



Artificial systematic attenuation in eta squared and some related consequences: attenuation-corrected eta and eta squared, negative values of eta, and their relation to Pearson correlation

Jari Metsämuuronen^{1,2}

Received: 22 April 2021 / Accepted: 3 March 2022
© The Author(s) 2022

Abstract

In general linear modeling (GLM), eta squared (η^2) is the dominant statistic for the explaining power of an independent variable. This article discusses a less-studied deficiency in η^2 : its values are seriously deflated, because the estimates by coefficient eta (η) are seriously deflated. Numerical examples show that the deflation in η may be as high as 0.50–0.60 units of correlation and in η^2 as high as 0.70–0.80 units of explaining power. A simple mechanism to evaluate and correct the artificial attenuation is proposed. Because the formulae of η and point-biserial correlation are equal, η can also get negative values. While the traditional formulae give us only the magnitude of nonlinear association, a re-considered formula for η gives estimates with both magnitude and direction in binary cases, and a short-cut option is offered for the polytomous ones. Although the negative values of η are not relevant when η^2 is of interest, this may be valuable additional information when η is used with non-nominal variables.

Keywords Coefficient eta · Eta squared · Pearson correlation · Product–moment correlation coefficient · Goodman–Kruskal gamma · Somers delta · Attenuation · Attenuation-corrected reliability

✉ Jari Metsämuuronen
jari.metsamuuronen@gmail.com

¹ Finnish Education Evaluation Centre (FINEEC), Hakaniemenranta 6, P.O. Box 380, 00531 Helsinki, Finland

² Centre for Learning Analytics, University of Turku, Turku, Finland

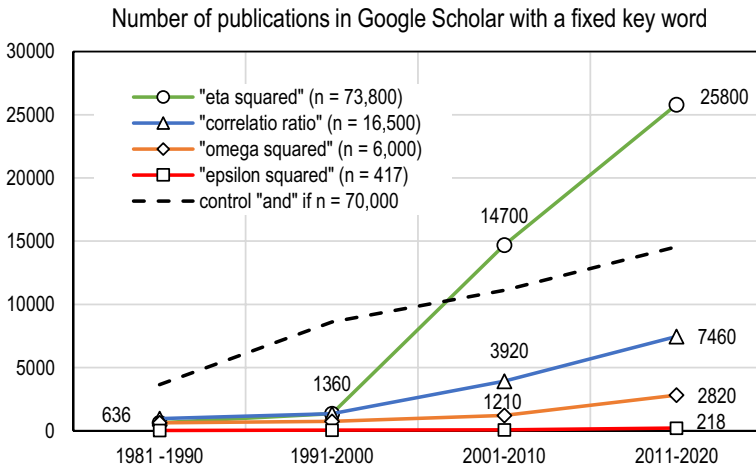


Fig. 1 Historical occurrence of the traditional measures of explaining power in Google Scholar

1 Introduction: coefficient eta and eta squared

Coefficient eta (η ; Pearson 1903, 1905), sometimes—specifically, in the early days—called the correlation ratio (e.g., Pearson 1911; Ayres 1920; Fisher 1925; Kelley 1935), is one of the oldest directional measures of association. Originally, Pearson proposed it as a measure of the relationship between a categorical and a continuous variable, although it can be used also with ordinal and interval-scaled variables. Fisher, however, was not convinced of the usefulness of correlation ratio and in his influential *Statistical Methods for Research Workers* he noted: “As a descriptive statistic the utility of the correlation ratio is extremely limited” (Fisher 1925, p. 219). However, η turned out to be the dominant measure to quantify a curvilinear association (e.g., Howell 2012; Sechrest and Yeaton 2011; see also Ayres 1920 and Peters and Van Voorhis 1940) and, as eta squared (η^2) and partial eta squared (η_p^2 ; Cohen 1965, 1973), it turned out to be the dominant indicator of the explaining power between two variables in settings related to the analysis of variance (ANOVA) and covariance (ANCOVA) or, more generally, in general linear modeling (GLM; e.g., Cohen 1965, 1973, 1988).

The rise of η^2 was influenced by the rise in the discussions about effect sizes related to the proportion of variance in the 1960s (e.g., Cohen 1965, 1969; Friedman 1968; Kerlinger 1964). Even though epsilon squared (ϵ^2 ; Kelley 1935), omega squared (ω^2 ; Hays, 1963) or, recently, adjusted (partial) eta squared ($\text{adj}\eta_p^2$, later η_{adj}^2 ; Mordkoff 2019)¹ are suggested to be used instead of η^2 for the inferential statistic related to population (see the history, differences, and literature in Glass and

¹ In the case where the association of just two variables is of interest, the attribute “partial” is not relevant. Then, Mordkoff’s coefficient could be called adjusted eta squared (η_{adj}^2). This notation is used in what follows.

Hakstian 1969; Okada 2013; 2017; Richardson 1996), η^2 is the most often used measure of these three (see Fig. 1). From the year 1990 onwards, the trend of fixed keyword “eta squared” has rocketed against the (trend of) proportion of all publications (as indicated by a common keyword “and” in Fig. 1).

Although ω^2 and, specifically, ε^2 are shown to give less biased estimates than η^2 (see Okada 2013, 2017; Mordkoff 2019), some positive characteristics of η are that, first, η^2 can be interpreted in the same manner as a squared partial correlation coefficient (r_p^2) from multiple linear regression: as the proportion of remaining variance in the dependent variable (Mordkoff 2019). Second, while all unbiased estimates ω^2 , ε^2 , and η_{adj}^2 are prone to give negative (out-of-range) values when η^2 is near 0, specifically, with the small sample sizes (Okada 2013, 2017), the estimates by η^2 always stay within the limits although with a considerable amount of bias. The positive bias in η and η^2 is caused by the fact that even if the true means in all categories of the categorical variable are equal leading to $\eta = \eta^2 = 0$, variance across the sample means is rarely zero. Random differences around zero will cause the value of η^2 to be greater than zero (Mordkoff 2019). In what follows, the *opposite* challenge of η and η^2 is also discussed: an obvious and grave underestimation of association and explaining power.

Of many indices of association between two variables, η is a truly directional measure in the same manner as Somers delta (Somers 1962). As a directional measure, η produces two estimates of the association: one of the two variables is dependent and the other, then, is independent. In some cases, either direction may be valuable to study—both directions may be calculated to conclude which direction is more dominant (e.g., whether the attitude better explains the achievement or it is the other way around).² However, often only one direction interests the researcher; in GLM settings, a metric variable (X) is usually taken as the dependent variable and the grouping variable (g) is taken as independent. Then, this direction is often notated as $\eta(X|g)$ which direction is named as “ X dependent” in most software packages; the thinking and naming are opposite in measurement modelling settings. The peculiarity in naming and an alternative notation are discussed later.

² An anonymous reviewer pointed correctly that the arbitrary direction may become an issue in the GLM settings from the statistical inference viewpoint: “under the general linear modeling framework, it wouldn’t make sense for the categorical variable g to be the outcome of which X explains, as GLM assumes conditionally normally distributed outcomes”, that is, in all conditional divisions of variables X and Y ($Y|X$), we should obtain a normal distribution to meet the assumptions of GLM. However, it seems that the *technicalities* related to calculating eta and eta squared is not based on conditionally normally distributed outcomes. That is, we may calculate eta and eta squared without assuming any distribution in the variables or outcomes and still interpret the result meaningfully; this is a relevant case when two ordinal or interval-scaled variables are of interest. However, the *statistical inference* requires conditional normality. Conditional normality and related diagnostic analyses of the assumptions of the modelling have traditionally been taken crucial for statistical inference and not that much in the calculation of the statistics. However, obviously, statistical inference is an important matter to discuss, because usually we are not interested in the dataset as such but to generalize the result to the population. Because this article does not discuss the statistical inference of the coefficients, the further discussion of the issue related to the challenges in interpretation is left for the writings to come.

Traditionally, η varies from 0 to 1 where $\eta(X|g) = 0$ indicates the special case of no dispersion among the means of the metric variable X with ordinal, interval, or continuous scale in different categories in the categorical or ordinal variable g . The value $\eta(X|g) = 1$ indicates that each category in X is related to only one category in g , which presumes that there are no crossing pairs in the dataset. Technically, then, η and η^2 can reach the value 1. However, this necessitates that the “metric” variable has only as many categories as the independent variable (see [Appendix](#)). The interpretation is parallel with $\eta(g|X) = 0$ and $\eta(g|X) = 1$.

This article discusses a specific challenge related to coefficient η and its close relative, product–moment correlation coefficient (PMC; Bravais 1844; Galton 1889; Pearson 1896 onwards) often seen in the form of item–total correlation ($R_{it} = \rho_{gX}$) in the measurement modelling settings: artificial systematic attenuation or deflation in the estimates of correlation. In the empirical section, it is seen that the deflation in η may be as high as 0.50–0.60 units of correlation and in η^2 as high as 0.70–0.80 units of explaining power. A deflation of this size is worth discussing and taking seriously.

Attenuation is a statistical concept that, in general, refers to underestimating the correlation between two different measures because of measurement error (Lavrakas 2008). Pearson himself was the first to offer a solution to the problem (1903) and many solutions have been offered after that (see the discussion related to the concept of restriction of range as the reason for the attenuation in Mendoza and Mumford 1987; Sacket et al. 2007; Schmidt and Hunter 2015). This matter is discussed later in Section 4.3. If the attenuation is understood broadly, the estimates by the traditional estimators of correlation are also be radically *deflated* caused by *artificial systematic errors* during the estimation (see the discussion of the terms in, e.g., Chan 2008; Gadermann et al. 2012; Lavrakas 2008; Metsämuuronen 2022a). Underestimation of a magnitude of 0.50–0.70 units of correlation (see Section 4.2) is no more attenuation as we understand it in measurement modelling—or attenuation is only a minor part of this radical deflation. However, most probably, the grave deflation also includes attenuation caused by the difference in scales. Separating these two is difficult. Hence, the term “*mechanical* attenuation” is used to cover both random error and deflation. Even though the focus in this article is on deflation and its correction, both the terms mechanical attenuation and deflation are in use.

2 Research questions and the course of study

Condensing the previous discussion, the main research questions are:

1. What is the nature of relation between η and PMC? This matter is discussed, first, based on the literature related to point-biserial correlation and, second, based on empirical dataset related to point–polyserial correlation.

2. What is the nature, algebraic reasons, and magnitude of attenuation and deflation in η and η^2 ? Because of the algebraic identity of η and PMC, η must be artificially attenuated because PMC is known to be radically deflated. The algebraic reasons are discussed in the theoretical section and [Appendix](#), and the empirical section illustrates the magnitude.
3. How to correct the obvious deflation in PMC, η , and η^2 ? A simple procedure is suggested to assess the magnitude of the deflation and to correct it.
4. How to obtain also the negative values of η by using the traditional procedures? The negative values of η are easy to obtain with binary g , because PMC and η are identical in dichotomous cases. For the polytomous ordinal case, a new procedure is suggested.

The course of the article can be as follows: Although η and PMC use, apparently, different information in their calculation, first, the identity of their formulae is discussed in the case of the dichotomous independent variable. Second, it is shown empirically in the polytomous ordinal case how close PMC follows a *certain* direction of η , namely, η directed, so that “ X dependent” although, practically always, $|\text{PMC}| < |\eta(X|g)|$. This leads to a somewhat disturbing note that Pearson correlation which is usually thought to be a symmetric measure is, factually, a directional measure when the scales of two variables are not identical. Third, because of the identity of η and PMC, and because PMC is known to be deflated in an obvious manner when the scales of two variables are not equal, the magnitude of underestimation by η and η^2 is discussed and shown by examples. Fourth, a relevant option of correcting the attenuation in η and, consequently, in η^2 are discussed. Fifth, a form of η that also allows negative values is discussed. The empirical section gives numerical examples of the underestimation in η and η^2 , how to obtain negative values for η , and how the negative values of η are seen in non-linear association with convex and concave patterns.

3 Forms of η and different traditions of naming the directions

Above, the notation of η relevant in GLM settings was discussed. Here, an alternative thinking relevant to measurement modelling settings is discussed. In measurement modelling settings, the relevant direction is the one where the variable with a wider scale (often, a score variable X) explains the response pattern or order of the test-takers in the variable with a narrower scale (usually, a grouping variable such as sex or a test item g). This direction is relevant also in nonparametric methodology when analyzing the variables using Mann–Whitney U test (Mann and Whitney 1947) or Jonckheere–Terpstra JT test (Jonckheere 1954; Terpstra 1952) The thinking, terminology, and the notations of the directions vary between GLM settings and measurement modelling settings. This is briefly discussed below.

