

A comparison of two indices for the intraclass correlation coefficient

Gwown Shieh

Published online: 7 March 2012
© Psychonomic Society, Inc. 2012

Abstract In the present study, we examined the behavior of two indices for measuring the intraclass correlation in the one-way random effects model: the prevailing ICC(1) (Fisher, 1938) and the corrected eta-squared (Bliese & Halverson, 1998). These two procedures differ both in their methods of estimating the variance components that define the intraclass correlation coefficient and in their performance of bias and mean squared error in the estimation of the intraclass correlation coefficient. In contrast with the natural unbiased principle used to construct ICC(1), in the present study it was analytically shown that the corrected eta-squared estimator is identical to the maximum likelihood estimator and the pairwise estimator under equal group sizes. Moreover, the empirical results obtained from the present Monte Carlo simulation study across various group structures revealed the mutual dominance relationship between their truncated versions for negative values. The corrected eta-squared estimator performs better than the ICC(1) estimator when the underlying population intraclass correlation coefficient is small. Conversely, ICC(1) has a clear advantage over the corrected eta-squared for medium and large magnitudes of population intraclass correlation coefficient. The conceptual description and numerical investigation provide guidelines to help researchers choose between the two indices for more accurate reliability analysis in multilevel research.

Keywords Multilevel modeling · Random effect · Reliability

Because of the inherently hierarchical nature of experimental randomization by cluster and clustered sampling designs, multilevel phenomena and modeling are common in behavioral, educational, and social sciences. Although the underlying structure of multilevel models is more complex than the traditional linear regression models, their basic concept and modeling framework provide a straightforward mechanism for examining the interrelation of individual and group influences. Thus, the class of hierarchical linear models has become the most widely accepted basis for multilevel analysis (Goldstein, 2002; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). The relevant methodological and theoretical issues have been further addressed throughout various areas of research by Cools, Van den Noortgate, and Onghena (2009); Hedge and Hedberg (2007); Hoffman and Rovine (2007); Hofmann (1997, 2002); Hofmann, Griffin, and Gavin (2000); Kozlowski and Klein (2000); Murray, Varnell, and Bliestein (2004); and Tasoluk, Droge, and Calantone (2009).

One of the distinctive features of multilevel data is that the primary outcomes for study participants within the same unit are more similar than the responses by individuals in different units. Hence, a problem of interest with regard to hierarchical data is the assessment of similarity or degree of resemblance among cluster members. The clustering can be expressed in terms of correlation among the measurements on individuals within the same group. The fundamental properties and general guidelines related to intraclass correlation coefficient (ICC) as an interrater reliability measure are described in Bartko (1976), McGraw and Wong (1996), and Shrout and Fleiss (1979). In particular, McGraw and Wong emphasized the distinction between various forms of the ICC developed using data from one-way and two-way random and mixed-effect ANOVA models. In a typical interrater reliability study, each of a random sample of g

G. Shieh (✉)
Department of Management Science,
National Chiao Tung University,
1001 Ta Hsueh Road,
Hsinchu, Taiwan 30050
e-mail: gwshieh@mail.nctu.edu.tw

participants is rated independently by n judges. For the simplest case, the one-way random effect model assumes that the random participant represents the only systematic source of variation (without judge effect). On the other hand, the two-way random or mixed effect model assumes there is a systematic source of random or fixed judge effect as well as the random participant variation. Moreover, the two-way random and mixed models can be extended to accommodate the additional participant–judge interaction effect. Notably, these four two-way models differ in whether judge represents random or fixed effects and in whether the model includes an interaction component (For ease of reference, the prescribed five models were noted as Cases 1, 2A, 3A, 2, and 3 in Table 1 of McGraw and Wong 1996). Consequently, the conceptual modeling difference among these five models ultimately leads to a distinct and unique definition of ICCs. The corresponding standard ANOVA estimators can be constructed with the mean square expectation (as shown in Tables 3 and 4 of McGraw & Wong, 1996). However, McGraw and Wong's primary focus was to provide a comprehensive account of procedures for computing confidence intervals and constructing significance tests. Hence, the existence and comparison of alternative estimators of ICC reliability measure were not considered. Note that there is a considerable amount of recent literature pertaining to both

the theoretical and practical problem of meaningfully and efficiently estimating correlations among individuals in nested designs (e.g., Alfères & Kenny, 2009; Beal & Dawson, 2007; Bliese, 1998, 2000; Bliese & Halverson, 1998; Castro, 2002; Courrieu, Brand-D'abrescia, Peereman, Spieler, & Rey, 2011; Hedges & Rhoads, 2011; LeBreton & Senter, 2008; O'Connor, 2004; Wampold & Serlin, 2000).

In the present article, we focused on the intraclass correlation coefficient that quantifies the homogeneity level in responses by individuals from the same group for the outcome of a one-way random-effects model, which is often useful as an initial step in the development of a multilevel analysis. Accordingly, the ICC(1) index, introduced by Fisher (1938), is the most frequently adopted measure of intraclass correlation. A comprehensive review of various inference procedures for the ICC in the one-way random-effects model was given by Donner (1986). Moreover, an extensive description of various methods for estimating and testing the within-groups and between-groups variances was presented in Searle, Casella, and McCulloch (1992). The purposes of the present study were (a) to explicate the essential issues concerning the estimation of ICC, and (b) to provide specific guidance for the use of proper ICC index in the light of new empirical evidences based on simulation technique.

Table 1 The bias of ICC indices for $g = 4$ and $N = 20$ ($\bar{n} = 5$)

