, 3, 4, 5 and 7 response levels. These levels were selected as they are the most commonly used in Social Science studies with Likert response scales. Previous studies analysed in this paper indicate that these response levels show the greatest differences between both correlation coefficients; the differences drop to a minimum as of 7 response levels.
- *Variable symmetry levels*: Given the issues indicated in the literature review regarding normality and symmetry, data were simulated with normal distributions (symmetry) and with accentuated negative asymmetry.
- *Sample size*: Given the variety of sample sizes used in studies in the field of Social Sciences and Health Sciences, the effect of three different sample sizes must be studied: small samples (N=30) adopting the limit commonly established by traditional statistics, medium samples (n=200) and large samples (n=1000).
- *Relationship level between variables*: The fact that the relationship between variables in survey studies tends to reach moderate values was taken into account, therefore it is understood that a low relationship level is r≈.2, medium relationship level is r≈.5, and high relationship level is r≈.7 (it is difficult to reach higher values in Social Sciences and they can be associated with problems of collinearity and multicollinearity). Note that the relationship levels between the 10 items in each of the experimental situations is not exactly the proposed value, but the package used to simulate data includes the criterion that the correlation between each pair of items should be close to the indicated value, with a low level of error. Thus, in all cases the errors in product-moment correlation estimates (the indicator used to simulate data to indicate the relationship between variables) are approximately ±0.1 points around the value of the forced relationship used in the simulation.

From the alternatives available to **simulate data**, R software was selected, specifically the GenOrd[1] package. This package offers the flexibility, reliability and tools necessary to implement simulated databases under the conditions indicated above.

Finally, regarding **data analysis,** it was necessary to use various statistical packages. The correlation and correlation difference matrices were calculated by generating a brief application in R (the Polycor[2] package was used to estimate the tetrachoric-polychoric correlation). Based on this information, a database was generated with information from the 4050 correlation differences (45 indicators

[1] Simulation of Discrete Random Variables with Given Correlation Matrix and Marginal Distributions. Documentation: https://cran.r-project.org/web/packages/GenOrd/GenOrd.pdf

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

for each of the 90 experimental situations) and from the different conditions applied in each situation.

The Analysis of Variance (ANOVA) technique was applied to analyse the effects of the different experimental conditions on the correlation differences, considering each experimental condition as a factor of the model, at a significance level of 5%. As a complementary measure to the hypothesis contrasts implemented, the effect size of the ANOVA model factors (statisticalη^2) and the contrasts of the post-hoc tests (Cohen's d statistic) were calculated. Finally, graphics were generated of the correlation differences. All calculations were completed using SPSS V.23 and Microsoft Excel 2010 software (both under the USAL Campus license).

Results

Firstly, given that it is a variable that did not directly affect the configuration and characteristics of the simulated ordinal variables, we decided to use three sample size levels (small, medium and large) as groupings to segment the database. As a result of this decision, three independent ANOVA models were generated and the number of response categories, asymmetry levels and level of relationship between variables were integrated as fixed factors. The final models also integrated the interaction effect of all the combinations of these factors. Also, based on this issue, descriptive results are presented for small, medium and large sample size simulated databases. Thus, table 1 shows the correlation difference in the different simulated response levels. It shows how in all cases the product-moment correlation estimated lower relationships between variables than the tetrachoric-polychoric correlation. This trend is reduced as the number of response options increases. The different sample sizes do not appear to have a clear effect on the correlation differences, although variability clearly falls as the sample size increases.

Table 1. Average correlation difference depending on response level

	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

Regarding the effect the different relationship levels between variables have on the average correlation difference (table 2), note how this increases as the relationship level between items becomes more important. Equally, all cases show that the product-moment correlation estimates lower-intensity relationships than the tetrachoric-polychoric correlation, and that the different sample sizes show no clear trend regarding average effects.