3.1 Different traditions in naming the directions

Assume a binary variable g and a metric variable X . Variable g could be, as an example, sex (0=female, 1=male or vice versa) or response in a binary test item (0=incorrect answer, 1=correct answer), and the metric variable may be a test score X . In the literature related to directional measures (e.g., IBM 2017; Metsämuuronen 2017; Newson 2002, 2006; Siegel and Castellan 1988) and, consequently, in the outputs of many generally known software packages such as IBM SPSS, SAS, as well as R libraries, the traditional direction of the analysis is the one where the categorical variable g is taken as an independent factor. This direction is called “ X dependent” and it is notated as $(X|g)$. This makes sense when η^2 is used in GLM settings where the metric variable (X) cannot explain the nominal variable, such as sex (g), and, hence, g must be an “independent” factor and, consequently, X must be a “dependent” factor.

The opposite thinking and naming of the same direction are relevant in settings related to measurement modelling and nonparametric testing. In what follows, this logic is used in the notation. In measurement modelling settings, the same factual direction as above means that the latent trait manifested as the score (X) explains the response pattern in the item (g), and the opposite direction does not make sense (e.g., Byrne 2016; Kim 1971; Metsämuuronen 2017; see the discussion of the directions with Table 3 and examples in Metsämuuronen 2020). Also, when analyzing the association of g and X with such procedures as Mann–Whitney U or Jonckheere–Terpstra JT tests, with manual calculation (see, e.g., Metsämuuronen 2017; Siegel and Castellan 1988), the subpopulations in g are first ordered by X after which the order is analyzed, that is, the *order* in g depends on X . In these settings, it is natural to think the relation of g and X from the condition’s viewpoint as “the pattern in g depends on X ”, that is, “ g dependent” or “ g given X ”, usually notated as $(g|X)$. Because of willingness to keep the discussion open for both interpretations, both viewpoints are kept equally relevant views in the text to interpret the outcome. From this on, the logic familiar from conditions is used in the notation: in what follows, $\eta(g|X)=\eta_{g|X}$ refers to “ η directed so that “ g given X ”” and $\eta_{g|X}^2$ refers to “ η^2 directed so that ‘ g given X ’” which would be labelled as “ X dependent” in generally known software packages. The alternative notation $\eta_{g|X}$ is preferred primarily for the cases that superscripts are used also as in $\eta_{g|X}^2$, $\eta_{g|X}^{Obs}$, or $\eta_{g|X}^{Max}$.

3.2 Calculation of coefficient η : a simple case with dichotomous categories

To lead to the general formula, let us first consider the simple case of a dichotomous g and a metric X . Assume a metric (ordinal, interval, or continuous³) variable X with observed values x_j and a binary variable g with observed values $y_j=0, 1$. The traditional direction of η in settings related to GLM (“ X dependent”) can be expressed as follows:

$$\eta(g|X) = \sqrt{\eta_{g|X}^2} = \sqrt{\frac{n_0(\bar{X}_{X0} - GM_X)^2 + n_1(\bar{X}_{X1} - GM_X)^2}{\sum_{j=1}^N (x_j - GM_X)^2}}, \tag{1}$$

where n_0 and n_1 refer to the number of cases in subpopulations 0 and 1 in g , \bar{X}_{X0} and \bar{X}_{X1} refer to the means of the variable X in these subpopulations, and GM_X is the grand mean of the variable X . If g is a test item with incorrect (0) and correct (1) responses, $\eta(g|X)$ indicates, in general, to what extent the higher score is related to the correct answer in the item. Notably, the interpretation for the same direction by Goodman–Kruskal gamma ($G=G(g|X)$; Goodman and Kruskal 1954; Metsämuuronen 2021a, b)⁴ and Somers’ delta ($D=D(g|X)$; Somers 1962; on the selected direction, see Metsämuuronen 2020) is known to correspond the proper direction in the measurement modelling settings. Then, consequently, it is known that $\eta(g|X)$ and related $\eta_{g|X}^2$ indicate to what extent the score variable X explains the behavior (or

³ The “continuous” nature of the metric variable X may be worth discussing, because it may become an issue in the direction selected for the analysis. In many cases, such as in econometrical, statistical, or engineering datasets, X may be a truly continuous variable. Then, the direction $\eta(X|g)$ does not make sense. However, in the contemporary settings of psychometrical testing, the score variable is very rarely totally continuous, but it is, factually, also a categorical one. To obtain a truly continuous scale would need an infinite number of items (or continuous scale in items which are very rare in the testing settings) and infinite sample size to form an infinite number of categories in the score. However, a raw score always forms a categorical (ordinal) score and, in the simplest case of a logistic transformation of the score by using item response theory (IRT) modelling, the number of categories in the score variable equals the raw score even though the “names” of the categories differ from the raw score numbers. Also, using factor analysis or related procedures, the number of categories is strictly dependent on the summed number of categories in the items; a test with ten Likert-scaled items with 5 categories means a maximum 50 categories in the scale regardless of the sample size. Then, factually, practically always, the score variable is a “categorical” variable with an unweighted or weighted ordinal scale in a sense that the names of the categories may look like they come from a continuous scale but the scale itself includes only limited number of categories that are in a weighted order. Hence, both directions in estimating eta and eta squared may make sense, because, in many cases, both directions are based on, factually, “categorical” variables. Notably, contrast to the polychoric correlation and truly non-parametric coefficients gamma and delta which, genuinely, transforms the weighted score into an ordinal scale, when calculating the values of eta and eta squared, no information of the scales is lost regardless of the direction taken.

⁴ Traditionally, G is taken as a symmetric coefficient, because it produces only one estimate in the same manner as PMC (e.g., IBM 2017; Sheskin 2011; Sirkin 2006; Wholey et al. 2015). However, Metsämuuronen (2021b) showed that, under certain conditions, $G = D(g|X)$ and not $D(\text{symmetric})$ or $D(X|g)$. Hence, G has a hidden directional nature: when the scales of two variables are not identical, G is unambiguously directional, so that the variable with a wider scale (X) explains the order in the variable with a narrower scale (g) the same manner as in $\eta(g|X)$. Then, a proper way to notate G may be $G(g|X)$.

the pattern) in variable g (and not the opposite way). From the GML viewpoint, $\eta(g|X)$ quantifies the proportional differences of the means of X between observations in different categories in g (e.g., males and females or between those who gave the correct and incorrect answer). From both viewpoints, the opposite direction is not meaningful. However, in some cases, both directions may be equally relevant.

3.3 Calculation of coefficient eta: a general case

Assume a nominal or ordinal variable g with observed values x_i with R categories, and a metric (ordinal, interval, or continuous) variable X with observed values y_i with C categories, and, often, $R << C$. The traditional direction of η in settings related to GLM (“ X dependent”) notated here as $\eta(g|X)$ can be expressed as square root of the sum of squares related to the difference between the means of X in the subpopulations in g ($SS_{between}$ or $SS_{treatment}$) divided by the sum of squares within the groups related to X (SS_{within} or SS_{total})

$$\eta(g|X) = \sqrt{\eta_{g|X}^2} = \sqrt{\frac{SS_{between}(g|X)}{SS_{total}(g|X)}} = \sqrt{\frac{\sum_{g=1}^R n_g (\bar{X}_{Xg} - GM_X)^2}{\sum_{i=1}^R \sum_{j=1}^C (x_{ij} - GM_X)^2}} \tag{2}$$

(e.g., Kerlinger 1964)⁵ where $\bar{X}_{Xg} = \sum_{j=1}^{n_g} \frac{y_j}{n_g}$ refers to the means of X in different categories in g , and

$GM_X = \sum_{g=1}^R n_g \bar{X}_{Xg} / \sum_{g=1}^R n_g$ is the grand mean of X . Correspondingly, η directed so that “ g is dependent” as in GLM settings or “ X given g ” as in settings related to conditions can be expressed as

$$\eta(X|g) = \sqrt{\eta_{X|g}^2} = \sqrt{\frac{SS_{between}(X|g)}{SS_{total}(X|g)}} = \sqrt{\frac{\sum_{X=1}^C n_X (\bar{X}_{gX} - GM_g)^2}{\sum_{i=1}^R \sum_{j=1}^C (y_{ij} - GM_g)^2}}, \tag{3}$$

⁵ Kennedy (1970, pp. 886–887) calls this the “classical formula” “proposed” by Kerlinger (1964). It seems a somewhat late proposal considering that Ayres (1920) already refers to the ratio of variances. On the other hand, Kennedy refers also to Peters and Van Voorhis (1940) while discussing epsilon squared, even though it was proposed 5 years earlier by Kelley (1935). Maybe, Kennedy just wanted to make the point by referring to the generally known textbooks while challenging Cohen’s (1965) and Friedman’s (1968) formulae and thinking of eta squared.

where $\bar{X}_{gX} = \sum_{i=1}^{n_X} \frac{y_i}{n_X}$ refers to the means of g in each category X and $GM_g = \sum_{X=1}^C n_X \bar{X}_{gX} / \sum_{X=1}^C n_X$ is the grand mean of g .

4 Relation of η and PMC and some consequences

4.1 Relation of η and PMC

In the case of the correlation between a dichotomous variable and metric variable, Eq. (1) can be expressed in a form

$$\eta(g|X) = (\bar{X}_{X1} - \bar{X}_{X0}) \frac{\sigma_g}{\sigma_X}, \tag{4}$$

where σ_g and σ_X are the standard deviation of g and X (see the algebraic proof in Wherry and Taylor 1946; Eikeland 1971). Notably, this form is identical with the simplified form for PMC found in textbooks for binary settings (e.g., Lane et al. 2016; Lord and Novick 1968; Metsämuuronen 2017), usually referred to as point-biserial correlation (ρ_{PB} ; Kuder 1937; Swineford 1936)⁶ or as item-total correlation (ρ_{gX}) for binary items in measurement modelling settings. Hence, with binary items or for example with dummy variables in GLM settings (see Cohen 1969), η directed so that “ g given X ” (in the measurement modelling settings) or “ X dependent” (in GLM settings) equals PMC

$$\eta(g|X) = \rho_{PB} = \rho_{gX} = \text{PMC}. \tag{5}$$

Notably, the opposite direction of η , $\eta(X|g)$ do not lead to the form of PMC.

Although the proof in the binary case is straightforward and simple, it is not trivial to derive—if not even possible—in the polytomous ordinal case, because, in general, PMC differs from $\eta(g|X)$ (see, however, formulae in Wherry and Taylor 1946). However, it is easy to show by empirical datasets that the closeness between PMC and $\eta(g|X)$ in Eq. (4) is a general one: with polytomous (ordinal) variables, the magnitude of the estimates by PMC is closer to $\eta(g|X)$ than to $\eta(X|g)$, however, such that always

$$|\text{PMC}| \leq |\eta(g|X)|. \tag{6}$$

This is caused by a small difference between the formulae of PMC and η (see Howell 2012). While the absolute value of PMC can be simplified as

⁶ It seems that Kuder (1937) was the first to use the term “point biserial correlation”. He also provided the computational forms. However, Swineford (1936) seems to be the first who studied the binary case of Pearson correlation in relation to biserial correlation (see Moses 2017).