ρ	Group Size Structure					
	Equal		Linear		Extreme	
	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$
0.00	0.063759	0.037186	0.067002	0.039329	0.086986	0.051589
0.05	0.041917	0.007477	0.044562	0.009193	0.061296	0.019523
0.10	0.023034	-0.018711	0.024420	-0.017895	0.036878	-0.010710
0.15	0.006223	-0.041795	0.007667	-0.041060	0.015692	-0.037663
0.20	-0.009490	-0.063138	-0.009490	-0.063435	-0.005943	-0.063879
0.25	-0.023619	-0.082057	-0.026342	-0.084796	-0.026351	-0.088472
0.30	-0.037888	-0.100120	-0.040228	-0.102587	-0.043018	-0.108729
0.35	-0.049613	-0.115218	-0.054630	-0.119937	-0.060379	-0.129223
0.40	-0.061150	-0.129277	-0.067756	-0.135419	-0.075783	-0.147017
0.45	-0.074410	-0.143853	-0.078682	-0.148031	-0.088404	-0.161189
0.50	-0.081787	-0.152420	-0.088148	-0.158560	-0.102256	-0.176245
0.55	-0.090286	-0.161014	-0.097730	-0.168433	-0.112250	-0.186532
0.60	-0.095447	-0.165508	-0.105590	-0.175620	-0.121990	-0.195793
0.65	-0.101594	-0.170155	-0.111976	-0.180779	-0.127277	-0.199802
0.70	-0.106299	-0.172178	-0.113498	-0.179889	-0.129185	-0.199079
0.75	-0.104033	-0.166232	-0.114432	-0.177056	-0.127992	-0.194387
0.80	-0.101521	-0.158582	-0.110571	-0.168241	-0.122973	-0.184157
0.85	-0.091493	-0.141216	-0.101177	-0.151939	-0.112605	-0.166284
0.90	-0.077342	-0.117232	-0.085953	-0.127172	-0.093294	-0.136623
0.95	-0.050889	-0.076458	-0.056408	-0.083098	-0.061256	-0.089314
0.99	-0.014940	-0.022268	-0.016491	-0.024252	-0.017316	-0.025345

Despite the widespread use and apparent utility of ICC in several different fields of research, the most fundamental interpretation is that it is a measure of the proportion of the total variance in the outcome that is accounted for by the group membership. In multiple linear regression, it is well known that the coefficient of determination R^2 represents the overall usefulness of the regression model, making it an appealing measure for the strength of association or the percentage of total variance that is explained by the predictor variables. However, the advocated contention regarding the role of R^2 in linear regression models does not generalize to a one way random-effects model in a straightforward manner. Interestingly, the corresponding measure for the proportionate reduction in error is denoted by $\hat{\eta}^2$ in the context of ANOVA designs, and both R^2 and $\hat{\eta}^2$ have the identical formulation, as was noted by Maxwell, Camp, and Arvey (1981). Although $\hat{\eta}^2$ is widely used in organizational research to draw inferences about group membership cohesion, it was shown in Bliese and Halverson (1998) that the sample eta-squared estimator $\hat{\eta}^2$ is an inaccurate measure of the amount of the total variance that is due to the group level properties. Specifically, their simulation study revealed that $\hat{\eta}^2$ provides a positively biased estimate of ICC and that the performance varies with group size and underlying magnitude of population intraclass correlation. Although the inappropriateness of $\hat{\eta}^2$ as an estimate of ICC was effectively demonstrated in the empirical results of Bliese and Halverson, there remains the problem: Which other indices should be used, and which ones are preferable under particular conditions?

Accordingly, Bliese and Halverson (1998) suggested the corrected eta-squared $\text{Eta}^2(\text{C})$ formula as a modification of $\hat{\eta}^2$ to provide more accurate estimates of the intraclass correlation. Although the proposed estimator $\text{Eta}^2(\text{C})$ may be useful in many settings, it was noted that $\text{Eta}^2(\text{C})$ systematically underestimates the population group level property and that the magnitude of underestimate is caused without adjusting the degrees of freedom associated with a sample when drawing inferences about the population intraclass correlation. Instead, Bliese and Halverson recommended the well-known ICC(1) formula because ICC(1) estimates are not biased by either group size or the number of groups in the sample. Since the empirical investigation of Bliese and Halverson focused on the behavior of $\hat{\eta}^2$, they did not present any numerical examinations regarding the properties of $\text{Eta}^2(\text{C})$ and ICC(1) in their estimation of intraclass correlation. From a methodological standpoint, the prescribed features of $\text{Eta}^2(\text{C})$ and ICC(1) in Bliese and Halverson, therefore, should be clarified.

In addition, the following three caveats seem to have been overlooked. First, although ICC(1) is strongly tied to normal theory and unbiased estimation of variance components, the derivation of $\text{Eta}^2(\text{C})$ is not clearly

conveyed relative to the existing estimation principles and formulas. Second, a serious disadvantage of the two estimators $\text{Eta}^2(\text{C})$ and ICC(1) is that they can assume negative values even though ICC is defined as a non-negative parameter. In practice, the estimate is often set equal to zero when this occurs. Although this simple and intuitive adjustment is of practical meaning, the fundamental behavior of the ICC estimator is inherently altered. Third, regarding the estimation appraisal, unbiasedness is certainly not the only criterion of theoretical importance. Mean square error (MSE) is another useful performance criterion obtained by incorporating the bias (accuracy) and variability (precision) of an estimator. Realistically, however, the existing results are arguably not detailed enough to demonstrate the explicit and relative estimation performance between the two indices. In the present research, we took up the aforementioned methodological and theoretical issues with technical clarifications and numerical investigations. Finally, Monte Carlo simulation study was conducted to assess the bias and MSE of the truncated counterparts of $\text{Eta}^2(\text{C})$ and ICC(1) for negative values under several model configurations in terms of varied group structure and magnitude of intraclass correlation. The present account helps to clarify their unique and contrasting behavior and to choose an appropriate ICC measure in the context of studying the degree of clustering.

Measures of intraclass correlation

In a two-level study design, suppose the response variable is measured on each of n_i individuals within each of g groups that arise from the one-way random effects model assuming the following form:

$$Y_{ij} = \mu + \gamma_i + \varepsilon_{ij}, i = 1, \dots, g; j = 1, \dots, n_i, \quad (1)$$

where Y_{ij} is the individual level outcome, μ is the grand mean, γ_i and ε_{ij} are independent random variables with $\gamma_i \sim N(0, \sigma_\gamma^2)$, and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. The variance of Y_{ij} is then given by $\sigma_\gamma^2 + \sigma_\varepsilon^2$, where σ_γ^2 represents the between-groups variance and σ_ε^2 represents the within-groups variance. Accordingly, the ICC ρ is defined as $\rho = \sigma_\gamma^2 / (\sigma_\gamma^2 + \sigma_\varepsilon^2)$, which can be interpreted as a simple correlation coefficient $\text{Corr}(Y_{ij}, Y_{ij'})$ between any two observations, Y_{ij} and $Y_{ij'}$, in the same group with $j \neq j'$. We are interested in assessing the magnitude of ICC, which is directly interpretable as the proportion of the total variance of the response that is accounted for by the clustering or group cohesion.