Table 2. Average correlation difference depending on relationship levels.

	r_{xy}≈.2		**r_{xy}≈.5**		**r_{xy}≈.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

Finally, the effect associated with the type of distribution (symmetric-asymmetric) of ordinal data. In this case the trend is not so clear, although it can be seen that asymmetric distributions tend to slightly increase the correlation difference in small and large samples. Note that the normality of variables is an assumption for calculating the tetrachoric-polychoric correlation coefficient, but breaching this assumption has no significant effect on bias in calculating this coefficient (Freiberg Hoffmann et al., 2013; Holgado–Tello et al., 2008; Morata-Ramírez & Holgado-Tello, 2013).

Table 3. Correlation difference depending on relationship levels.

	SYMMETRIC DISTRIBUTION		ASYMMETRIC DISTRIBUTION	
	$\bar{X}$	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

Regarding the inferential analyses conducted, table 4 gives an overview of the adjustment of the three models with and without the effects of interactions. In all cases they are highly significant models with very important explained variance percentages (R^2, coefficient of determination), which increase as the simulated sample sizes get bigger. These results offer us an internally valid indicator and consistency in the perspective used in the study, confirming that the factors selected are essential in the configuration of the different estimates of product-moment and tetrachoric-polychoric indicators.

Table 4. Significance and explained variance fixed effect and complete ANOVA models.

	F	p.	R^2
Small sample (n=30). Fixed effects	276.11	<.001	.588
Small sample (n=30). Complete model	88.10	<.001	.652
Average sample (n=200). Fixed effects	1092.72	<.001	.850
Average sample (n=200). Complete model	600.70	<.001	.928
Large sample (n=1000). Fixed effects			
Large sample (n=1000). Complete model			

The coefficient of determination is significantly higher in complete models, raising the possibility of significant interactions between factors. Therefore, we deemed appropriate to incorporate these interactions into the three final models. The first model, for simulated databases with a sample size of 30, is shown in table 5. In this case, the symmetry factor is insignificant with a very small effect size (eta-squared statistic). On the other hand, most interactions are significant, with low or moderate effect sizes in significant interactions, and high in significant fixed factors (Hopkins, 2000).

Table 5. ANOVA for correlation differences. Small sample size (n=30).

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

Given that the model offers highly significant interactions it could be difficult to interpret the main factors. Therefore, figure 1 shows the average marginals of the correlation differences of the number of response options depending on the factors of symmetry and relationship. Although some interactions can be observed in the diagrams, it is evident that the estimates of both correlation coefficients clearly tend to get closer as the number of response options increases and the relationship between the variables is less intense. The approximation between the estimates of the correlation coefficients is important for 2, 3, 4 and 5 response levels, and the trend stabilised between 5 and 7 levels.