Table 1 Selected characteristics of 14,880 test items

Range in the scale of g	$df(g) = R - 1$	Average PMC	Average $\eta_{g X}^1$	Average $\eta_{X g}^2$	$\eta_{g X}$ - PMC	Number of items
0-1	1	0.473	0.473	0.682	0.000	7948
0-2	2	0.612	0.625	0.755	0.013	3056
0-3	3	0.701	0.717	0.807	0.017	1390
0-4	4	0.766	0.784	0.843	0.018	729
0-5	5	0.809	0.828	0.873	0.019	474
0-6	6	0.848	0.865	0.897	0.017	366
0-7	7	0.878	0.893	0.912	0.016	255
0-8	8	0.901	0.916	0.933	0.015	140
0-9	9	0.916	0.929	0.940	0.014	160
0-10	10	0.929	0.939	0.947	0.010	136
0-11	11	0.941	0.950	0.956	0.010	93
0-12	12	0.945	0.954	0.959	0.009	67
0-15	13-14	0.944	0.950	0.956	0.006	66
Total		0.596	0.600	0.716	0.004	14,880

$$|PMC| = \sqrt{\frac{SS_{total} - SS_{residual}}{SS_{total}}} = \sqrt{\frac{\sum (x_{ij} - GM_X)^2 - \sum (x_{ij} - \hat{X}_{ij})^2}{\sum (x_{ij} - GM_X)^2}}, \tag{7}$$

the absolute value of η is

$$|\eta(g|X)| = \sqrt{\eta_{g|X}^2} = \sqrt{\frac{SS_{total} - SS_{error}}{SS_{total}}} = \sqrt{\frac{\sum (x_{ij} - GM_X)^2 - \sum (x_{ij} - \bar{X}_{Xg})^2}{\sum (x_{ij} - GM_X)^2}}. \tag{8}$$

If the association between two variables is perfectly linear, which in general is a rare special case but which is always the case with the binary and dichotomous g , the predicted value by the regression model (\hat{X}_{ij}) equals the means of X in the subpopulations of g (\bar{X}_{Xg}). In that specific case, $|\eta(g|X)| = |PMC|$, because $\sum (x_{ij} - \hat{X}_{ij})^2 = \sum (x_{ij} - \bar{X}_{Xg})^2$. In any other condition, that is, practically always with polytomous g , $\sum (x_{ij} - \hat{X}_{ij})^2 < \sum (x_{ij} - \bar{X}_{Xg})^2$, because the residuals related to the data in hand ($\sum (x_{ij} - \hat{X}_{ij})^2$) are constructed to be *minimum* and, hence, $\sum (x_{ij} - \bar{X}_{Xg})^2$ would always be greater than the minimum, causing

$$|\eta(g|X)| > |PMC|. \tag{9}$$

The factual difference between $\eta(g|X)$ and PMC may be subtle depending on the degree of non-linear nature of the association. In the empirical dataset related to scores (X) and ordinal items (g) in measurement modelling settings with finite or

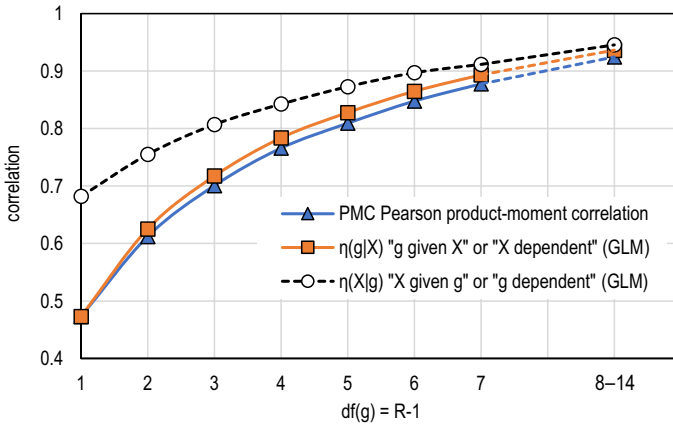


Fig. 2 Connection of PMC and eta in real-life datasets ($n = 14,880$ estimates)

small sample sized ($n \leq 200$) with 14,880 estimates of correlation between items and score variables,⁷ the difference between $\eta(g|X)$ and PMC appeared to be 0.007 units of correlation, while the difference between $\eta(X|g)$ and PMC appeared to be 0.159 units of correlation on average (Tables 1 and Fig. 2). The former difference seems to be relatively small. Hence, in what follows, the value of PMC is taken as a benchmark for the values of $\eta(g|X)$ also in polytomous (ordinal) cases.

- 1) $\eta_{g|X}$ = eta directed, so that “g given X” (as conditions) or “X dependent” (as in GLM)
- 2) $\eta_{X|g}$ = eta directed, so that “X given g” (as conditions) or “g dependent” (as in GLM)

4.2 Related result 1: Pearson correlation is a directional measure

Traditionally, PMC is taken as a symmetric measure, because it produces only one estimate for the association (e.g., Walk and Rupp 2010). A consequence related to Eqs. (4) and (5) with a binary and metric variable and empirical findings with polytomous case, the classic Pearson correlation is not a symmetric measure of association as has been traditionally thought. Instead, PMC has a hidden directional nature in the same direction as $G (=G(g|X)$; Metsämuuronen 2021b), such that the variable with a wider scale (X) explains the response pattern or order in the variable with a narrower scale (g) and not the opposite way nor symmetrically (or as “X dependent” as in GLM settings). Then, because $\eta(g|X) \neq \eta(X|g)$

⁷ See the characteristics and forming of the dataset in Metsämuuronen (2022a). The dataset is available at <http://dx.doi.org/10.13140/RG.2.2.17594.72641>.

$$\text{PMC} \cong \eta(g|X) \neq \eta(X|g). \quad (10)$$

This connection of PMC and η is not elaborated on here except to the extent needed for proposing a new type of attenuation correction to η and η^2 and a short-cut to reach the negative values of η in a polytomous g with an ordinal scale.

4.3 Related result 2: η and η^2 underestimate the association and explaining power in an obvious manner

The identity and relation of PMC and η make it clear why the estimates by η and η^2 must underestimate the true association and the true explaining power in an obvious manner. The attenuation in PMC and, consequently, in η and η^2 is artificial and systematic in nature and may be partly related to the phenomenon called restriction of range or range restriction (RR; see literature in Meade 2010; Sackett and Yang 2000; Sackett et al. 2007; Schmidt et al. 2008; Schmidt and Hunter 2015; Walk and Rupp 2010). Pearson himself was the first to offer a solution to the problem (1903) and many solutions have been offered to correct the attenuation in X variable (see the typology in Sackett et al. 2007; see also Mendoza and Mumford 1987). However, even if there is no manifestation of RR in X , PMC itself is very vulnerable to several sources of mechanical error in the estimates of correlation (MEC; see simulations in Metsämuuronen 2021a, 2022a). Notably, the reason for the deflation is not in the imperfect scales; if two variables both are measured by 5-point Likert scale reflecting continuous latent variables, but all the cases are in the diagonal of the crosstable, $\text{PMC}=1$ without error. The question is, why the correlation cannot be 1 if one scale is 5-point scale and the other 4- or 6-point scale? This is no more a matter of attenuation but caused by technical reasons.

In general, it is known that the number of categories, among others, influences the magnitude of the estimates by PMC (see simulations in Martin 1973, 1978; Olsson 1980). In practical terms, assume two identical continuous variables with obvious perfect correlation. Let one be dichotomized (g) and the other polytomized (X). Under this condition, PMC and, consequently, $\eta(g|X)$ cannot reach the (obvious) perfect (latent) correlation (see simulations in Metsämuuronen 2021a, 2022a; see the algebraic reasons in Metsämuuronen 2016; see also Appendix). As an example, let us take a vector of 1000 normally distributed cases, duplicate it, truncate the one version into a form of ordinal variable (g) with three categories (0, 1, 2; $df(g)=2$) and the other into a form of 21 ordinal categories (X). Let the proportion of the three categories be so that the 'difficulty level' is $p=0.15$, that is, most of the cases fall into the categories 0 or 1. In the case, the highest value of the observed PMC is around 0.745 and of the observed $\eta(g|X)=0.749$ even if the latent variables, obviously, correlated perfectly ($\rho=1$). Notably, such estimators of correlation as polychoric correlation (RPC) and G detect the perfect association and D almost perfectly ($D(g|X)=0.985$; see Fig. 3).

By modifying the example related to the obvious perfect latent correlation and binary g , Metsämuuronen (2021a, 2022a) showed that PMC and, consequently, η are sensitive to, at least, five sources of artificial systematic attenuation causing deflation in the estimates. First, if there is a discrepancy in the scales, in general, η tends to underestimate the true association between g and X . Second, η is sensitive to the

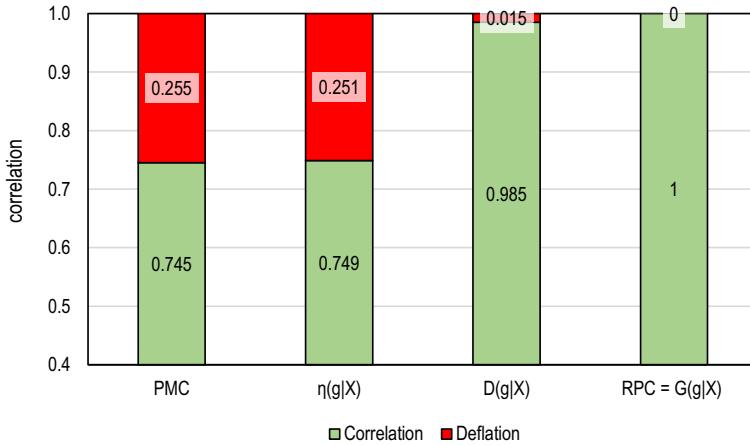


Fig. 3 Mechanical attenuation in eta and PMC; $df(g)=2$, $df(x)=20$, $p=0.15$, latent normality

number of categories in g ; the true association tends to be underestimated more the less categories there are in g . Third, η is sensitive to the distribution of the latent variable; the true association tends to be underestimated more when the distribution of the latent variable is normal or skewed than when it is even. Fourth, η is sensitive to the division of subpopulations in g (or item difficulty in the measurement modelling settings); true association is underestimated more, the more extreme is the division (or the difficulty level) in g . Fifth, η is sensitive to the number of categories in X ; the true association tends to be underestimated more the less categories there are in X .

Notably, $\eta(g|X)$ can reach the perfect correlation only in one specific theoretical case discussed above: that the variances of X in the subpopulations of g are equally zero. In the binary case, this can be inferred strictly from an alternative form of $\eta(g|X)$ (see Appendix)

$$\eta(g|X) = \frac{(\bar{X}_{X1} - GM_X)}{\sqrt{(1-p)\sigma_{X1}^2 + (1-p)\left(\frac{1}{p} - 1\right)\sigma_{X0}^2 + (\bar{X}_{X1} - GM_X)^2}}, \quad (11)$$

where σ_{X1}^2 and σ_{X0}^2 are the variances of X in the subpopulations 1 and 0. From Eq. (11), it is seen that $\eta(g|X)$ and, consequently, $\eta_{g|X}^2$ can reach the perfect 1 only in the theoretical case of $\sigma_{X1}^2 = \sigma_{X0}^2 = 0$.⁸ This condition implies that each category in X is connected to only one category in g without crossing pairs between the variables,

⁸ In this respect, both PMC and $\eta(g|X)$ require the condition of $\sigma_{Xi}^2 = \sigma_{Xj}^2 = 0$ to reach the value 1 (see Metsämuuronen 2016 and Appendix). This requires equal number of categories, because the condition necessary for $\sigma_{Xi}^2 = \sigma_{Xj}^2 = 0$ obtains only when the number of categories is equal; if there are more categories in X than in g , the variance of X in the categories of g would not be zero. The difference between PMC and $\eta(g|X)$ is that, in the polytomous case, the latter does not require any specific order to reach the perfect $\eta(g|X) = 1$ (see the examples of the attenuation-corrected eta in the section of Numeric examples).

that is, there are equal number of categories in g and X . This is, however, usually not true when η and η^2 are used in normal settings related to η^2 . In the binary case, the highest magnitude in Eq. (20) is achieved when $p=0.5$

$$\eta_{g|X}^{Max} = \frac{(\bar{X}_{X1} - GM_X)}{\sqrt{0.5 \times (\sigma_{X1}^2 + \sigma_{X0}^2) + (\bar{X}_{X1} - GM_X)^2}} \tag{12}$$

and, assuming symmetric distribution of X ,

$$\eta_{g|X}^{Max} = \frac{(\bar{X}_{X1} - GM_X)}{\sqrt{\sigma_{X1}^2 + (\bar{X}_{X1} - GM_X)^2}} = \frac{(\bar{X}_{X0} - GM_X)}{\sqrt{\sigma_{X0}^2 + (\bar{X}_{X0} - GM_X)^2}} \tag{13}$$

(see [Appendix](#)).

The numeric examples below show that the underestimation in η and η^2 may be notable. For instance, in deterministic patterns where G and D can detect (correctly) the extreme association ($G=D=1$), the maximum magnitudes of $\eta(g|X) = 0.701 - 0.769$ leading to $\eta_{g|X}^2 = 0.491 - 0.591$ (see [Tables 2 and 3](#) below in Section “[Numeric examples...](#)”) indicate directly how much η and η^2 underestimate the association and the explaining power. In terms of explaining power, around 50% of the information is lost, that is, η^2 cannot even reach more than 50–60% of the remaining variance in the other variable. With η^2 and η_p^2 as well as with ρ_p^2 , this leads us to conclude that the explaining power traditionally defined as “the proportion of remaining variance in the dependent variable” means, factually, “the proportion of remaining variance of which the coefficient can reach in the dependent variable”. This relates with the note by Hays (1963, p. 505; see also Richardson 1996) that the proportion of the total variation in the dependent variable is what *can* be predicted or explained based on its regression on the independent variable within the sample being studied. This phenomenon is discussed later with numerical examples.