From the one-way random effects model defined in Eq. 1, the estimated measure of proportionate reduction in error is

$$\hat{\eta}^2 = \frac{SSB}{SST}, \tag{2}$$

where *SSB* is the between-groups sum of squares, and *SST* is the total sum of squares. Although $\hat{\eta}^2$ accurately describes the effect of group membership in a sample, it systematically overestimates ICC (Bliese & Halverson, 1998). The well-established ANOVA estimator $\hat{\rho}_A$ is obtained by replacing variance parameters in population ICC ρ with corresponding unbiased estimators:

$$\hat{\rho}_A = \frac{MSB - MSW}{MSB + (n_0 - 1)MSW} = \frac{F^* - 1}{F^* + n_0 - 1}, \tag{3}$$

where *MSB* is the between-groups mean square, *MSW* is the within-groups mean square, $n_0 = \left(N - \sum_{i=1}^g n_i^2 / N\right) / (g - 1)$, $N = \sum_{i=1}^g n_i$ is the total number of observations, and $F^* = MSB /$

MSW. For a group size of equal number, $n_i = n$, it is readily apparent that $\hat{\rho}_A$ reduces to the familiar formula

$$ICC(1) = \frac{MSB - MSW}{MSB + (n - 1)MSW} = \frac{F^* - 1}{F^* + n - 1} \tag{4}$$

Although $\hat{\rho}_A$ is the most frequently adopted measure of intraclass correlation, it is not unbiased (Donner, 1986). Unfortunately, this general result contradicts the position of Bliese and Halverson that ICC(1) is not biased by either group size or the number of groups. Such conflicting results will be discussed later in the numerical investigation. Note that Olkin and Pratt (1958) derived the minimum variance unbiased estimator of the intraclass correlation, but its use has been impeded by the lack of a closed form expression. The corresponding computation requires a special purpose computer program (see, e.g., Donoghue & Collins, 1990). Apparently the computational complexity of the unbiased estimator has resulted in limited acceptance for practical use. Thus, it is worthwhile to consider alternative formulas that might yield similar results with less computation.

Alternatively, the maximum likelihood estimator $\hat{\rho}_M$ of ρ can be derived, although an explicit expression generally

Table 2 The bias of ICC indices for $g = 4$ and $N = 200$ ($\bar{n} = 50$)

ρ	Group Size Structure					
	Equal		Linear		Extreme	
	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$
0.00	0.006112	0.003033	0.006619	0.003305	0.059756	0.031544
0.05	-0.000862	-0.015910	-0.001105	-0.016212	0.034656	-0.001356
0.10	-0.006034	-0.030454	-0.006738	-0.031093	0.012869	-0.030265
0.15	-0.012225	-0.044496	-0.013321	-0.045407	-0.007387	-0.056856
0.20	-0.020230	-0.059061	-0.021807	-0.060301	-0.028802	-0.083135
0.25	-0.027562	-0.072111	-0.030141	-0.074218	-0.046770	-0.105845
0.30	-0.036033	-0.085420	-0.040701	-0.089263	-0.063765	-0.126897
0.35	-0.045367	-0.098606	-0.050266	-0.102756	-0.079014	-0.145713
0.40	-0.054033	-0.110586	-0.060300	-0.115918	-0.093468	-0.162646
0.45	-0.063636	-0.122564	-0.068735	-0.126966	-0.105734	-0.176998
0.50	-0.071081	-0.131833	-0.078283	-0.138156	-0.118549	-0.190907
0.55	-0.077796	-0.139577	-0.086244	-0.147152	-0.127686	-0.200781
0.60	-0.086447	-0.148293	-0.094372	-0.155643	-0.134336	-0.207366
0.65	-0.091229	-0.152451	-0.098157	-0.158948	-0.140529	-0.212266
0.70	-0.094891	-0.154479	-0.102811	-0.162163	-0.141644	-0.211475
0.75	-0.095397	-0.152197	-0.103688	-0.160549	-0.140476	-0.206726
0.80	-0.093118	-0.145744	-0.101316	-0.154073	-0.133989	-0.195347
0.85	-0.086670	-0.133180	-0.094413	-0.141377	-0.122011	-0.176291
0.90	-0.073959	-0.111822	-0.081207	-0.119888	-0.101866	-0.146213
0.95	-0.048859	-0.073378	-0.054263	-0.079711	-0.066561	-0.095479
0.99	-0.014521	-0.021695	-0.016146	-0.023725	-0.019153	-0.027661

Table 3 The bias of ICC indices for $g = 10$ and $N = 50$ ($\bar{n} = 5$)

ρ	Group Size Structure					
	Equal		Linear		Extreme	
	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$
0.00	0.040430	0.029886	0.041376	0.030547	0.058236	0.043240
0.05	0.021726	0.006245	0.022530	0.006808	0.036237	0.016995
0.10	0.008721	-0.011003	0.009178	-0.010744	0.017448	-0.005492
0.15	-0.000804	-0.023969	-0.000361	-0.023660	0.002056	-0.024008
0.20	-0.007738	-0.033598	-0.008178	-0.034064	-0.010448	-0.039045
0.25	-0.013300	-0.041106	-0.013904	-0.041737	-0.021566	-0.052167
0.30	-0.018631	-0.047839	-0.019630	-0.048793	-0.031613	-0.063575
0.35	-0.022554	-0.052614	-0.025247	-0.055284	-0.039272	-0.072217
0.40	-0.025994	-0.056485	-0.029714	-0.060165	-0.047902	-0.081294
0.45	-0.029491	-0.060053	-0.034535	-0.065062	-0.052596	-0.086023
0.50	-0.032805	-0.063009	-0.037516	-0.067721	-0.057896	-0.090879
0.55	-0.034672	-0.064180	-0.040376	-0.069915	-0.061135	-0.093320
0.60	-0.035881	-0.064279	-0.041911	-0.070355	-0.063899	-0.094875
0.65	-0.037051	-0.063948	-0.042699	-0.069691	-0.063305	-0.092622
0.70	-0.036726	-0.061684	-0.043349	-0.068463	-0.062957	-0.090171
0.75	-0.034275	-0.056751	-0.040767	-0.063466	-0.059778	-0.084397
0.80	-0.032257	-0.051805	-0.037458	-0.057228	-0.054829	-0.076286
0.85	-0.027931	-0.043893	-0.033449	-0.049722	-0.047159	-0.064807
0.90	-0.021124	-0.032745	-0.025105	-0.037005	-0.035804	-0.048769
0.95	-0.012217	-0.018616	-0.014946	-0.021584	-0.020691	-0.027929
0.99	-0.002759	-0.004153	-0.003358	-0.004816	-0.004688	-0.006291

does not exist. However, much of the complexity is considerably simplified if each group is of the same size. As was noted by Donner (1986) concerning the special case of $n_i = n$, the maximum likelihood estimator $\hat{\rho}_M$ coincides with the pairwise estimator $\hat{\rho}_P$, defined as the Pearson product-moment correlation computed over all possible pairs of observations that can be constructed within groups:

$$\hat{\rho}_P = \frac{F^* - 1 - 1/(g-1)}{F^* + n - 1 + (n-1)/(g-1)}. \quad (5)$$