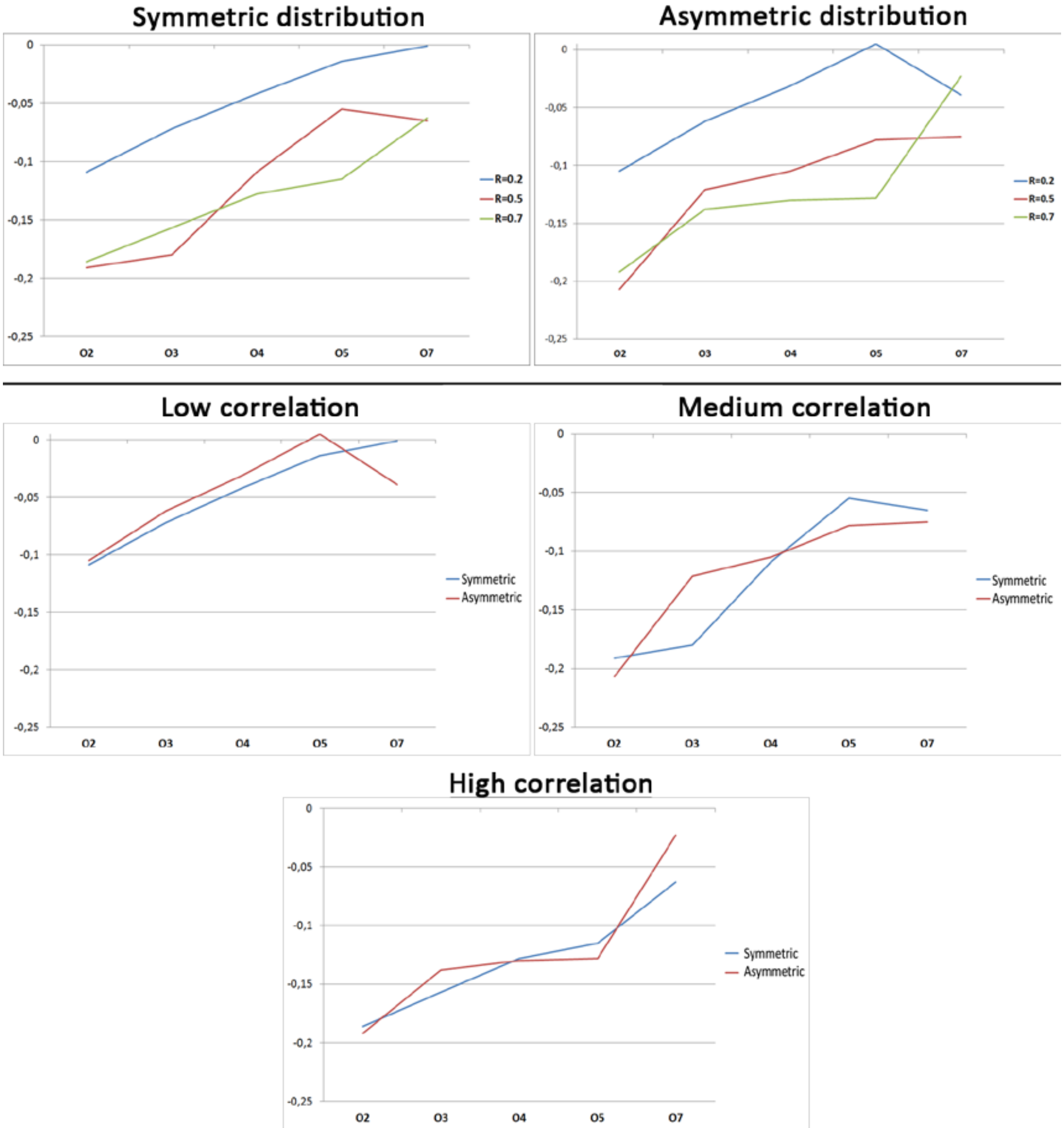

Figure 1. Averages for factors with greater interaction (n=30)

The model applied for samples with 200 subjects (table 6) shows significant effects in all factors and interactions, with effect sizes ranging from moderate (symmetry) to very high (options and relationship). It appears that for this sample size the estimates of correlation coefficients stabilise and the errors are minimised. Interaction continues to have a significant effect.

Table 6. ANOVA for correlation differences. Average sample size (n=200).

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

Figure 2 shows how this interaction is less pronounced than in the case of small sample sizes, and the trends are more stable.