The severity of mechanical attenuation or deflation in η and η^2 comes from the fact that, in the binary case, Eq. (11) implies that the loss of information in $\eta(g|X)$ and $\eta_{g|X}^2$ approximates 100% when the division of subpopulations (or, difficulty level) is extreme and, hence, the variance in g approximates zero irrespective of the fact that the true correlation between the variables may be perfect. This underestimation of association can be benchmarked when the behavior of directional coefficients such as D and G is known: these are capable of detecting the deterministic

patterns correctly (see Newson 2002; Metsämuuronen 2020, 2021a, b; see Fig. 3 and Tables 2 and 3 and the related discussion).⁹

4.4 Related result 3: need for attenuation correction to PMC and eta

To conclude the discussion of the underestimation in the estimates by η and η^2 , a possible new kind of attenuation correction in PMC, η , and η^2 is discussed. The traditional coefficients suggested to correct the inaccuracy of η^2 , that is, ϵ^2 , ω^2 , and η^2_{adj} were developed to correct the *positive* bias in η^2 , specifically, with values near zero (see Okada 2013, 2017; Mordkoff 2019). However, because PMC and η , in general, tend to be severely affected by mechanical error causing bias toward zero, it may be worth developing potential correction factors that may correct also the *negative* bias in η and η^2 .

Many corrections of attenuation for PMC are available (see a typology in Sacket and Yang 2000), although these are developed for different purposes than what is discussed here. Possible corrections for attenuation have been studied from Pearson (1903) and Spearman (1904) onward. The well-known corrections based on works of Pearson (1903; see also the notes by Aitken 1934 and Lawley 1943) and Thorndike (1949) are based on correcting the error when a restriction has occurred in one variable (usually in X). The idea is to enhance the concurrent validity of the test score of this restricted sample by altering it either by understanding or modelling the behavior of unrestricted population variance (see the mechanics in, e.g., Sacket and Yang 2000; Schmidt et al. 2008). This approach seems not the best option in correcting the deflation in the settings related to η . Another type of simple option for the correction of the mechanical attenuation is discussed below.

An essential characteristic of PMC and η is that, given the observed dataset, their maximum values are fixed. For PMC, because of the basic formula of PMC ($\rho_{gX} = \sigma_{gX} / \sigma_g \sigma_X$) and given the observed values of σ_g^2 and σ_X^2 , the only element affecting the magnitude of correlation is the covariance between g and X (σ_{gX}). The maximum value of σ_{gX} is obtained when g and X are in the *same order*. Hence, the maximal possible correlation (ρ_{gX}^{Max}) in the given set of variables is

$$\rho_{gX}^{Max} = \frac{\sigma_{gX}^{Max}}{\sigma_g \sigma_X} \tag{14}$$

⁹ Possible explanations for the obviously contradicting results between the earlier researchers of the *overestimation* by eta squared and the results of the radical *underestimation* discussed in this article may disturb the reader. The discrepancy could be explained by two distinct quantities estimated: eta squared may reflect the association between the observed measured variables, while the new estimator may reflect the association between latent continuous versions of these variables. To confirm this, we need more theoretical and empirical works. Another, simpler, explanation is that while PMC was originally created for the association of two continuous variables, the way we use it today with variables with different scales and specifically with binary and metric variables is not the intended environment for PMC and, hence, mechanical error in the estimates. The correction suggested in this article is *one* option to rectify the deflation.

(see Metsämuuronen 2022c). For $\eta(g|X)$, because of Eq. (8), the only element affecting the magnitude of the estimate is the sum of squared deviances of the score in the subsamples in g , $SS_{Error} = \sum (x_{ij} - \bar{X}_{Xg})^2$; the smaller is this element, the higher gets the value of $\eta(g|X)$. Given the number or proportions of cases in the subpopulations in g , the minimum value of $\sum (x_{ij} - \bar{X}_{Xg})^2$ is obtained when X is in order and all cases in a subpopulation in g are as close as possible to each other. One of these options is the condition that both g and X are in the same order irrespective of the nominal or ordinal nature of g . Hence, the same logic of finding the maximum correlation can be used with PMC and $\eta(g|X)$ irrespective of the nature of scale in $\eta(g|X)$.

A simple attenuation correction related to $\rho_{gX}(\rho_{AC}$, later R_{AC}) is the proportion of the observed correlation (ρ_{gX}^{Obs}) with ρ_{gX}^{Max} given the observed g and X

$$\rho_{AC} = \frac{\rho_{gX}^{Obs}}{\rho_{gX}^{Max}} = \frac{\rho_{gX}}{\rho_{gX}^{Max}} \tag{15}$$

(see Metsämuuronen 2022c). Similarly, the attenuation-corrected η (η_{AC} , later E_{AC}) is the proportion of observed eta ($\eta_{g|X}^{Obs}$) and the maximal value ($\eta_{g|X}^{Max}$), which, in the binary case, is the maximum value of PMC

$$\eta_{AC} = \frac{\eta_{g|X}^{Obs}}{\eta_{g|X}^{Max}} \left(= \frac{\eta_{g|X}}{\rho_{gX}^{Max}} \right). \tag{16}$$

Consequently, attenuation-corrected η^2 (η_{AC}^2 , later $E2_{AC}$) is

$$\eta_{AC}^2 = \left(\frac{\eta_{g|X}^{Obs}}{\eta_{g|X}^{Max}} \right)^2 \left(= \left(\frac{\eta_{g|X}}{\rho_{gX}^{Max}} \right)^2 \right). \tag{17}$$

In polytomous ordinal g , the latter part of formulae (16) and (17) leads to a mild overestimation in practical settings, because the absolute magnitude of the estimates by PMC is somewhat lower than the ones by $\eta_{g|X}$ (see Eq. 9). However, in empirical datasets, the difference between $\eta_{g|X}$ and PMC may be nominal (see above, 0.007 units of correlation), while the attenuation may raise as high as 0.13 to 0.99 units of correlation depending on the characteristics of g (see Metsämuuronen, 2021a, 2022a). Obviously, using ρ_{gX}^{Max} as a benchmark would not make sense with genuinely nominal-scaled polytomous g . Numeric examples of R_{AC} , E_{AC} , and $E2_{AC}$ are given later.

Equations (15) and (16) seem to solve all five challenges of PMC and η discussed above: (1) The general characteristic of PMC and η of being artificially attenuated is solved; R_{AC} and E_{AC} can reach the extremes of correlation also when the number of

categories of the variables is not equal. (2) The effect of low variance in g caused by extreme division of cases of subpopulations (or item difficulty) in g is solved; R_{AC} and E_{AC} can reach the perfect 1 irrespective of the difficulty and variance in g and, by solving this challenge, the radical deflation related to items with extreme difficult level is solved. (3) The effect of the number of categories in g is solved; R_{AC} and E_{AC} can reach value 1 irrespective of the number of categories in g . (4) The effect of small number of categories in X is solved, R_{AC} and E_{AC} can reach value 1 irrespective of the number of categories in X . (5) The latent variable is not a challenge; R_{AC} and E_{AC} can reach value 1 irrespective of the form of the variable latent to X and g .

This article does not study further the characteristics of R_{AC} , E_{AC} , and E_{2AC} ; simulations of their behavior would be beneficial (see comparison of R_{AC} , and E_{AC} in Metsämuuronen 2022a). However, these estimators reach the value 1 when the maximum possible value of ρ_{gX} and $\eta_{g|X}$ is achieved, that is, when the item and score are in the same order. Value 0 is obtained when the observed correlation is 0. R_{AC} and E_{AC} also can reach negative values; because the maximum possible value is always positive, the value of R_{AC} is negative when the observed ρ_{gX} is negative (see the next section). Hence, unlike PMC and η , R_{AC} and E_{AC} reach the limits of correlation ($-1 \leq R_{AC}, E_{AC} \leq +1$) also when the number of categories in two variables is not equal.

4.5 Related result 4: η can reach negative values

Traditionally, η does not reach negative values. This is caused by the traditional formulae based on squares or variances which are always positive. However, in the form in Eq. (4), η can reach negative values; if the mean of X in subpopulation 1 in g is lower than in subpopulation 0, ($\bar{X}_{X1} < \bar{X}_{X0}$), the factual correlation is negative, although the traditional way of calculating η will lead to signal (falsely) a positive association. Then, the value of η we usually see is, in fact, the *absolute value* of the estimate of the association between two variables. Hence, coefficient η calculated by using the traditional formulae gives us just the *magnitude* of the association and not necessarily the true association. As a parallelism, the traditional way of thinking and calculating η seems to follow the same logic as with the coefficient phi (φ) originally suggested as “mean square contingency” by Pearson (1904, p. 6) for a 2×2 cross-table: $\varphi^2 = \frac{\chi^2}{N} \Rightarrow \varphi = \sqrt{\frac{\chi^2}{N}}$. The original definition leads us to the positive values, even though the factual association could be negative. The negative values are found by going back to the basic elements in the calculation: $\varphi = \frac{AD-BC}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}$ (Yule 1912; see Glen 2016; see also Gibbons 1993).

Hence, to obtain the true value of eta, Eq. (4) can be used in the binary or dichotomous case. For the polytomous ordinal case, because of the close relation between PMC and $\eta(g|X)$, we could use the sign of PMC to indicate both the direction and magnitude of η

$$\eta(g|X) = \text{sign}(PMC) \times |\eta(g|X)|. \quad (18)$$

Numeric examples will clarify these results.

5 Numeric examples of artificial mechanical attenuation and negative values in eta and eta squared

5.1 Examples of attenuation in η and η^2 and related attenuation corrections

As a simple numeric example of R_{AC} , E_{AC} , and $E2_{AC}$, assume a test with a score X and five items ($g1$ – $g5$) as part of a longer test (or five conditions of different proportions of males and females in an independent variable) with descendent levels of proportion of 1 s (p_i) in g_i as in Table 2. Two of the items ($g1$ and $g5$) discriminate the lower performing test takers from the higher performing ones (or males from females) in a deterministic manner. This is indicated as the perfect correlation by G and D in Table 2, and it would be obtained also using polychoric correlation coefficient albeit asymptotically.

Notably, an apparent challenge with the extreme items $g1$ and $g5$ is that the extreme division of the responses in g causes that only the extreme values in X are reached and, hence, this may cause the reduced eta and eta squared. However, if letting all 1 s in $g5$, as an example, be related to every value in X (10 different categories) but keeping the proportion of 0 s and 1 s the same as it was (8.3% of the cases being 1s) we would have a dataset with 146 test takers with 134 0s and we would obtain almost zero correlation between variables $g5$ and X ; this is easy to confirm by forming such a dataset. That is, the reduction in correlation has not to do with non-extreme X scores and the problem remains—if not exacerbates—with a larger sample size. Because of the deflation, neither ρ_{gX}^{Obs} and η_{gX}^{Obs} cannot reach the perfect correlation, and the maximal possible correlations ρ_{gX}^{Max} and η_{gX}^{Max} differ item-wise. In items $g2$ – $g4$, the patterns include stochastic error to different extent.

The right-hand side of Table 2 shows the pattern in single items leading to maximal correlations; with each g , both g and X are in the same order leading to maximal covariance.

Three points are highlighted based on Table 2. First, with deterministic patters, R_{AC} , E_{AC} , and $E2_{AC}$ can reach the extreme value ($R_{AC}=E_{AC}=E2_{AC}=1$) in the same manner as G and D do. Second, in comparison with R_{AC} and E_{AC} , the deflation in PMC and η varies 0.03–0.56 units of correlation. The latter indicates a notable deflation in correlation, and it has a strict effect on the deflation in η^2 : 0.03–0.80 units of explaining power. In the case of $g1$, η^2 was able to reach only 20% of the total variance in the score because of deflation. Third, the loss of information in η^2 is not symmetric; based on Table 2, we see somewhat more loss of information when the proportion of 1s is extremely high than when it is extremely low. These need in-depth studies when exploring the boundaries of R_{AC} , E_{AC} , and $E2_{AC}$.