Moreover, in line with the estimation of ICC, the corrected eta-squared estimator $\text{Eta}^2(\text{C})$ suggested by Bliese and Halverson (1998, Eq. 4) can be written in our notation as

$$\text{Eta}^2(\text{C}) = \frac{\hat{\eta}^2 - 1/n}{1 - 1/n}. \quad (6)$$

It is straightforward, then, to show from Eq. 2 that $\hat{\eta}^2 = \text{SSB}/\text{SST} = F^*/\{F^* + g(n-1)/(g-1)\}$ for $\text{SST} = \text{SSB} + \text{SSW}$, $\text{MSB} = \text{SSB}/(g-1)$, and $\text{MSW} = \text{SSW}/\{g(n-1)\}$ under balanced design. Hence, an alternative form of $\text{Eta}^2(\text{C})$ given in Eq. 6 can be readily established by using the new expression of $\hat{\eta}^2$ that $\text{Eta}^2(\text{C}) = \{F^* - 1 - 1/(g-1)\}/\{F^* + n - 1 + (n-1)/(g-1)\}$. Consequently, the formulation of

$\text{Eta}^2(\text{C})$ yields the equivalence property for the three estimators $\text{Eta}^2(\text{C}) = \hat{\rho}_P = \hat{\rho}_M$. Interestingly, this fundamental correspondence among the three indices of $\text{Eta}^2(\text{C})$, $\hat{\rho}_P$ and $\hat{\rho}_M$ was not addressed in Bliese and Halverson. In view of the strong connection of $\text{Eta}^2(\text{C})$ with the maximum likelihood principle, it is natural to present a simplified approximation for the case of unequal group sizes with the advantage of giving accessible expression. Accordingly, the following formula with computational ease and general accessibility is considered

$$\hat{\eta}_c^2 = \frac{\hat{\eta}^2 - 1/n_0}{1 - 1/n_0} = \frac{F^* - 1 - 1/(g-1)}{F^* + n_0 - 1 + (n_0 - 1)/(g-1)}. \quad (7)$$

Although a similar notion was presented in Bliese and Halverson, no numerical evidence was provided to demonstrate its estimation property.

Thus, it is easy to see from Eqs. 4, 5 and 6 that $\text{Eta}^2(\text{C})$ ($= \hat{\rho}_P$) and $\text{ICC}(1)$ are virtually equivalent if the number of groups g is large, although $\text{Eta}^2(\text{C})$ is always slightly less than $\text{ICC}(1)$ for all values of $g > 1$. Although there are several comparative studies about the finite-sample properties of the aforementioned estimators (e.g., Bliese & Halverson, 1998; Donner & Koval, 1980; Swiger, Harvey, Everson, & Gregory, 1964), they appear to overlook the fact

Table 4 The bias of ICC indices for $g = 10$ and $N = 500$ ($\bar{n} = 50$)

ρ	Group Size Structure					
	Equal		Linear		Extreme	
	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$
0.00	0.003703	0.002560	0.003831	0.002650	0.040675	0.028632
0.05	-0.000724	-0.007134	-0.000868	-0.007291	0.019723	0.002813
0.10	-0.002105	-0.012461	-0.002893	-0.013201	0.002781	-0.018112
0.15	-0.004339	-0.018084	-0.005363	-0.019019	-0.010845	-0.034973
0.20	-0.006790	-0.023463	-0.008356	-0.024884	-0.022985	-0.049626
0.25	-0.010553	-0.029654	-0.012434	-0.031351	-0.033868	-0.062462
0.30	-0.013988	-0.035135	-0.015945	-0.036894	-0.043239	-0.073184
0.35	-0.016939	-0.039748	-0.020141	-0.042685	-0.051911	-0.082846
0.40	-0.020367	-0.044405	-0.024371	-0.048134	-0.059563	-0.091035
0.45	-0.024954	-0.049819	-0.028059	-0.052668	-0.065700	-0.097389
0.50	-0.026395	-0.051703	-0.031578	-0.056631	-0.071498	-0.102944
0.55	-0.029165	-0.054514	-0.035008	-0.060129	-0.075914	-0.106819
0.60	-0.031594	-0.056522	-0.036857	-0.061628	-0.078591	-0.108556
0.65	-0.032192	-0.056264	-0.039413	-0.063400	-0.080164	-0.108808
0.70	-0.033590	-0.056331	-0.039046	-0.061762	-0.078759	-0.105668
0.75	-0.032443	-0.053317	-0.038557	-0.059484	-0.074874	-0.099467
0.80	-0.030226	-0.048613	-0.036352	-0.054905	-0.069559	-0.091335
0.85	-0.026428	-0.041664	-0.031510	-0.046959	-0.060039	-0.078190
0.90	-0.020785	-0.032068	-0.024820	-0.036346	-0.045925	-0.059506
0.95	-0.012264	-0.018580	-0.014728	-0.021246	-0.026552	-0.034276
0.99	-0.002750	-0.004140	-0.003377	-0.004832	-0.006122	-0.007875

that $\hat{\rho}_A$ and $\hat{\eta}_C^2$ permit negative values whenever $F^* < 1$ or $F^* < g/(g - 1)$, respectively. In practice, the estimate is set to zero when this occurs. This modification incurs a truncated form of estimator, and the associated estimation performance differs from the original counterpart, especially when the underlying ICC is small. In fact, it was reported in Bliese (2000) that the ICC(1) values are typically between 0.05 and 0.20, which implies that the true magnitudes of ICC tend to be small in practical applications. Consequently, the existing numerical results for untruncated formulas do not necessarily apply to the truncated versions of the two estimators. To extend the concept and applicability of the ICC indices, it is prudent to investigate the estimation behavior of their truncated forms.