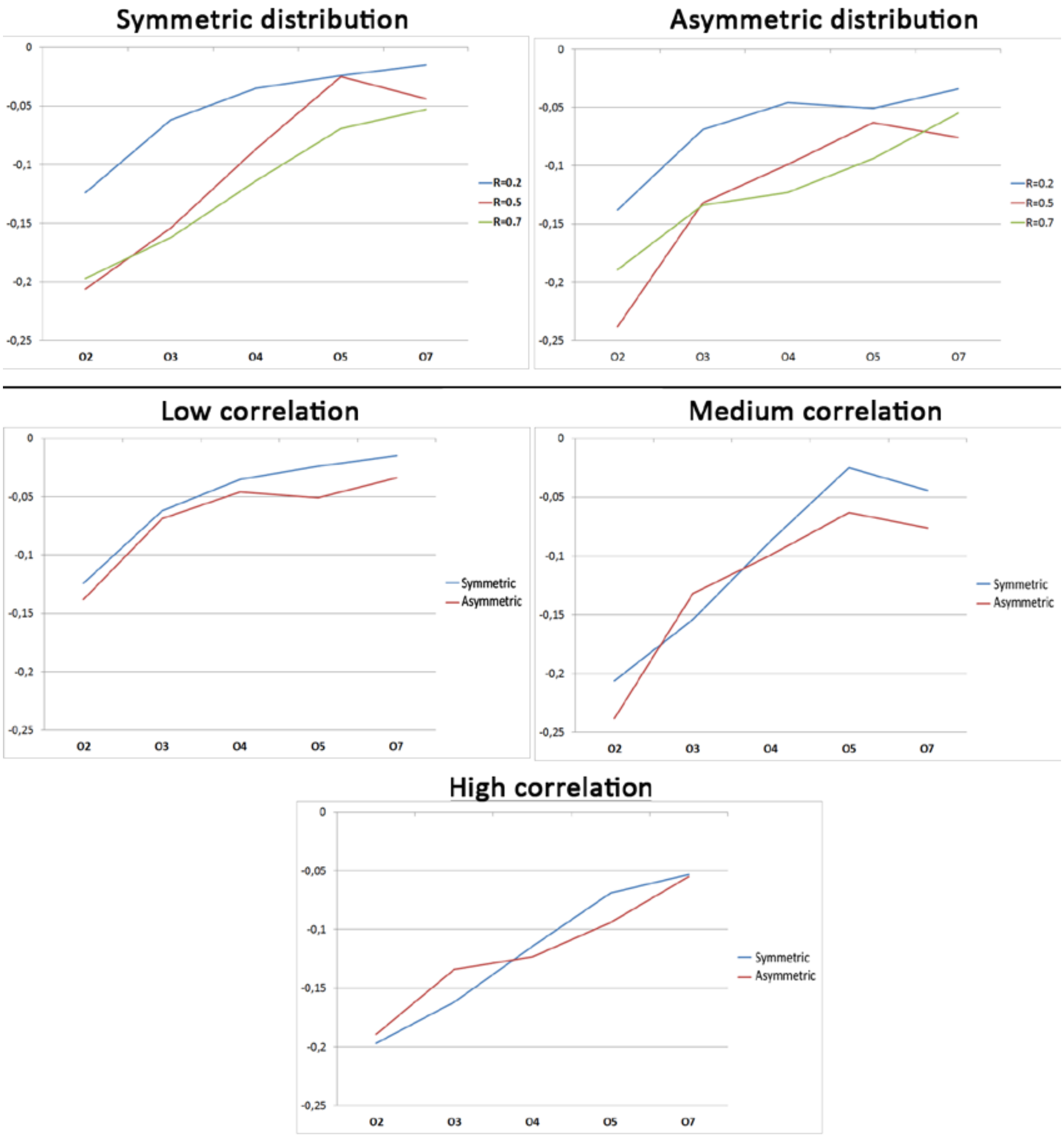

Figure 2. Averages for factors with greater interaction (n=200)

The last model, which incorporates the simulated data of large samples (n=1000), is shown in table 7. This confirms the trends observed above. While the number of response options, followed closely by the relationship level between variables, is the factor with the most important effects, the symmetry-asymmetry of the marginal distributions of simulated variables has significant effects with much more limited effect sizes. Interactions are still significant and the effect of the interaction between the number of response options and the relationship between variables is especially important.

Table 7. ANOVA for correlation differences. Large sample size (n=1000).

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

Given that the significant interaction effects indicated are maintained, figure 3 details these issues in the case of large samples. The trend continues to stabilise, showing lighter interactions. This makes it possible to directly interpret the main factors, except symmetry, which continues to show symptoms of significant interaction with the number of response options, although more moderately.

In any case, these figures and ANOVA techniques with various applied factors confirm that the product-moment correlation coefficient estimates, in absolute terms, that the relationship between ordinal variables is lower (compared to tetrachoric-polychoric estimation) when these have few response options in the scale and when the relationship between the variables is more intense. Meanwhile, it seems that when the number of response options is 5 or more, the differences between both correlation coefficients are minimised, approaching 0. The opposite occurs with the level of relationship between variables. It seems that, while there are very few differences between the estimates of both correlations when the relationship between the variables is low (r=2), this increases greatly when the relationship reaches .5; stagnating at this level and remaining similar when the relationship is .7. In fact, in the post-hoc tests shown below, the only pair in which the differences are not highly significant is that comprising small sample simulated databases with relationships between medium and high variables.