Table 2 Hypothetic binary dataset with the descendent proportion of 1 s

ID	Observed pattern					Score (X)	Pattern for maximal correlation				
	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5		ξ_1	ξ_2	ξ_3	ξ_4	ξ_5
1	0	1	0	0	0	5	0	0	0	0	0
2	1	0	1	0	0	6	1	0	0	0	0
3	1	1	0	0	0	7	1	0	0	0	0
4	1	0	1	0	0	8	1	1	0	0	0
5	1	1	0	0	0	8	1	1	0	0	0
6	1	1	0	0	0	8	1	1	0	0	0
7	1	0	0	1	0	11	1	1	1	0	0
8	1	1	1	0	0	12	1	1	1	0	0
9	1	1	1	0	0	13	1	1	1	0	0
10	1	1	1	0	0	14	1	1	1	1	0
11	1	1	0	1	0	15	1	1	1	1	0
12	1	1	1	1	1	16	1	1	1	1	1
p_i	0.917	0.750	0.500	0.250	0.083		0.917	0.750	0.500	0.250	0.083
$\sigma_{gi}^2 = p_i(1 - p_i)$	0.076	0.188	0.250	0.188	0.076		0.076	0.188	0.250	0.188	0.076
σ_{gX}^{Obs}	0.438	0.479	0.625	0.938	0.479	σ_{gX}^{Max}					
$\rho_{gX}^{Obs} = \rho_{gX} = \eta_{g X}$	0.444	0.311	0.351	0.608	0.487	ρ_{gX}^{Max}	0.125	0.208	0.333	0.292	0.208
$\eta_{g X}^2$	0.197	0.097	0.123	0.369	0.237	η_{gX}^{Max}	0.444	0.689	0.912	0.770	0.487
$\rho_{gX}^{Max} = \eta_{gX}^{Max}$	0.444	0.689	0.912	0.770	0.487		0.444	0.689	0.912	0.770	0.487
$G = G(g X)$	1	0.440	0.412	0.778	1						
$D(g X)$	1	0.407	0.389	0.778	1						
$\rho_{AC} = \eta_{AC} = \frac{\rho_{gX}}{\rho_{gX}^{Max}}$	1	0.451	0.385	0.789	1						
$\rho_{AC}^2 = \eta_{AC}^2 = \left(\frac{\rho_{gX}}{\rho_{gX}^{Max}} \right)^2$	1	0.203	0.148	0.623	1						

Table 2 (continued)

ID	Observed pattern					Pattern for maximal correlation					
	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	Score (X)	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5
$\eta_{AC} - \eta_{g X}$	0.556	0.140	0.034	0.182	0.513						
$\eta_{AC}^2 - \eta_{g X}^2$	0.803	0.107	0.025	0.254	0.763						

5.2 Simple example of detecting negative values of eta

As a simple numeric example of the difference between the traditional formulae (Eqs. 2 and 3) and the better formula of η (Eqs. 4 and 18), let us consider a set of variables with the deterministic association as in Table 3. Of the two binary variables A1 and A2, one correlates positively with X, while the other correlates negatively. Similarly, of the two polytomous variables B1 and B2, one correlates positively with X, while the other correlates negatively. Notably, because of deterministic patterns, the correlations in variables A1 and B1 are the maximum possible

Table 3 Negative and positive eta in a hypothetic dataset with deterministic patterns

ID	A1	A2	B1	B2	X
1	0	1	0	5	10
2	0	1	0	4	11
3	0	1	0	3	12
4	0	1	0	2	13
5	0	1	0	1	14
6	0	0	0	0	15
7	0	0	0	0	15
8	1	0	1	0	17
9	1	0	2	0	18
10	1	0	3	0	19
11	1	0	4	0	20
12	1	0	5	0	21
PMC	0.878	-0.837	0.886	-0.851	
$\eta(g X)$ Eq. (2)	0.878	0.837	0.917	0.878	
$\eta(X g)$ Eq. (3)	1	1	1	1	
\bar{X}_{X0}	12.9	17.9			
\bar{X}_{X1}	19	12			
σ_g	0.493	0.493	1.738	1.738	
σ_X	3.451	3.451	3.451	3.451	
ρ_{gX} Eq. (12)	0.878	-0.837	0.886	-0.851	
$\eta(g X)$ Eqs. (4), (18)	0.878	-0.837	0.917	-0.878	
ρ_{gX}^{Max}	0.878	0.878	0.886	0.886	
η_{gX}^{Max}	0.878	0.878	0.917	0.917	
$G=D(g X)$	1	-1	1	-1	
ρ_{AC}	1	-0.953	1	-0.961	
η_{AC}	1	-0.953	1	-0.957	
$\eta_{g X}^2$	0.770	0.700	0.841	0.771	
ρ_{gX}^2	0.770	0.700	0.784	0.724	
$G^2=D^2$	1	1	1	1	
ρ_{AC}^2	1	0.909	1	0.923	
η_{AC}^2	1	0.909	1	0.917	

values that ρ_{gX} and $\eta(g|X)$ can reach given the observed g and X because g_i and X are in the same order. The scale in B1 and B2 may be ordinal or nominal.

In the binary case with positive association (A1), $\eta(g|X)$ acquired both in the traditional (Eq. 2) and with the enhanced formula (Eq. 4) equals PMC. The difference is with negative association (A2) where PMC and Eq. (14) reach the real correlation (-0.837), while the traditional formula gives us an absolute value ($+0.837$). The outcome is obvious because of Eq. (5): in the binary case, $PMC = \rho_{PB} = \rho_{gX} \equiv \eta(g|X)$. The same is seen also with polytomous items, however, so that Eq. (4) cannot be used because of a lack of proper formula; $PMC = -0.851$ while $|\eta(g|X)| = +0.878$ and $|PMC| < |\eta(g|X)|$, as known from Eq. (6). In this case, however, we could use Eq. (14): the sign of PMC indicates the direction of η : $\eta(g|X) = \text{sign}(PMC) \times |\eta(g|X)|$, that is, $\eta(g|X) = -0.878$.

Notably, in Table 3, in comparison with estimators that *can* detect the deterministic patterns and strictly indicate the proportion of the logically ordered cases in g after they are ordered by X (here, D and G), $\eta(g|X)$ underestimates the association and the explaining power in an obvious manner ($\eta_{g|X}^2 = 0.70 - 0.84$). From this viewpoint, the attenuation-corrected explaining powers by $\rho_{AC}^2 = \eta_{AC}^2 = 1$ as well as $G^2 = D^2 = 1$ intuitively feel to be an interesting subject to study more. Notably also, the estimates by $\eta(X|g)$ do not refer to the characteristics of g but to those of X . Because of no changes in X and because, in all items, the two tied scorers with $X = 15$ are both from the subpopulation 0 in all items and, hence, no crossing observations between the categories, the values are intact even if the patterns in g differ item-wise.

Until we have more accurate forms of η for a polytomous ordinal case, because of the algebraic connection of PMC and $\eta(g|X)$ (see Eqs. 4 and 5), a possible short-cut for the sign of η is to use the sign of PMC obtained for the same variables (Eq. 18). Hence, η would be estimated in the traditional way but, on the side, PMC also is calculated, and the sign of PMC is given to η to point the direction of the association. This may serve as an intermediate solution to reach the real η in the case that g has an ordinal scale; with truly nominal polytomous g , PMC does not have a meaningful interpretation. Some limits of this option are studied in what follows.

5.3 Examples with convex and concave patterns

Being used as a measure for non-linear association, the behavior of η is also illustrated with convex and concave patterns based on 5-point ordinal variables. The convex pattern *could* lead us to obtain η with a positive sign, while the concave pattern would lead to η with a negative sign. This is not, however, true if we use the sign of PMC as an indicator of the sign of η as seen in Table 4 and a set of illustrations in Fig. 4.

Table 4 includes three sets of polytomous ordinal variables ($g1$ to $g5$) with a common X . Variables $g1$ and $g2$ represent symmetric convex and concave situations where $PMC = G = D = 0$. Variables $g3$ and $g4$ represent non-symmetric convex and concave situations where $|PMC|$ is identical although opposite in direction. Variable $g5$ represents convex patterns where PMC has a negative sign.

Table 4 Negative and positive values of eta in hypothetic convex and concave patterns

ID	g1	g2	g3	g4	g5	Score X
1	1	5	3	3	5	20
2	1	5	3	3	5	20
3	1	5	3	3	5	21
4	2	4	4	2	3	21
5	2	4	4	2	3	22
6	3	3	5	1	1	22
7	3	3	5	1	1	23
8	4	2	4	2	1	23
9	4	2	4	2	1	24
10	5	1	3	3	2	24
11	5	1	3	3	2	25
12	4	2	2	4	2	25
13	4	2	2	4	2	26
14	3	3	1	5	3	26
15	3	3	1	5	3	27
16	2	4	1	5	3	27
17	2	4	1	5	3	28
18	1	5	1	5	3	28
19	1	5	1	5	3	29
20	1	5	1	5	3	29
ρ_{gX}	0	0	-0.787	0.787	-0.212	
ρ_{gX}^2	0	0	0.620	0.620	0.045	
$\eta(g X)$ Eq. (2)	0	0	0.886	0.886	0.748	
$\eta(X g)$ Eq. (3)	0.944	0.944	0.961	0.961	0.913	
$G(g X)$	0	0	-0.724	0.724	0.038	
$D(g X)$	0	0	-0.583	0.583	0.036	
$f''(X)$	-0.333	0.333	-0.110	0.110	0.246	

The first thing to note from Table 4 is that high values of η do not indicate that η detects the curvilinear pattern better than other coefficients but, instead, the fact that η is sensitive to the small number of crossing pairs between variables. The same magnitudes of η would be obtained if the pattern would have been strictly linear, but the pattern of paired cases would be identical. Second, to simplify the assessment of convexity of the pattern, curvilinearity is indicated by a function with second power—notably, a function with third power would fit better to the patterns (compare graphs for $g5$ in Fig. 4). From the function of the second power, convexity is easy to note as the positive sign of the second derivate of function ($f''(X) > 0$), while the concave pattern gives a negative sign ($f''(X) < 0$). Notably, the complexity comes from the fact that the pattern in variable $g5$, as an example, is both convex and concave; convex when $X=1-6$ and concave when $X=7-10$ (see the right-hand side graph related to $g5$ in Fig. 4). From this viewpoint, the second derivate of $g5$