Numerical study

In this section, we consider the performance appraisal in point estimation of ICC following the notion of choosing a profound index for the level of reliability. In view of the common practice of reporting zero estimates of ρ when ICC measures yield negative values, it is natural to consider the truncated version for such occurrence. The corresponding

modified formulas $\hat{\rho}_{TA}$ and $\hat{\eta}_{TC}^2$ for the two prominent indices $\hat{\rho}_A$ and $\hat{\eta}_C^2$ defined in Eqs. 3 and 7 are defined as

$$\hat{\rho}_{TA} = \hat{\rho}_A \text{ if } \hat{\rho}_A \geq 0, \text{ and } \hat{\rho}_{TA} = 0 \text{ if } \hat{\rho}_A < 0, \tag{8}$$

And

$$\hat{\eta}_{TC}^2 = \hat{\eta}_C^2 \text{ if } \hat{\eta}_C^2 \geq 0 \text{ and } \hat{\eta}_{TC}^2 = 0 \text{ if } \hat{\eta}_C^2 < 0, \tag{9}$$

respectively. Although this natural modification is intuitive and heuristic, the theoretical properties of $\hat{\rho}_{TA}$ and $\hat{\eta}_{TC}^2$ are substantially different from those of their counterparts, $\hat{\rho}_A$ and $\hat{\eta}_C^2$, when the underlying coefficient parameter ρ is small. However, a unified and rigorous presentation of these truncated measures of ρ does not exist to our knowledge. For pedagogical purposes, it is constructive to provide informative results that not only permit new insights into their relationships but also allow clear representations of various methodological issues.

Note that the estimators $\hat{\rho}_A$ and $\hat{\eta}_C^2$ are functions of F^* , as are the two modifications $\hat{\rho}_{TA}$ and $\hat{\eta}_{TC}^2$. It follows from the model assumption with balanced group sizes $n_i = n$ that the F^* statistic is distributed as a multiple of an F distribution $F^* \sim (n\sigma_\gamma^2/\sigma_\varepsilon^2 + 1)F$, where F is an F distribution with $(g - 1)$ and

Table 5 The root mean squared error of ICC indices for $g = 4$ and $N = 20$ ($\bar{n} = 5$)

Group Size Structure						
ρ	Equal		Linear		Extreme	
	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$
0.00	0.126871	0.089515	0.133112	0.094334	0.167778	0.120544
0.05	0.140086	0.104359	0.145052	0.107948	0.174133	0.129706
0.10	0.156674	0.126649	0.160605	0.129519	0.184015	0.146398
0.15	0.174914	0.151800	0.178628	0.154568	0.197357	0.168087
0.20	0.191161	0.175451	0.195257	0.178874	0.211774	0.191747
0.25	0.206616	0.198060	0.210428	0.201982	0.225851	0.214796
0.30	0.220869	0.219414	0.224894	0.223384	0.241047	0.237872
0.35	0.232773	0.238065	0.238514	0.243910	0.254138	0.259121
0.40	0.242599	0.254201	0.249982	0.261740	0.266261	0.278439
0.45	0.252932	0.270399	0.259515	0.276870	0.277532	0.295883
0.50	0.258661	0.281411	0.266743	0.289664	0.286215	0.311136
0.55	0.263807	0.291430	0.272634	0.300768	0.292647	0.322756
0.60	0.265233	0.296829	0.276465	0.308705	0.297787	0.332836
0.65	0.265266	0.300634	0.276603	0.312825	0.298265	0.337392
0.70	0.263710	0.301729	0.271952	0.310909	0.295369	0.337194
0.75	0.253199	0.293074	0.266337	0.306847	0.285841	0.329905
0.80	0.240586	0.281121	0.252732	0.294049	0.271348	0.316285
0.85	0.218076	0.257323	0.230615	0.271103	0.249634	0.293104
0.90	0.187181	0.222667	0.197833	0.234880	0.212656	0.252089
0.95	0.132226	0.159427	0.140040	0.168739	0.151303	0.181783
0.99	0.051079	0.062428	0.053840	0.065997	0.057369	0.070089

$(N - g)$ degrees of freedom. Thus, the bias and MSE of an estimator $\hat{\rho}(F^*)$ of ρ are defined as

$$\begin{aligned} \text{Bias}(\hat{\rho}, \rho) &= E[\hat{\rho}(F^*) - \rho] \text{ and } \text{MSE}(\hat{\rho}, \rho) \\ &= E\left[\left\{\hat{\rho}^2(F^*) - \rho\right\}^2\right], \end{aligned} \tag{10}$$

where the expectation is taken with respect to the distribution of F^* . However, the distributional result does not extend directly to unbalanced group size settings; see Searle et al., 1992. Because of the complexity of the estimation problem, analytical justifications of the theoretical discrepancies of the two estimators are generally not feasible. Thus, a large-scale simulation study is employed to assess the estimation properties of the prescribed truncated formulas $\hat{\rho}_{TA}$ and $\hat{\eta}_{TC}$ given in Eqs. (8) and (9), respectively.

Study design

To demonstrate the potential extent of characteristics that an applied work may reflect in clustering research, the number

of groups and average group size are set as $g = 4$ and 10 , and $\bar{n} = N/g = 5$ and 50 , respectively. Moreover, the group sizes are designated to have three different characteristics: equal, linear, and extreme structures so as to exemplify notable circumstances of practical importance. Specifically, the group sizes $\{n_i, i = 1, \dots, g\}$ for the combination of $g = 4$ and $N = 20$ ($\bar{n} = 5$) are:

- Equal structure: {5, 5, 5, 5};
- Linear structure: {2, 4, 6, 8};
- Extreme structure: {14, 2, 2, 2}.

In the case of $g = 4$ and $N = 200$ ($\bar{n} = 5$), the group sizes are chosen as:

- Equal structure: {50, 50, 50, 50};
- Linear structure: {20, 40, 60, 80};
- Extreme structure: {194, 2, 2, 2}.

On the other hand, the group sizes for the setting of $g = 10$ and $N = 50$ ($\bar{n} = 5$) are:

- Equal structure: {5, 5, 5, 5, 5, 5, 5, 5, 5, 5};
- Linear structure: {1, 2, 3, 4, 5, 5, 6, 7, 8, 9};
- Extreme structure: {32, 2, 2, 2, 2, 2, 2, 2, 2, 2}.