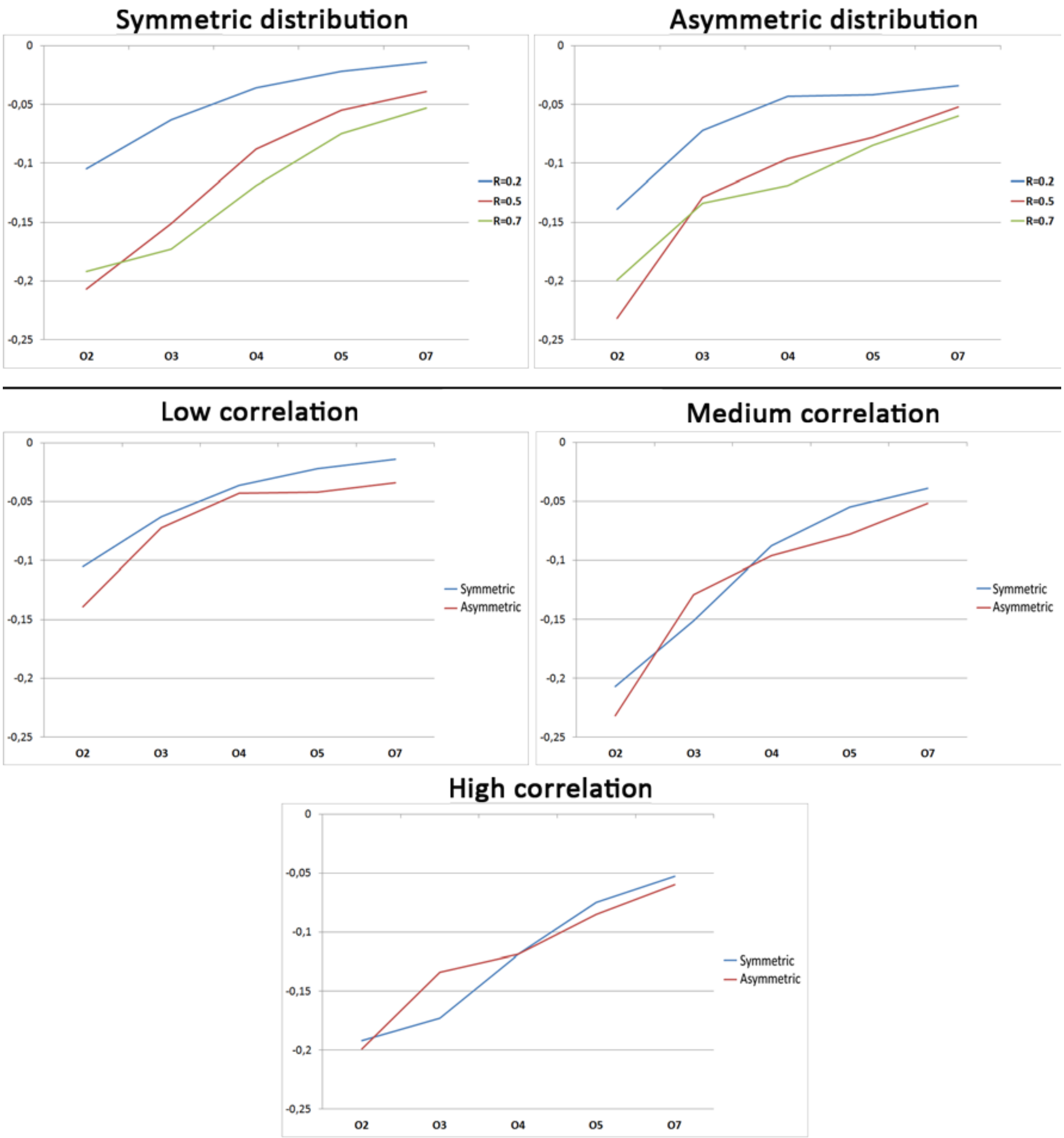

Figure 3. Averages for factors with greater interaction (n=1000)

Regarding post-hoc tests, as anticipated above, all contrasted pairs are significant. For a more detailed analysis of this significance we calculated the statistics of Cohen's d effect size. Table 8 shows the effect sizes for post-hoc tests between all possible pairs of the number of category responses factor of the ordinal variable. As anticipated in the above graphic analysis, it is confirmed that the effect sizes become more important when comparing pairs that are distant in terms of the number of response options, and that, when comparing consecutive pairs with a high number of response options, the effect size is lower. For example, whereas when comparing samples with 2 or 3 response options obtains effect sizes over 1 in the three sample size levels, when samples with 5 and 7 response options are compared, two of these values do not achieve a value of the effect size over .5, lower than the values we could interpret as having moderate effects.