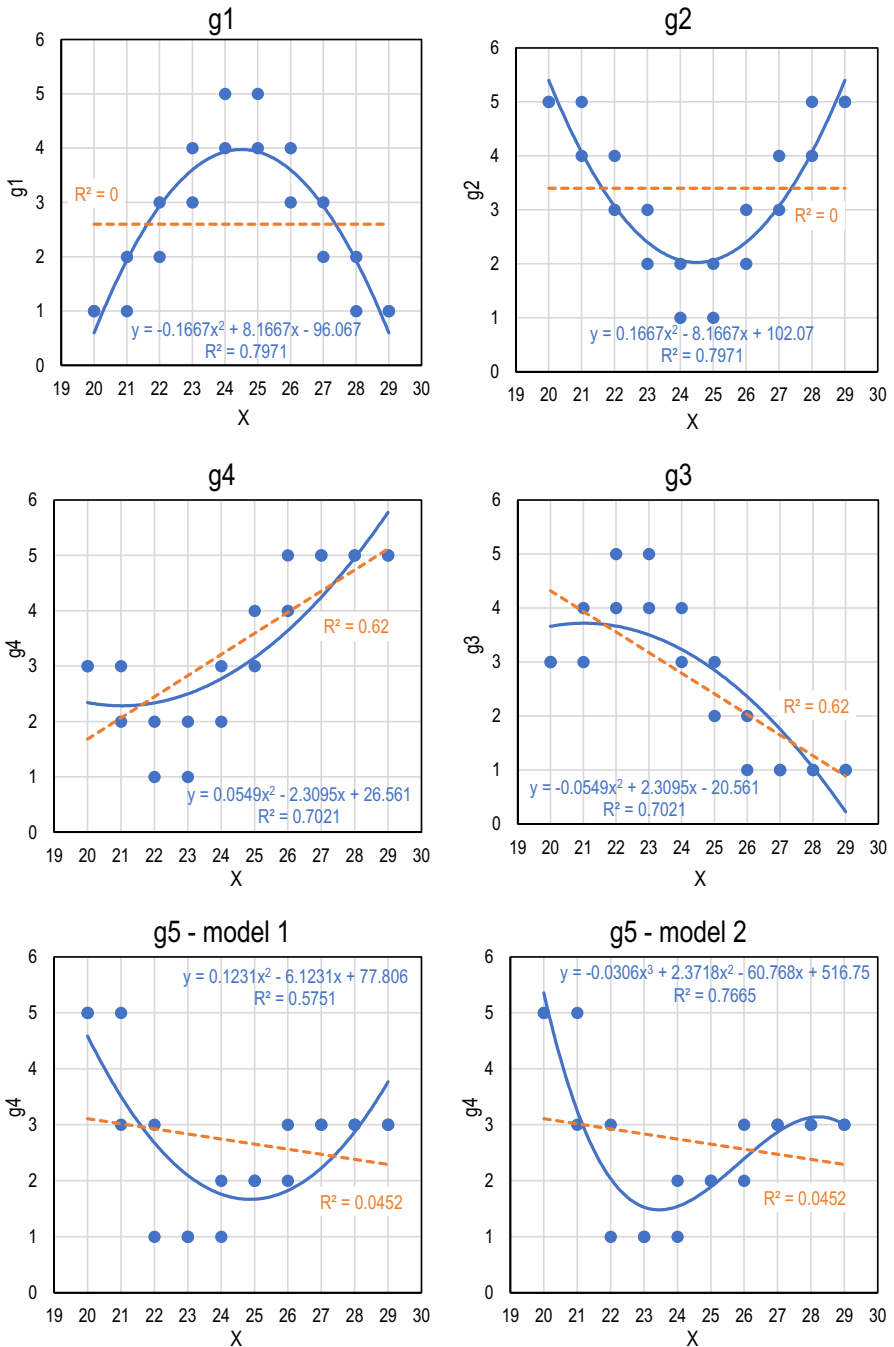


Fig. 4 Convex and concave patterns

is worth noting: while $f''(X) = 2 \times 0.1231 = 0.2462 > 0$ indicates convex patterns, PMC indicates negative correlation that seems to point to a negative $\eta(g|X)$. This indicates that the possible negative or positive sign of η is not an indicator of convexity or concavity in the pattern.

In the graphs in Fig. 4, the explaining powers (ρ_{gX}^2) are calculated for PMC and curvilinear association based on residuals related to the non-linear prediction. Hence, η^2 is not seen in the graphs. Notably, in the case of symmetric convex and concave patterns where PMC equals 0 (items $g1$ and $g2$), the short-cut method cannot be used to indicate whether η should have a positive or negative sign. In the case of non-symmetric patterns ($g3$ and $g4$), the sign of PMC may indicate the sign of η properly (based on the close connection of the forms in Eqs. 4 and 5).

6 Conclusions and limitations

6.1 General discussion

The main result concerning coefficients η and η^2 is that their values are artificially and systematically attenuated in an obvious manner. This is known from the identity of $\eta(g|X)$ and PMC, of which the latter tends to include notable mechanical error in estimates of correlation leading to mechanical attenuation or deflation when the number of categories in the variables is not equal. The magnitude of deflation in $\eta(g|X)$ and PMC may be remarkable, specifically when the division of cases in sub-populations on g or difficulty level in g is extreme.

Because of the obvious deflation, it is worth noting the basic deficiency in $\eta_{g|X}^2$ as an indicator of explaining power in practical settings related to GLM. In terms of explaining power, in some cases, as much as 50%—or even more—of the information may get lost, that is, η^2 cannot even reach more than 50–20% of the remaining variance in the other variable. With η^2 and η_p^2 as well as with ρ_p^2 , this leads us to conclude that the attribute for the explaining power, “the proportion of remaining variance in the dependent variable” should be rephrased as “the proportion of remaining variance of which the coefficient can reach in the dependent variable”.

The identity of PMC and η evokes the need of deflation correction for PMC and η . The issue is not necessarily relevant when the true association is near zero; in these settings, the bias-corrected estimators ε^2 , ω^2 , and η_{adj}^2 are developed to correct the *positive* bias in η^2 . However, this article discussed the opposite challenge in η and η^2 : the radical *negative* bias. To combine these obviously contradicting views seems interesting and worth studying more. Logically, because η is connected to PMC, it always underestimates the correlation—even near zero—and this is a strict characteristic of PMC and η and this has nothing to do with the proposed attenuation-corrected η . A possible direction to go to seek the answer to this practical challenge may be related to the traditional practicality in estimation of *not to consider the negative values* as real values; near zero, the factual values of eta may be negative ones, and this squared make them always positive which, apparently, is seen as an apparent or real overestimation in explaining power. Simulations with the

near-zero correlations from a viewpoint of negative values may enrich our knowledge of the phenomenon.

A simple option to correct attenuation-corrected coefficients for dichotomous and ordinal-scaled g by proportioning the observed estimates by PMC and $\eta(g|X)$ with the maximum possible estimates given the dataset. This kind of correction, leading to attenuation-corrected ρ_{gX} ($R_{AC} = \rho_{gX}^{Obs} / \rho_{gX}^{Max}$) and attenuation-corrected $\eta(g|X)$ ($E_{AC} = \eta_{g|X}^{Obs} / \eta_{g|X}^{Max}$) of which the latter leads to attenuation-corrected $\eta_{g|X}^2$ ($E2_{AC} = (\eta_{g|X}^{Obs} / \eta_{g|X}^{Max})^2$), seems to solve most of the challenges related to attenuation and deflation in PMC, $\eta(g|X)$, and $\eta_{g|X}^2$. R_{AC} and E_{AC} can reach the extremes of correlation ± 1 and $E2_{AC}$ can reach values 0–1 even when the categories in the items differ from each other in an obvious manner and irrespective of the division of cases in sub-populations of g or difficulty level in g . In some cases, the attenuation may be corrected by 0.20–0.50 units of correlation or even more. The article did not discuss the characteristics of R_{AC} , E_{AC} , and $E2_{AC}$ in-depth and simulations of their behavior is needed (see comparisons in Metsämuuronen 2022a). Notably, the attenuation correction to η and η^2 presented in the article also fits with nominal-scaled g .

6.2 Some practical possibilities to use RAC and EAC

The characteristics of R_{AC} and E_{AC} were not studied in-depth, although some limits were discussed. More studies in this respect would be beneficial. However, if R_{AC} and E_{AC} and the related maximal Rit and eta are found stable and useful tools in evaluating and correcting attenuation in correlations, the maximal possible Rit and eta given the observed dataset—if not R_{AC} or E_{AC} —may be worth considering reporting routinely as a related statistic with the observed correlation. In case the widths of the scales differ from each other in an obvious manner, this may help assess the magnitude of possible deflation in the observed estimates. Maybe, R_{AC} could be considered when choosing the correction formulae for the r^2 effect sizes (see, e.g., Skidmore and Thompson 2011; Vacha-Haase and Thompson 2004).

R_{AC} and E_{AC} have two specific applications in measurement modelling settings where the correlation between an item with a narrower scale and a score with a wider scale is of interest and where Rit and η always underestimate the true correlation. First, Rit has been classically used as an estimator of item discrimination power (see Swinford 1936; Kuder 1937; Moses 2017) and η also could be used; η would react more efficiently to the non-linearity in the association. Because Rit and η underestimate the true correlation between an item and a score in an obvious manner, R_{AC} and E_{AC} could be used instead to reflect closer the true association. More studies in this regard could be beneficial.

Second, on one hand, Rit is embedded in the traditional estimators of reliability of the score such as coefficient alpha (Cronbach 1951) based in Rit , Armor's theta (Armor 1973) based on principal component loadings, McDonald's omega total (McDonald 1970), and rho or maximal reliability (e.g., Li 1997; Li et al. 1996; Raykov 2004), both based on factor loadings. Notably, the principal component or factor loadings (λ_i) are, essentially, correlations between items and the score variable, that is, their essence is

Rit, see, e.g., Yang 2010). On the other hand, empirical results show that, using the traditional estimators of reliability, the reliability may be underestimated as much as 0.40–0.60 units of reliability (see Gadermann et al. 2012; Metsämuuronen 2022b, c; Metsämuuronen and Ukkola 2019) because of attenuation in PMC. Hence, Metsämuuronen (2022b, c) discusses attenuation-corrected estimators of or reliability (ACER) by changing *Rit* or principal component- or factor loading in the estimators by attenuation-corrected estimate of *Rit* (or attenuation-corrected λ_i). Then, the attenuation-corrected coefficient alpha based on R_{AC} would be

$$\rho_{\alpha_RAC\theta} = \alpha_{RAC} = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_{gi}^2}{\left(\sum_{i=1}^k \sigma_{gi} \times RAC_{i\theta} \right)^2} \right) \tag{19}$$

and based on E_{AC}

$$\rho_{\alpha_EAC\theta} = \alpha_{EAC} = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_{gi}^2}{\left(\sum_{i=1}^k \sigma_{gi} \times EAC_{i\theta} \right)^2} \right), \tag{20}$$

where k is the number of items in the test, σ_{gi}^2 refers to item variances, and $RAC_{i\theta}$ and $EAC_{i\theta}$ are attenuation-corrected correlations between the item i and the score variable θ . Similarly, attenuation-corrected theta based on R_{AC} would be

$$\rho_{TH_RAC\theta} = \frac{k}{k-1} \left(1 - \frac{1}{\sum_{i=1}^k RAC_{i\theta}^2} \right) \tag{21}$$

and based on E_{AC}

$$\rho_{TH_EAC\theta} = \frac{k}{k-1} \left(1 - \frac{1}{\sum_{i=1}^k EAC_{i\theta}^2} \right), \tag{22}$$

(22) where $RAC_{i\theta}^2$ and $EAC_{i\theta}^2$ are attenuation corrected principal component loadings. Correspondingly, attenuation-corrected omega total based on R_{AC} would be

$$\rho_{\omega_RAC\theta} = \frac{\left(\sum_{i=1}^k RAC_{i\theta}\right)^2}{\left(\sum_{i=1}^k RAC_{i\theta}\right)^2 + \sum_{g=1}^k (1 - RAC_{i\theta}^2)}, \tag{23}$$

and based on E_{AC}

$$\rho_{\omega_EAC\theta} = \frac{\left(\sum_{i=1}^k EAC_{i\theta}\right)^2}{\left(\sum_{i=1}^k EAC_{i\theta}\right)^2 + \sum_{g=1}^k (1 - EAC_{i\theta}^2)}, \tag{24}$$

and attenuation-corrected maximal reliability (ρ_{MAX}) based on R_{AC} would be

$$\rho_{MAX_RAC\theta} = \frac{1}{1 + \frac{1}{\sum_{i=1}^k (RAC_{i\theta}^2 / (1 - RAC_{i\theta}^2))}}, \tag{25}$$

and based on E_{AC}

$$\rho_{MAX_EAC\theta} = \frac{1}{1 + \frac{1}{\sum_{i=1}^k (EAC_{i\theta}^2 / (1 - EAC_{i\theta}^2))}}, \tag{26}$$

where $RAC_{i\theta}^2$ and $EAC_{i\theta}^2$ are attenuation-corrected factor loadings. The characteristics of the estimators 19 to 26 are not studied here (see closer Metsämuuronen 2022b, c). However, it is known that, except for binary items where $R_{AC} = E_{AC}$, the estimates based on E_{AC} would give somewhat higher estimates than those by R_{AC} . Also, it is known that theta maximizes alpha (e.g., Greene and Carmines 1980) and rho maximizes omega (e.g., Cheng et al. 2012) and that rho tends to overestimate reliability with small sample sizes (Aquirre-Urreta et al. 2019). Hence, using R_{AC} with the alpha formula (Eq. 19) would give us a more conservative estimate of the reliability than the other estimators. Correspondingly, using E_{AC} with the form of maximal reliability (Eq. 26) may lead to overestimate reliability with small sample sizes. In-depth studies of the estimators would enrich the discussion. It is good to note also the critical note by Chalmers (2017) that if the estimator of correlation does not refer to the observed variables as *RPC* does not, the estimator of reliability does not refer to the observed score. From this viewpoint, estimators based on R_{AC} and E_{AC} refer to the observed score the same manner as *G* and *D* do. Hence, their use could be justified in estimators or reliability (see Metsämuuronen 2022c).

6.3 Some possibilities to use the negative values of η

One of the results was that coefficient η can indicate a negative association between two variables, although the traditional ways of calculating the estimates cannot detect this. This has its own value in extending our knowledge of the traditional coefficients of association. We may relevantly ask: Would it be valuable to start using a similar kind of “back to the basic elements” type of version of coefficient η as we have for the coefficient φ to indicate both the tendency and the magnitude of the association between variables and not just the magnitude? For a binary case, we can use Eq. (4), and parallel version(s) for the polytomous cases could be derived. While waiting for the derivation, Eq. (18) could be used as a shortcut to the direction of the association. Alternatively, a relevant question for the binary case is whether it would be easier, at first hand, to use PMC or, more precisely, the point-biserial correlation coefficient (Eq. 4) instead of η to obtain strictly both the magnitude as well as direction of the association?