Table 6 The root mean squared error of ICC indices for $g = 4$ and $N = 200$ ($\bar{n} = 50$)

Group Size Structure						
ρ	Equal		Linear		Extreme	
	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$
0.00	0.013350	0.008141	0.014501	0.008901	0.119797	0.078759
0.05	0.049347	0.041277	0.051273	0.042786	0.130562	0.092691
0.10	0.078526	0.070023	0.082135	0.073059	0.148244	0.118187
0.15	0.103505	0.096094	0.107847	0.099914	0.167467	0.146282
0.20	0.124526	0.119775	0.129959	0.124637	0.186498	0.174265
0.25	0.144244	0.142323	0.149043	0.146963	0.205312	0.201024
0.30	0.160106	0.162077	0.166379	0.168416	0.222946	0.226225
0.35	0.175440	0.181537	0.181137	0.187437	0.238632	0.249005
0.40	0.187202	0.197743	0.194575	0.205328	0.253912	0.270426
0.45	0.198804	0.213728	0.205096	0.220308	0.266164	0.288560
0.50	0.207423	0.226570	0.215882	0.235208	0.278007	0.305756
0.55	0.214131	0.237218	0.223845	0.247168	0.285798	0.318484
0.60	0.220902	0.247812	0.229392	0.256710	0.290336	0.327359
0.65	0.223116	0.253217	0.231298	0.261747	0.294199	0.334536
0.70	0.222967	0.255745	0.231800	0.265094	0.290770	0.334144
0.75	0.218452	0.253016	0.227135	0.262532	0.285366	0.330067
0.80	0.209001	0.244480	0.219206	0.255252	0.271318	0.316553
0.85	0.193769	0.228458	0.203304	0.238759	0.249731	0.293567
0.90	0.167460	0.199434	0.176145	0.209235	0.214228	0.254353
0.95	0.120026	0.144552	0.127292	0.153045	0.153682	0.184417
0.99	0.046116	0.056583	0.049520	0.060565	0.058767	0.071774

The group structures are expanded to $N = 500$ with $g = 10$ and $\bar{n} = 50$ as follows:

Equal structure: {50, 50, 50, 50, 50, 50, 50, 50, 50, 50};

Linear structure: {10, 20, 30, 40, 50, 50, 60, 70, 80, 90};

Extreme structure: {482, 2, 2, 2, 2, 2, 2, 2, 2, 2}.

Note that the linear structures are approximately but not exactly linear in group sizes because of a minor adjustment made so that their sum meets the selected total number of observations. We also conducted a simulation study for $g = 100$ with $\bar{n} = 5$ and 50; however, the general phenomenon between the two indices $\hat{\rho}_{TA}$ and $\hat{\eta}_{TC}^2$ was similar to that of $g = 10$, and the discrepancies were rather small. To conserve space, the details are not provided here.

Accordingly, the performance of bias and MSE of $\hat{\rho}_{TA}$ and $\hat{\eta}_{TC}^2$ are examined for $\rho = 0$ to 0.95, with an increment of 0.05 and 0.99 for each of the 12 combined model configurations of two numbers of groups, two average group sizes, and three group size structures. Without loss of generality, the one-way random effects model with parameter values $\mu = 1$, $\sigma_\gamma^2 = \rho/(1 - \rho)$, and $\sigma_\varepsilon^2 = 1$ is used as the base for Monte Carlo assessment. With the selected model configurations, the estimates of bias and MSE defined in Eq. 10 are computed through

simulation of 100,000 replicate data sets. For each replicate, N observations are generated from the designated one-way random effects model. These values in turn determine the F^* statistic and the estimates $\hat{\rho}_{TA}$ and $\hat{\eta}_{TC}^2$. Then, the resulting errors $(\hat{\rho}_{TA} - \rho)$ and $(\hat{\eta}_{TC}^2 - \rho)$, and squared errors $(\hat{\rho}_{TA} - \rho)^2$ and $(\hat{\eta}_{TC}^2 - \rho)^2$, are computed. The simulated bias and MSE associated with the two ICC indices are the arithmetic means of the corresponding 100,000 replicated values. The computed biases are presented in Tables 1, 2, 3 and 4 for $(g, \bar{n}) = (4, 5)$, $(4, 50)$, $(10, 5)$, and $(10, 50)$, respectively. In addition, the corresponding root mean squared errors (RMSE = $\text{MSE}^{1/2}$) are summarized in Tables 5, 6, 7 and 8. These numerical results reveal unfamiliar and essential relations between the competing formulas.

Empirical results

Although it was noted previously that the $\hat{\eta}_{TC}^2$ index tends to underestimate ρ , it can be readily seen from Tables 1, 2, 3 and 4 that the modified version $\hat{\eta}_{TC}^2$ incurs slightly positive

Table 7 The root mean squared error of ICC indices for $g = 10$ and $N = 50$ ($\bar{n} = 5$)

Group Size Structure						
ρ	Equal		Linear		Extreme	
	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$
0.00	0.077203	0.063381	0.078705	0.064584	0.108642	0.089752
0.05	0.091038	0.078514	0.092546	0.079784	0.117766	0.100542
0.10	0.107809	0.098571	0.109342	0.099922	0.131370	0.118227
0.15	0.122816	0.116985	0.125067	0.118975	0.146871	0.138197
0.20	0.134992	0.132239	0.138483	0.135523	0.161665	0.157158
0.25	0.144663	0.144556	0.148650	0.148351	0.173859	0.173307
0.30	0.151184	0.153640	0.157313	0.159502	0.184608	0.187593
0.35	0.156988	0.161452	0.163122	0.167694	0.192277	0.198410
0.40	0.160238	0.166538	0.167632	0.174158	0.198188	0.207274
0.45	0.160818	0.168937	0.169142	0.177598	0.201242	0.212624
0.50	0.161210	0.170844	0.169768	0.179784	0.201992	0.215301
0.55	0.158012	0.168918	0.167367	0.178782	0.199958	0.214901
0.60	0.153497	0.165404	0.163898	0.176282	0.196233	0.212428
0.65	0.147126	0.159791	0.157069	0.170288	0.187682	0.204586
0.70	0.137813	0.150885	0.149000	0.162696	0.178417	0.195627
0.75	0.125649	0.138510	0.136353	0.149902	0.163600	0.180551
0.80	0.111141	0.123478	0.120557	0.133540	0.145377	0.161328
0.85	0.092836	0.103935	0.102261	0.114079	0.121473	0.135759
0.90	0.068708	0.077609	0.075967	0.085474	0.091298	0.102710
0.95	0.039268	0.044737	0.044259	0.050241	0.052779	0.059823
0.99	0.008913	0.010250	0.010148	0.011628	0.012127	0.013872

bias when ρ is near zero because of truncation for negative values. In general, however, it remains negatively biased, and the estimation behavior varies with group structures. Unlike the documented claim that modified ICC(1) is not affected by group size and the number of groups (Bliese & Halverson, 1998, p. 168), the performance of $\hat{\rho}_{TA}$ differs across the aforementioned four diverse group characteristics. Explicitly, the extensive empirical results showed that the performance of the truncated indices improves with an increasing number of groups and increasing average group size. On the other hand, it is noteworthy that their accuracy deteriorates as the group size allocation changes from equal, linear to extreme structures. The same phenomenon can be seen for the RMSE results in Tables 5, 6, 7 and 8 as well.