Table 8. Post-hoc test effect sizes (Cohen's d). Response options.

		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

Likewise, table 9 shows the effect sizes of post-hoc tests in the case of the relationship between variables factor. Once again, we observe how effect sizes increase systematically as more distant relationship levels are compared and when comparing the low relationship with the moderate. In fact, when comparing average and high relationship levels, the effects are low for two of the three sample sizes.

Table 9. Post-hoc test effect sizes (Cohen's d). Relationship between variables.

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

Therefore, the evidence shows the following general questions:

- Symmetry-asymmetry levels do not appear to have major effects on the differences between tetrachoric-polychoric estimation and product-moment estimation, and the differences located do not show a clear trend. This must also be noted with regards the subsequent proposal that the bias associated with the symmetry-asymmetry factor may be due to the effects of interaction between this and the other factors in the estimation of the product-moment correlation coefficient, and with the breach of the assumed normality in tetrachoric-polychoric estimation.
- While the number of response levels of ordinal variable seems to achieve the highest effects, no large effect sizes are found in post-hoc tests when comparing 5 response levels with 7 levels.
- The starting level of relationship between the variables also appears to reach highly significant effects, with high effect sizes (eta-squared value), in the models proposed, although the effect sizes of post-hoc test indicate that there are no major differences between the medium and high relationship levels.

Having developed this analysis set, we are in conditions to offer a reasoned and well-defined proposal of the scenarios in which, given the significant underestimation the product-moment correlation coefficient offers of the true relationship between variables (understanding the tetrachoric-polychoric coefficient as the unbiased estimator), it may be recommendable to use tetrachoric-polychoric estimation. On the contrary, some other cases are determined in which, as there are no major differences between both correlation coefficients (based on moderate or low effect sizes), and based on the principle of parsimony which must be present in any statistical analysis, the use of the product-moment coefficient is deemed more recommendable. In this regard, we must not forget that in this case we can assume the perspective of some authors (Morales Vallejo, 2000; Nunnally, 2010) who stress that if the use of classic techniques entails minimum bias, given the loss of benefits associated with applying the tetrachoric-polychoric estimator, its use is not justified. Table 10 summarises the proposal. As it is necessary to gather more in-depth information on this regard from various sources of information to provide a more grounded and specific proposal, we offer a conservative general recommendation.

Table 10. Recommended correlation coefficient based on the data matrix.

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

* N=The use of any type of multivariate technique is not recommended
** T-P=Tetrachoric-polychoric correlation coefficient
*** P-M=Product-moment correlation coefficient

The criteria follow to establish the proposal were the following:

1. As analyses have proven that the existence or lack of symmetry has no differentiating effect in both estimations, this variable is not considered in the proposal.
2. In cases with a small sample (n=30), we must consider that tetrachoric-polychoric estimation is not recommended with samples sizes under 150 subjects (Freiberg Hoffmann et al., 2013), due to the high instability this estimate can entail. Meanwhile, regarding the use of product-moment estimation, given the major differences found in both estimations in cases with fewer than 5 response levels, and in cases with medium and high relationships between variables, its use in not recommended in these cases. Therefore, this estimator is only recommended with small samples when the relationship between variables is low and the response levels are no less than 5.
3. When sample sizes are medium to large, the same recommendation applies based on the indicators obtained in prior analyses. As the differences in product-moment and tetrachoric-polychoric estimations, as shown in the figures and models applied, are very large when the original variables have fewer than 4 response levels, it is understood that the tetrachoric-polychoric correlation coefficient must be used in all such cases. Nevertheless, as the graphic analysis shows that the difference in correlations in minimal when the

relationship between variables is low and we have 4 or 5 response levels, it is reasonable to use the product-moment correlation in these cases.