The negative values in η may expand the usefulness of this coefficient as a descriptive statistic. First, generally, η gives a more credible estimate of association in comparison with PMC in the case of curvilinear association with ordinal g . Second, the enhanced η could be used to indicate strictly whether the curvilinear trend between the variables is, overall, ascending or descending. However, although the enhanced η tells the overall trend with higher accuracy than PMC under the condition of non-linearity in the association, it does not tell whether the pattern of association is convex or concave. Also, η seems not to reach the same accuracy of the fit as do the methods based on residuals related to the predicted values. Third, η has been used in measurement modelling settings rarely. Maybe the possibility of detecting curvilinearity in the item responses could be utilized more in the screening phase of items?

6.4 Limitations of the study

An obvious limitation of the study is the lack of further simulation of the proposed corrections of attenuation, R_{AC} , E_{AC} , and E_{AC}^2 (see, however, Metsämuuronen 2022a). Another is that an enhanced form of η is available only for a binary case, and not for the polytomous ordinal case. A possible short-cut method for the sign of η for polytomous cases was discussed: to use the sign of PMC for the same variables as a sign for η . Until better formulae are developed, this may serve as an intermediate solution to reach the real η . However, if this method is applied, one needs to be aware of three limitations in the interpretations of the results.

First, the short-cut method makes sense only when the categorical variable is ordinal (or better); PMC does not get any relevant interpretation with truly nominal-scaled variables with polytomous categories. Second, the sign of η does not tell whether the pattern between the variables is convex or concave, but it indicates the overall (linear) trend in the pattern. In these settings, the estimate of the true association between the variables is, most probably, reached closer by η than by PMC, because the latter detects the linear association, while η is not restricted to linear correlation. Third, if PMC equals zero, irrespective of the pattern between the variables, the sign of η will always be positive, because the sign of PMC cannot be used as an indicator of the sign of η .

From this point of view, testing of the null hypothesis related to $\rho_{gX}=0$ is a relevant side procedure in the short-cut method; only when the estimate by PMC truly differs zero, the sign of PMC should be used in the process.

Appendix

Algebraic reasons why η underestimates association in the binary case

Assume a metric variable X with observed values x_j and a binary variable g with observed values $y_i=0,1$. In binary case, eta can be expressed in a form

$$\eta(g|X) = \frac{(\bar{X}_{X1} - \bar{X}_{X0}) \times \sigma_g}{\sigma_X}, \tag{1}$$

where \bar{X}_{X0} and \bar{X}_{X1} refer to the means of the variable X in the subpopulations $y=0$ and $y=1$, and σ_g and σ_X are standard deviations of g and X . By denoting the grand mean in X by GM_X , the variance of X can be manipulated as follows:

$$\begin{aligned} \sigma_X^2 &= \frac{1}{N} \left(\sum_{j=1}^{n_0} (x_j - GM_X)^2 + \sum_{j=1}^{n_1} (x_j - GM_X)^2 \right) \\ &= \frac{1}{N} \left(\sum_{j=1}^{n_0} (x_j - \bar{X}_{X0})^2 + n_0(\bar{X}_{X0} - GM_X)^2 + \sum_{j=1}^{n_1} (x_j - \bar{X}_{X1})^2 + n_1(\bar{X}_{X1} - GM_X)^2 \right) \\ &= \frac{1}{N} \left(n_0 \sum_{j=1}^{n_0} \frac{(x_j - \bar{X}_{X0})^2}{n_0} + n_1 \sum_{j=1}^{n_1} \frac{(x_j - \bar{X}_{X1})^2}{n_1} + n_0(\bar{X}_{X0} - GM_X)^2 + n_1(\bar{X}_{X1} - GM_X)^2 \right) \\ &= \frac{1}{N} \left(n_0\sigma_{X0}^2 + n_1\sigma_{X1}^2 + n_0(\bar{X}_{X0} - GM_X)^2 + n_1(\bar{X}_{X1} - GM_X)^2 \right) \\ &= (1 - p)\sigma_{X0}^2 + p\sigma_{X1}^2 + (1 - p)(\bar{X}_{X0} - GM_X)^2 + p(\bar{X}_{X1} - GM_X)^2. \end{aligned} \tag{2}$$

The term $(\bar{X}_{X1} - \bar{X}_{X0}) \times \sigma_g$ can be manipulated as follows:

$$(\bar{X}_{X1} - \bar{X}_{X0}) \times \sigma_g = \left((\bar{X}_{X1} - GM_X) - (\bar{X}_{X0} - GM_X) \right) \times \sigma_g. \tag{3}$$

Then, $\eta(g|X)$ can be expressed in a form

$$\eta(g|X) = \frac{\left((\bar{X}_{X1} - GM_X) - (\bar{X}_{X0} - GM_X) \right) \times \sigma_g}{\sqrt{p\sigma_{X1}^2 + (1 - p)\sigma_{X0}^2 + p(\bar{X}_{X1} - GM_X)^2 + (1 - p)(\bar{X}_{X0} - GM_X)^2}}. \tag{4}$$

Because of having only two means, they are related to each other through GM_X

$$p(\bar{X}_{X1} - GM_X) = (1 - p)(GM_X - \bar{X}_{X0}) = -(1 - p)(\bar{X}_{X0} - GM_X). \tag{5}$$

leading to

$$(\bar{X}_{X0} - GM_X) = -\frac{p}{(1 - p)}(\bar{X}_{X1} - GM_X) \tag{6}$$

and

$$(\bar{X}_{X0} - GM_X)^2 = \frac{p^2}{(1 - p)^2}(\bar{X}_{X1} - GM_X)^2. \tag{7}$$

Because of Eqs. (5), (6), and (7), and because $\sigma_g = \sqrt{p(1 - p)}$

$$\begin{aligned} \eta(g|X) &= \frac{\left((\bar{X}_{X1} - GM_X) + \frac{p}{(1-p)}(\bar{X}_{X1} - GM_X) \right) \times \sqrt{p(1 - p)}}{\sqrt{p\sigma_{X1}^2 + (1 - p)\sigma_{X0}^2 + p(\bar{X}_{X1} - GM_X)^2 + \frac{(1-p)p^2}{(1-p)^2}(\bar{X}_{X1} - GM_X)^2}} \\ &= \frac{\sqrt{\frac{p}{(1-p)}(\bar{X}_{X1} - GM_X)}}{\sqrt{\frac{(1-p)}{(1-p)}p\sigma_{X1}^2 + \frac{p(1-p)}{p(1-p)}(1 - p)\sigma_{X0}^2 + \frac{p}{(1-p)}(\bar{X}_{X1} - GM_X)^2}} \\ &= \frac{\sqrt{\frac{p}{(1-p)}(\bar{X}_{X1} - GM_X)}}{\sqrt{\frac{p}{(1-p)}\sqrt{(1 - p)\sigma_{X1}^2 + \frac{(1-p)^2}{p}\sigma_{X0}^2 + (\bar{X}_{X1} - GM_X)^2}}} \\ &= \frac{(\bar{X}_{X1} - GM_X)}{\sqrt{(1 - p)\sigma_{X1}^2 + (1 - p)\left(\frac{1}{p} - 1\right)\sigma_{X0}^2 + (\bar{X}_{X1} - GM_X)^2}}. \tag{8} \end{aligned}$$

Equation (8) shows strictly that the magnitude of $\eta(g|X)$ and, consequently, of $\eta_{g|X}^2$ can reach the perfect 1 only in the theoretical case of $\sigma_{X1}^2 = \sigma_{X0}^2 = 0$. From the behavior of $\eta(g|X)$, we know that this generalizes to polytomous cases too ($\sigma_{Xi}^2 = \sigma_{Xj}^2 = 0$), although showing it algebraically is not obvious. The highest magnitude in Eq. (8) is achieved when $p=0.5$, that is, when σ_g is the highest

$$\eta_{g|X}^{Max} = \frac{(\bar{X}_{X1} - GM_X)}{\sqrt{0.5 \times (\sigma_{X1}^2 + \sigma_{X0}^2) + (\bar{X}_{X1} - GM_X)^2}}. \tag{9}$$

If we assume symmetric distribution of X around $p=0.5$, the symmetricity with two means determinates that $\sigma_{X1}^2 = \sigma_{X0}^2$. Then, the maximum magnitude of $\eta(g|X)$ is

$$\eta_{g|X}^{Max} = \frac{(\bar{X}_{X1} - GM_X)}{\sqrt{\sigma_{X1}^2 + (\bar{X}_{X1} - GM_X)^2}} = \frac{(\bar{X}_{X1} - GM_X)}{\sqrt{\sigma_{X0}^2 + (\bar{X}_{X1} - GM_X)^2}}. \quad (10)$$

Again, the maximum value is strictly dependent on the variance of X in subpopulations of g (0 and 1), and $\eta_{g|X}^{Max} = 1$ can be reached only when $\sigma_{X_i}^2 = \sigma_{X_j}^2 = 0$, that is, when all cases in the sub-group i in g share the identical category in X . This condition necessitates a common number of categories in both variables, although the categories do not need to be in a linear order (cf. PMC).

Acknowledgements The author would sincerely thank an anonymous reviewer for pointing valuable sources of the algebraic connection of eta and PMC. These references also led to valuable secondary sources regarding the phenomenon. Discussions with all reviewers were valuable, and those shaping the text in a good way; some parts of the discussion are incorporated in the footnotes of the article.

Funding Open Access funding provided by University of Turku (UTU) including Turku University Central Hospital. No funding was received for preparing the article. University of Turku paid the publication fee

Declarations

Conflict of interest There is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aitken AC (1934) Note on selection from a multivariate normal population. *Proc Edinb Math Soc* 4(2):106–110. <https://doi.org/10.1017/S0013091500008063>
- Aquirre-Urreta M, Rönkkö M, McIntosh CN (2019) A cautionary note on the finite sample behavior of maximal reliability. *Psychol Methods* 24(2):236–252. <https://doi.org/10.1037/met0000176>
- Armor D (1973) Theta reliability and factor scaling. *Sociol Methodol* 5:17–50. <https://doi.org/10.2307/270831>
- Ayres LP (1920) The correlation ratio. *J Educ Res* 2(1):452–456. <https://doi.org/10.1080/00220671.1920.10879073>
- Bravais A (1844) *Analyse Mathématique. Sur les probabilités des erreurs de situation d'un point.* (Mathematical analysis of the error probabilities of a point). *Mémoires présentés par divers savants à l'Académie Royale des Sciences de l'Institut de France* (Memoirs presented by various scholars to the Royal Academy of Sciences of the Institute of France), 9, 255–332. https://books.google.fi/books?id=7g_hQAACAAM&redir_esc=y