Furthermore, there is an important distinction between the two estimators. The absolute bias of $\hat{\eta}_{TC}^2$ is smaller than that of $\hat{\rho}_{TA}$ at very small values of ρ . Specifically, the simulated biases associated with four groups and 20 total observations in Table 1 show that $\text{Bias}(\hat{\eta}_{TC}^2, \rho) < \text{Bias}(\hat{\rho}_{TA}, \rho)$ for all three group-size structures when $\rho \leq 0.10$. In additional, Table 3 (10

groups and 50 total observations) shows the same dominance at $\rho = 0$ and 0.05 for equal and linear structures, and at $\rho \leq 0.10$ for extreme structure. In the cases of four and 10 groups with the average sample size $\bar{n} = 50$ of Tables 2 and 4, the situation occurs only at $\rho = 0$ for equal and linear structures, and at $\rho = 0$ and 0.05 for extreme structure. In contrast, it can be concluded that $\hat{\rho}_{TA}$ is consistently more accurate than $\hat{\eta}_{TC}^2$ at values of $\rho \geq 0.15$ for all three group size structures with respect to bias assessment. On the other hand, the computed RMSE results in Tables 5, 6, 7, and 8 basically reveal the same phenomenon that $\hat{\eta}_{TC}^2$ performs better than $\hat{\rho}_{TA}$ for small values of ρ , whereas the opposite is true for moderate and large values of ρ . However, the condition is slightly more prevalent than that in the evaluation of bias. For example, $\text{RMSE}(\hat{\eta}_{TC}^2, \rho)$ is less than $\text{RMSE}(\hat{\rho}_{TA}, \rho)$ when $\rho \leq 0.30$ for all three group size structures with the number of groups $g = 4$ and $N = 20$ in Table 5. The dominance situations only occurred at the values of $\rho \leq 0.25$ in Tables 6 and 7. Moreover, the results in Table 8 basically maintain almost the same pattern for $\rho \leq 0.25$ with the only exception at $\rho = 0.25$ for the extreme group size structure. Overall, the truncated

Table 8 The root mean squared error of ICC indices for $g = 10$ and $N = 500$ ($\bar{n} = 50$)

ρ	Group Size Structure					
	Equal		Linear		Extreme	
	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$	$\hat{\rho}_{TA}$	$\hat{\eta}_{TC}^2$
0.00	0.007438	0.005794	0.007657	0.005962	0.077493	0.061586
0.05	0.030336	0.028572	0.032134	0.030201	0.091051	0.077255
0.10	0.049069	0.046975	0.052162	0.049924	0.110191	0.100805
0.15	0.064754	0.062939	0.069436	0.067387	0.130046	0.124842
0.20	0.078536	0.077332	0.084285	0.082879	0.147426	0.145995
0.25	0.090571	0.090470	0.097413	0.097075	0.162264	0.164287
0.30	0.100408	0.101553	0.107773	0.108672	0.175663	0.180589
0.35	0.108305	0.110756	0.117004	0.119351	0.185883	0.193601
0.40	0.115122	0.119002	0.124124	0.128040	0.193650	0.203788
0.45	0.119736	0.125354	0.129420	0.134884	0.198600	0.210923
0.50	0.123002	0.129764	0.133404	0.140309	0.202120	0.216182
0.55	0.123577	0.131766	0.134870	0.143272	0.202617	0.218257
0.60	0.123818	0.133182	0.134154	0.143737	0.200319	0.217084
0.65	0.120291	0.130537	0.132308	0.143012	0.195680	0.213270
0.70	0.116203	0.127183	0.126309	0.137650	0.185344	0.203324
0.75	0.107326	0.118561	0.118666	0.130334	0.172579	0.190281
0.80	0.096749	0.107689	0.107089	0.118609	0.155069	0.172040
0.85	0.081501	0.091549	0.090588	0.101184	0.132144	0.147391
0.90	0.062190	0.070459	0.069551	0.078380	0.100334	0.112714
0.95	0.036210	0.041422	0.040916	0.046556	0.058804	0.066561
0.99	0.008223	0.009512	0.009527	0.010961	0.013900	0.015871

estimator $\hat{\eta}_{TC}^2$ has smaller RMSE for small values of ρ (≤ 0.20) for all group structures, whereas $\hat{\rho}_{TA}$ gives better RMSE for moderate and large values of ρ (≥ 0.35). Hence, the dominance relationship between $\hat{\rho}_{TA}$ and $\hat{\eta}_{TC}^2$ is closely related to group structures when the true value of ρ is between 0.20 and 0.35. In other words, each estimator is better only for certain combined configurations of ρ , numbers of groups, average group sizes, and group size structures.

Implication for intraclass correlation analysis

In the present article, we focused our attention on the two general formulas $\hat{\rho}_{TA}$ and $\hat{\eta}_{TC}^2$ for their appealing features of overall accuracy and computational ease. The detailed numerical investigation further helps differentiate the use of a proper ICC index from others under various circumstances. First, the ICC(1) or $\hat{\rho}_A$ index is used by authors and requested by reviewers almost reflexively in multilevel analysis. The empirical evidence demonstrates that further improvement may be obtained by adopting the corrected eta-squared estimator

when the magnitude of an intraclass correlation is small. Second, according to the body of accumulated knowledge in field research (Bliese, 2000), the underlying magnitudes of ICC are typically less than 0.20. Thus, $\text{Eta}^2(\text{C})$ or $\hat{\eta}_C^2$ appears to be a more appropriate estimate of the strength of intraclass correlation. With the reported values of F^* , the number of groups, and actual group sizes in actual practice, the corrected eta-squared estimators can be readily computed with the prescribed formulas. On the other hand, it can be shown that $\text{Eta}^2(\text{C})$ and $\hat{\eta}_C^2$ can be recovered directly from ICC(1) and $\hat{\rho}_A$ through the identity

$$\text{Eta}^2(\text{C}) = \frac{n \cdot \text{ICC}(1) - (1 - \text{ICC}(1))/(g - 1)}{n + (1 - \text{ICC}(1))(n - 1)/(g - 1)} \quad (11)$$

And

$$\hat{\eta}_C^2 = \frac{n_0 \hat{\rho}_A - (1 - \hat{\rho}_A)/(g - 1)}{n_0 + (1 - \hat{\rho}_A)(n_0 - 1)/(g - 1)}, \quad (12)$$

respectively. Thus, for example, an ICC(1) of 0.20 with the number of groups $g = 10$ and group size $n = 5$ translates into having an $\text{Eta}^2(\text{C})$ value of 0.17 with Eq. 11. Similarly, the

correspondence between $\hat{\rho}_A$ and $\hat{\eta}_C^2$ can be established through Eq. 12 for a given value n_0 . In short, Bliese and Halverson (1998) asserted that ICC(1) and Eta²(C) provide significant improvements over $\hat{\eta}^2$ for describing the effect of group membership. In the present article, we presented a comprehensive treatment to provide operational guidelines and practical implications for choosing between the two indices ICC (1) and Eta²(C) and their extended formulas.