4. Finally, given that the evidence shows that the difference in correlations in minimal in all cases with 7 response levels in the variables (achieving correlation differences close to or lower than .05 points), it is reasonable to use the product-moment correlation with scales of this type (or with more response levels).

Discussion

Using instruments to measure traits, characteristics, attitudes, etc. is common practice in the field of Social and Health Sciences (Abad, 2011; Morales Vallejo, 2000), and non-quantitative measurement scales such as, primarily, Likert response scales, are widespread (García Cueto et al., 2000; Pearse, 2011; Preston & Colman, 2000; Shafel et al., 2012). Therefore, the widespread use of statistical techniques inherent to interval or reasoning scales in these non-metric variables is in question, at least from a mathematical perspective (Marcus-Roberts & Roberts, 1987; Stevens, 1946). However, while it is difficult to locate clear, shared criteria on the most suitable multivariate techniques based on the scales and items available, the scientific community has not reached a consensus on the suitability of replacing classic techniques.

It is clear that, regarding the study of estimation properties of the tetrachoric-polychoric correlation compared to the product-moment correlation, the literature review does not generally consider the different characteristics of the set of ordinal variable in the scale (Lozano et al., 2008; Maydeu-Olivares et al., 2009; Oliden & Zumbo, 2008). Meanwhile, most studies directly analyse the effects the different estimations of these two coefficients have on the results of factorial analyses in these scales (Burga León, 2012; Freiberg Hoffmann et al., 2013; Gilley & Uhlig, 1993; Holgado–Tello et al., 2008; Muthen & Kaplan, 1992; Panter et al., 1997; Richaud, 2005). Thus, no papers that directly analyse the differences between these two estimators have been found, contemplating the main factors that can bias estimations.

In this context, the paper aims to contribute useful information for applied researchers regarding the bias in product-moment correlation coefficient estimation, based on various key characteristics of ordinal variables from which Likert response scales are taken, understanding that the tetrachoric-polychoric estimator offers an unbiased index in cases where a latent continuous scale is understood as underlying the ordinal scale observed. The objectives and hypotheses proposed initially have been met: a procedure for the simulation of ordinal data controlling many factors that characterise ordinal variables which, based on the literature review, affect product-moment estimation has been studied and applied; a set of simple, clear data analysis techniques have been applied to the basic level of the difference in correlations to enable comprehensive interpretation of the results; and finally, a specific, operational proposal was conducted, based on the premise that tetrachoric-polychoric estimation is initially the most appropriate for calculating the relationship between variables from a Likert response scale, indicating cases in which the difference in correlations between the two estimators are considered so insignificant that, based on the principle of parsimony and the perspective of some leading authors (Morales Vallejo, 2000; Nunnally, 2010), use of the product-moment correlation is recommended.

Regarding the evidence obtained, the robustness of the variance analysis models developed is worthy of note; these attained goodness of fit levels over 90%, and even very close to 100%, when the sample sizes were sufficiently large. The results indicate that, while the number of response categories of variables and the level of relationship between them are the most important factors, asymmetry levels have limited, undefined influence. Meanwhile, sample size only implies different levels of variability in the estimations. Note that, despite the fact that these goodness of fit levels are infrequent, even with large samples, they may be largely influenced by the algorithms incorporated in the statistical package used to implement the data simulation procedure; this issue could be subject to analysis in future studies.

More specifically, evidence confirms that in variables with more than 5 response levels, unless the relationship between the items is very high (extremely uncommon in this type of scale), the use of the product-moment correlation coefficient does not noticeably undermine the relationship between variables compared to tetrachoric-polychoric estimation. These results concur with those obtained in the previous studies analysed, which identify a critical point in variables as of 5 response levels (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). Furthermore, the use of the tetrachoric-polychoric correlation coefficient appears to be mandatory in the case of variables with 5 or less response levels, unless the relationship between them is very low. In this sense, as the relationship between variables increases, so do the differences between the two correlation estimations. Thus, in the case of variables with 5 response options, if the relationship is low to medium-low, using the product-moment correlation coefficient could be considered appropriate; this affirmation could be controversial in all other circumstances.