- Byrne BM (2016) Structural equation modeling with AMOS. Basic concepts, applications, and programming, 3rd edn. Routledge
- Chalmers RP (2017) On misconceptions and the limited usefulness of ordinal alpha. *Educ Psychol Measur* 78(6):1056–1071. <https://doi.org/10.1177/0013164417727036>
- Chan D (2008) So why ask me? Are self-report data really that bad? In: Lance CE, Vanderberg RJ (eds) *Statistical and methodological myths and urban legends*. Routledge, pp 309–326
- Cheng Y, Yuan K-H, Liu C (2012) Comparison of reliability measures under factor analysis and item response theory. *Educ Psychol Measur* 72(1):52–67. <https://doi.org/10.1177/0013164411407315>
- Cohen J (1965) Some statistical issues in psychological research. In: Wolman BB (ed) *Handbook of clinical psychology*. McGraw-Hill
- Cohen J (1969) *Statistical power analysis for the behavioral sciences*. Academic press
- Cohen J (1973) Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educ Psychol Measur* 33(1):107–112. <https://doi.org/10.1177/001316447303300111>
- Cohen J (1988) *Statistical power analysis for the behavioral sciences*, 2nd edn. Erlbaum
- Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* 16(3):297–334. <https://doi.org/10.1007/BF02310555>
- Eikeland HM (1971) On the generality of univariate eta. *Scand J Educ Res* 15(1):149–167. <https://doi.org/10.1080/0031383710150109>
- Fisher R (1925) *Statistical methods for research workers*. Oliver and Boyd
- Friedman H (1968) Magnitude of experimental effect and a table for its rapid estimation. *Psychol Bull* 70(4):245–251. <https://doi.org/10.1037/h0026258>
- Gademann AM, Guhn M, Zumbo BD (2012) Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Pract Assess Res Eval* 17(3):1–13. <https://doi.org/10.7275/n560-j767>
- Galton F (1889) Kinship and correlation. *Stat Sci* 4(2):80–86. <https://doi.org/10.1214/ss/1177012581> (Also, 1890 in *North American Review*, 150, 419–431)
- Gibbons JD (1993) *Nonparametric statistic. An introduction. Quantitative applications for social sciences*. SAGE Publications, Inc, p 90
- Glass GV, Hakstian AR (1969) Measures of association in comparative experiments: their development and interpretation. *Am Educ Res J* 6(3):403–414. <https://doi.org/10.2307/1161859>
- Glen S (2016) Phi coefficient (mean square contingency coefficient). From StatisticsHowTo.com. <https://www.statisticshowto.com/phi-coefficient-mean-square-contingency-coefficient/>
- Goodman LA, Kruskal WH (1954) Measures of association for cross classifications. *J Am Stat Assoc* 49(268):732–764. <https://doi.org/10.1080/01621459.1954.10501231>
- Greene VL, Carmines EG (1980) Assessing the reliability of linear composites. *Sociol Methodol* 11:160–217. <https://doi.org/10.2307/270862>
- Hays WL (1963) *Statistics for psychologists*. Holt, Rinehart & Winston
- Howell DG (2012) *Statistical methods for psychology*, 8th edn. Wadsworth
- IBM (2017) IBM SPSS Statistics 25 algorithms. IBM. ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/25.0/en/client/Manuals/IBM_SPSS_Statistics_Algorithms.pdf
- Jonckheere AR (1954) A distribution-free k-sample test against ordered alternatives. *Biometrika* 41(1–2):133–145. <https://doi.org/10.1093/biomet/41.1-2.133>
- Kelley TL (1935) An unbiased correlation ratio measure. *Proc Natl Acad Sci* 21:554–559
- Kennedy JJ (1970) The eta coefficient in complex ANOVA designs. *Educ Psychol Measur* 30(4):885–889. <https://doi.org/10.1177/001316447003000409>
- Kerlinger FN (1964) *Foundations of behavioral research*. Holt, Rinehart & Winston
- Kim J-O (1971) Predictive measures of ordinal association. *Am J Sociol* 76(5):891–907
- Kuder GF (1937) Nomograph for point biserial r , biserial r , and fourfold correlations. *Psychometrika* 2:135–138. <https://doi.org/10.1007/BF02288067>
- Lawley DN (1943) A note on Karl Pearson's selection formulae. *Proc R Soc Edinb Sect A: Math* 61(1):28–30. <https://doi.org/10.1017/S0080454100006385>
- Lane S, Raymond MR, Haladyna TM (2016) *Handbook of test development*, 2nd edn. Routledge
- Lavrakas PJ (2008) Attenuation. In: Lavrakas PJ (ed) *Encyclopedia of Survey Methods*. Sage Publications Inc
- Li H (1997) A unifying expression for the maximal reliability of a linear composite. *Psychometrika* 62(2):245–249. <https://doi.org/10.1007/BF02295278>
- Li H, Rosenthal R, Rubin DB (1996) Reliability of measurement in psychology: from Spearman-Brown to maximal reliability. *Psychol Methods* 1(1):98–107. <https://doi.org/10.1037/1082-989X.1.1.98>

- Lord FM, Novick MR (1968) Statistical theories of mental test scores. Addison-Wesley Publishing Company
- Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* 18(1):50–60. <https://doi.org/10.1214/aoms/1177730491>
- Martin WS (1973) The effects of scaling on the correlation coefficient: A test of validity. *J Market Res* 10(3):316–318. <https://doi.org/10.2307/3149702>
- Martin WS (1978) Effects of scaling on the correlation coefficient: Additional considerations. *J Market Res* 15(2):304–308. <https://doi.org/10.1177/002224377801500219>
- McDonald RP (1970) Theoretical canonical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *Br J Math Stat Psychol* 23:1–21. <https://doi.org/10.1111/j.2044-8317.1970.tb00432.x>
- Meade AW (2010) Restriction of range. In: Salkind NJ (ed) *Encyclopedia of research design*. SAGE Publications, pp 1278–1280
- Metsämuuronen J (2017) *Essentials of research methods in human sciences*, vol 1–3. SAGE Publications
- Mordkoff JT (2019) A simple method for removing bias from a popular measure of standardized effect size: adjusted partial eta squared. *Adv Methods Pract Psychol Sci* 2(3):228–232. <https://doi.org/10.1177/2515245919855053>
- Moses T (2017) A review of developments and applications in item analysis. In: Bennett R, von Davier M (eds) *Advancing human assessment. The methodological, psychological and policy contributions of ETS*. Springer Open, pp 19–46
- Mendoza JL, Mumford M (1987) Corrections for attenuation and range restriction on the predictor. *J Educ Stat* 12(3):282–293. <https://doi.org/10.3102/10769986012003282>
- Metsämuuronen J (2016) Item–total correlation as the cause for the underestimation of the alpha estimate for the reliability of the scale. *Global J Res Anal* 5(1):471–477
- Metsämuuronen J (2020) Somers’ D as an alternative for the item–test and item–rest correlation coefficients in the educational measurement settings. *Int J Educ Measure* 6(1):207–221
- Metsämuuronen J (2021a) Goodman-Kruskal gamma and dimension-corrected gamma in educational measurement settings. *Int J Educ Methodol* 7(1):95–118
- Metsämuuronen J (2021b) Directional nature of Goodman-Kruskal gamma and some consequences—Identity of Goodman-Kruskal gamma and Somers delta, and their connection to Jonckheere-Terpstra test statistic. *Behaviormetrika*. <https://doi.org/10.1007/s41237-021-00138-8>
- Metsämuuronen J (2022a) Effect of various simultaneous sources of mechanical error in the estimators of correlation causing deflation in reliability. Seeking the best options of correlation for deflation-corrected reliability. *Behaviormetrika* 49:91–130. <https://doi.org/10.1007/s41237-022-00158-y>
- Metsämuuronen J (2022b) Deflation-corrected estimators of reliability. *Front Psychol* 12:748672. <https://doi.org/10.3389/fpsyg.2021.748672>
- Metsämuuronen J (2022c) Attenuation-corrected estimators of reliability. *Appl Psychol Measure*. <https://doi.org/10.3389/fpsyg.2021.748672>
- Metsämuuronen J, Ukkola A (2019) Alkumittauksen menetelmällisiä ratkaisuja (Methodological solutions of zero level assessment). *Publications* 18:2019. Finnish Education Evaluation Centre. https://karvi.fi/app/uploads/2019/08/KARVI_1819.pdf
- Newson R (2002) Parameters behind “nonparametric” statistics: Kendall’s tau, Somers’ D and median differences. *The Stata J* 2(1):45–64
- Newson R (2006) Confidence intervals for rank statistics: Somers’ D and extensions. *The Stata J* 6(3):309–334
- Okada K (2013) Is omega squared less biased? A comparison of three major effect size indices in one-way ANOVA. *Behaviormetrika* 40:129–147. <https://doi.org/10.2333/bhmk.40.129>
- Okada K (2017) Negative estimate of variance-accounted-for effect size: how often it is obtained, and what happens if it is treated as zero. *Behav Res Methods* 49:979–987. <https://doi.org/10.3758/s13428-016-0760-y>
- Olsson U (1980) Measuring correlation in ordered two-way contingency tables. *J Mark Res* 17(3):391–394. <https://doi.org/10.1177/002224378001700315>
- Pearson K (1896) VII. Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. *Philos Trans R Soc Lond* 187:253–318. <https://doi.org/10.1098/rsta.1896.0007>
- Pearson K (1903) I. Mathematical contributions to the theory of evolution. —XI. On the influence of natural selection on the variability and correlation of organs. *Philos Trans R Soc A Math Phys Eng Sci* 200(321–330):1–66

- Pearson K (1904) On the theory of contingency and its relation to association and normal correlation. Drapers' Company Research Memoirs. Biometric Series I, XIII. Soho Square, W.: Dulau & Co. <http://archive.org/details/cu31924003064833>
- Pearson K (1905) On the general theory of skew correlation and non-linear regression. London. Dulau & Co. <https://archive.org/details/ongeneraltheory00peargoog/page/n3>
- Pearson K (1911) On a correction to be made to the correlation ratio η . *Biometrika* 8(1/2):254–256. <https://doi.org/10.2307/2331454>
- Peters CC, Van Voorhis WR (1940) Statistical procedures and their mathematical bases. McGraw-Hill
- Raykov T (2004) Estimation of maximal reliability: a note on a covariance structure modeling approach. *Br J Math Stat Psychol* 57(1):21–27. <https://doi.org/10.1348/000711004849295>
- Richardson JTE (1996) Measures of effect size. *Behav Res Methods Instrum Comput* 28(1):12–22. <https://doi.org/10.3758/BF03203631>
- Sackett PR, Lievens F, Berry CM, Landers RN (2007) A cautionary note on the effect of range restriction on predictor intercorrelations. *J Appl Psychol* 92(2):538–544. <https://doi.org/10.1037/0021-9010.92.2.538>
- Sackett PR, Yang H (2000) Correction for range restriction: an expanded typology. *J Appl Psychol* 85(1):112–118. <https://doi.org/10.1037/0021-9010.85.1.112>
- Sechrest L, Yeaton WH (2011) Magnitudes of experimental effects in social science research. In: Salkind NJ (ed) *SAGE directions of educational psychology*, vol IV. SAGE Publications Inc., pp 3–22
- Sheskin DJ (2011) *Handbook of parametric and nonparametric statistical procedures*, 5th edn. Chapman & Hall/CRC
- Siegel S, Castellan NJ Jr (1988) *Nonparametric statistics for the behavioural sciences*, 2nd edn. McGraw-Hill, Singapore
- Sirkin MR (2006) *Statistics of the social science*, 3rd edn. SAGE Publications Inc.
- Skidmore ST, Thompson B (2011) Choosing the best correction formula for the Pearson r^2 effect size. *J Exp Educ* 79(3):257–278. <https://doi.org/10.1080/00220973.2010.484437>
- Schmidt FL, Hunter JE (2015) *Methods of meta-analysis: correcting error and bias in research findings*, 3rd edn. SAGE Publications
- Schmidt FL, Shaffer JA, Oh I-S (2008) Increased accuracy for range restriction corrections: implications for the role of personality and general mental ability in job and training performance. *Pers Psychol* 61(4):827–868. <https://doi.org/10.1111/j.1744-6570.2008.00132.x>
- Somers RH (1962) A new asymmetric measure of association for ordinal variables. *Am Sociol Rev* 27(6):799–811. <https://doi.org/10.2307/2090408>
- Spearman C (1904) The proof and measurement of association between two things. *Am J Psychol* 15(1):72–101. <https://doi.org/10.2307/1422689>
- Swineford F (1936) Biserial r versus Pearson r as measures of test-item validity. *J Educ Psychol* 27(6):471–472. <https://doi.org/10.1037/h0052118>
- Thorndike RL (1949) *Personnel selection*. Wiley
- Terpstra TJ (1952) The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. *Indag Math* 14(3):327–333. [https://doi.org/10.1016/S1385-7258\(52\)50043-X](https://doi.org/10.1016/S1385-7258(52)50043-X)
- Vacha-Haase T, Thompson B (2004) How to estimate and interpret various effect sizes. *J Couns Psychol* 51(4):473–481. <https://doi.org/10.1037/0022-0167.51.4.473>
- Walk MJ, Rupp AA (2010) Pearson product-moment correlation coefficient. In: Salkind NJ (ed) *Encyclopedia of research design*. SAGE Publications, pp 1022–1026
- Wherry RJ, Taylor EK (1946) The relation of multiserial η to other measures of correlation. *Psychometrika* 11:155–161. <https://doi.org/10.1007/BF02289296>
- Wholey JS, Hatry HP, Newcomer KE (eds) (2015) *Handbook of practical program evaluation*, 4th edn. Jossey-Bass, Berlin
- Yang H (2010) Factor loadings. In: Salkind NJ (ed) *Encyclopedia of research design*. SAGE Publications, pp 480–483
- Yule GU (1912) On the methods of measuring association between two attributes. *J R Stat Soc* 75(6):579–652