Conclusion

The present article concerns the use of two ICC indices as strength of association measures for clustering studies. Despite their routine and common application in empirical studies, the fundamental properties of ICC indices are not sufficiently illustrated in the literature. The well-known ANOVA estimator and the previously suggested corrected eta-squared formulas were closely examined with respect to their point estimation principle, unbalanced data extension, and practical truncation consideration. In view of estimation principle, ICC(1) was obtained by substituting the variance components in population ICC with corresponding unbiased estimators. In contrast, the corrected eta-squared estimator was identical to the maximum likelihood estimator and the pairwise estimator under equal group sizes. Although their expressions were tractable and easy to compute, the intrinsic estimation complexity prevented analytical justification. Therefore, contemporary computer capabilities were used to conduct intensive simulation study for the bias and MSE performance of the two competing formulas. The numerical examinations revealed their critical and subtle discrepancy in estimating population ICC: Their estimation properties varied with the number of groups, average group size, and group size structure. Moreover, the modification of corrected eta-squared estimator performed better for small values of ICC, and the adjusted ICC(1) was preferred for moderate and large ICC values. Recognizing the different behavior of the two estimators helps to clarify the issue of evaluating the strength of the group property and how to choose an appropriate estimate in multilevel analysis. This information may be useful in selecting an appropriate measure of the intraclass correlation when a researcher has a basic conceptual idea about the underlying ICC in conjunction with the fundamental configurations of the obtained sample, such as the number of groups, average group size, and group size structure.

Author Note The author thanks Professor T. K. Peng of I-Shou University for helpful comments on an earlier version of this manuscript, and the action editor, Ira Bernstein, and two anonymous referees for constructive suggestions, which resulted in a clearer exposition.

References

- Alferes, V. R., & Kenny, D. A. (2009). SPSS programs for the measurement of nonindependence in standard dyadic designs. *Behavior Research Methods*, *41*, 47–54.
- Bartko, J. J. (1976). On various intraclass correlation reliability. *Psychological Bulletin*, *83*, 762–765.
- Beal, D. J., & Dawson, J. F. (2007). On the use of Likert-type scales in multilevel data: Influence on aggregate variables. *Organizational Research Methods*, *10*, 657–672.
- Bliese, P. D. (1998). Group size, ICC values and group-level correlations: A simulation. *Organizational Research Methods*, *1*, 355–373.
- Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). San Francisco, CA: Jossey-Bass.
- Bliese, P. D., & Halverson, R. R. (1998). Group size and measurements of group-level properties: An examination of eta-squared and ICC values. *Journal of Management*, *24*, 157–172.
- Castro, S. L. (2002). Data analytic methods for the analysis of multilevel questions: A comparison of intraclass correlation coefficients, $r_{wg}(f)$, hierarchical linear modeling, within- and between-analysis, and random group resampling. *The Leadership Quarterly*, *13*, 69–93.
- Cools, W., Van den Noortgate, W., & Onghena, P. (2009). Design efficiency for imbalanced multilevel data. *Behavior Research Methods*, *41*, 192–203.
- Courrieu, P., Brand-D'abrescia, M., Peereeman, R., Spieler, D., & Rey, A. (2011). Validated intraclass correlation statistics to test item performance models. *Behavior Research Methods*, *43*, 37–55.
- Donner, A. (1986). A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. *International Statistical Review*, *54*, 67–82.
- Donner, A., & Koval, J. J. (1980). The estimation of intraclass correlation in analysis of family data. *Biometrics*, *36*, 19–25.
- Donoghue, J. R., & Collins, L. M. (1990). A note on the unbiased estimation of the intraclass correlation. *Psychometrika*, *55*, 159–164.
- Fisher, R. A. (1938). *Statistical Methods for Research Workers* (7th ed.). Edinburgh, Scotland: Oliver and Boyd.
- Goldstein, H. (2002). *Multilevel statistical models* (3rd ed.). New York, NY: Wiley.
- Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*, 60–87.
- Hedges, L. V., & Rhoads, C. H. (2011). Correcting an analysis of variance for clustering. *British Journal of Mathematical and Statistical Psychology*, *64*, 20–37.
- Hoffman, L., & Rovine, M. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, *39*, 101–117.
- Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management*, *23*, 723–744.
- Hofmann, D. A. (2002). Issues in multilevel research: Theory development, measurement and analysis. In S. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 247–274). New York, NY: Blackwell.
- Hofmann, D. A., Griffin, M. A., & Gavin, M. B. (2000). The application of hierarchical linear modeling to organizational research. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 467–511). San Francisco, CA: Jossey-Bass.
- Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.),

- Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 467–511). San Francisco, CA: Jossey-Bass.
- LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*, 815–852.
- Maxwell, S. E., Camp, C. J., & Arvey, R. D. (1981). Measures of strength of association: A comparative examination. *Journal of Applied Psychology, 66*, 525–534.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30–46.
- Murray, D. M., Varnell, S. P., & Blitstein, J. L. (2004). Design and analysis of group-randomized trials: A review of recent methodological developments. *American Journal of Public Health, 94*, 423–432.
- O'Connor, B. P. (2004). SPSS and SAS programs for addressing interdependence and basic levels-of-analysis issues in psychological data. *Behavior Research Methods, 36*, 17–28.
- Olkin, I., & Pratt, J. W. (1958). Unbiased estimation of certain correlation coefficients. *Annals of Mathematical Statistics, 29*, 201–211.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance Components*. New York, NY: Wiley.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, England: Sage.
- Swiger, L. A., Harvey, W. R., Everson, D. O., & Gregory, K. E. (1964). The variance of intraclass correlation involving groups with one observation. *Biometrics, 20*, 818–826.
- Tasoluk, B., Droge, C., & Calantone, R. J. (2009). Interpreting interrelations across multiple levels in HGLM models: An application in international marketing research. *International Marketing Review, 28*, 34–56.
- Wampold, B. E., & Serlin, R. C. (2000). The consequence of ignoring a nested factor on measures of effect size in analysis of variance. *Psychological Methods, 5*, 425–433.