Regarding variable asymmetry-symmetry levels, although asymmetric distributions entail a slight increase in differences between both correlation coefficients, it is not clear that this increase is due to the underestimation of the product-moment correlation coefficient on this type of variables, or to the small bias in the tetrachoric-polychoric correlation coefficient estimation when there is no univariate-multivariate normality (Freiberg Hoffmann et al., 2013; Holgado–Tello et al., 2008; Jöreskog, 1994; Morata-Ramírez & Holgado-Tello, 2013; Olsson, 1979). Therefore, the proposal in this paper does not consider a difference between the symmetry-asymmetry levels of the reference variables.

Finally, we must highlight the strengths and weaknesses of this study. Positive aspects that endorse this paper include, as indicated above, the multitude of factors taken into account during simulation and the simplicity of the statistical analysis of the differences. This made it easier and simpler to interpret the results in-depth, increasing the internal validity of the process. However, the selection of multiple factors and not having included more complex, global techniques, such as statistical hypothesis testing to directly compare the pairs of correlation matrices, could also be considered a weakness. In this sense, not including more global techniques and the variety of factors analysed are issued that could, from our perspective, threaten the external validity of the results obtained. In fact, it should be noted that the proposal, despite basically concurring with the literature review and being based on the evidence gathered, could be considered partial and in some way insufficient as it is only based on the difference in correlations obtained from both estimators, without taking other significant statistical-theoretical questions into account. Finally, the fact that this study is limited to data simulation level is an added difficulty for the external validity of the results as it entails omitting some biases and casuistries inherent to providing real subjects with measurement scales.

The weaknesses analysed provide new possibilities for future studies, such as replicating similar studies based on samples of real subjects, moving away from the laboratory situation presented here. It would also be very interesting to study bias and the most appropriate statistical techniques more in-depth in cases where variables have high levels of asymmetry. Finally, we must not forget that this paper has been developed under the assumption of one-dimensional scales so it is therefore possible to replicate the study by simulating or applying multidimensional scales.

References

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. In *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. (in press). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, inflluence 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

Ferreyra, 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. Retrieved from http://www.sportsci.org/resource/stats/

Lara, S. A. D. (2014). ¿Matrices Policóricas/Tetracóricas o Matrices Pearson? Un estudio metodológico. Retrieved June 16, 2016, from 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

Muthen, 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

Panter, 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

Richaud, M. C. (2005). Desarrollos del analisis factorial para el estudio de item dicotomicos y ordinales. *Interdisciplinaria*, *22*(2), 237-251.

Saris, W. E., & Coenders, G. (1995). Categorization and measurement quality. The choice between Pearson and Polychoric correlations. In W. Saris & A. Munnich, *The multitrait-multimethod approach to evaluate measurement instruments* (pp. 125-145). Budapest: Eotvos University Press. Retrieved from 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. In 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). Presented in 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
Martínez-Abad, Fernando (fma@usal.es). Doctorate in Educational Sciences with Outstanding Doctorate Award by the University of Salamanca, he is currently an Assistant Professor and Academic Secretary for the Department of Didactics, Organisation and Research Methods at the same University. His main research focuses on quantitative methods in Educational Sciences, and assessing academic performance and skills. His postal is: Instituto Universitario de Ciencias de la Educación. Paseo de Canalejas, 169. 37008 Salamanca (Spain)	ORCID 0000-0002-1783-8198 Google académico ResearchGate
Rodríguez-Conde, María José (mjrconde@usal.es). Director of the Institute of Educational Sciences and Chair at the University of Salamanca, her teaching and research activity has focused on the field of educational assessment and quality. Her postal is: Instituto Universitario de Ciencias de la Educación. Paseo de Canalejas, 169. 37008 Salamanca (Spain)	ORCID 0000-0002-2509-1901 Google académico ResearchGate

RELIEVE

Revista **EL**ectrónica de **I**nvestigación y **EV**aluación **E**ducativa

E-Journal of Educational Research, Assessment and Evaluation

[ISSN: 1134-4032]