ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2026, Volume 21, Issue 1

Simulative and empirical comparison of health scales' scores derived by confirmatory factor analysis and sum of items

Cristian Ricci[(1)], *Sara Isernia*[(2)], *Stefano Terzoni*[(3)], *Francesca Baglio*[(2)], *Carlo La Vecchia*[(4)]

(1) Africa Unit for Transdisciplinary Health Research (AUTHeR), North-West University, Potchefstroom, South Africa (ROR: 010f1sq29)
(2) IRCCS Fondazione Don Carlo Gnocchi ONLUS, Milan, Italy (ROR: 02e3ssq97)
(3) Department of Biomedical Sciences for Health, University of Milan, Milan, Italy (ROR: 00wjc7c48)
(4) Department of Clinical Sciences and Community Health, Faculty of Medicine, University of Milan, Milan, Italy (ROR: 00wjc7c48)

CORRESPONDING AUTHOR: Cristian Ricci, Africa Unit for Transdisciplinary Health Research (AUTHeR), North-West University, Potchefstroom, South Africa. Email: cristian.ricci@nwu.ac.za

SUMMARY

We focus on how combining questionnaire's items to derive a score. Data simulation and real data were used to compare questionnaires' scores derived by sum of items and scores derived from a confirmatory factor analysis. We investigated the equivalence of the sum of items and the latent factor scores from confirmatory factor analysis in simulative and empirical studies. We investigated to what extent the derivation of the score may have influenced the assumptions of linear model analysis. Deriving the questionnaire's scores using the sum of items or the application of the confirmatory factor analysis is equivalent. This evidence is confirmed in the simulation settings and by empirical data.

Keywords: Health scales; psychometric scores; sum of items; confirmatory factor analysis; simulation study

INTRODUCTION

The use of scales, questionnaires and tests is common in many fields of health research, social sciences, and behavioural research [1]. The spread of these tools' use in health research is justified for these measures are an apparently simple instrument and their use seems as limitless [1–3]. Moreover, numerous valid and reliable scales, are available nowadays. These scales can be administered in electronic form making the collection of large amounts data easier than ever [4]. However, using scales in health research is not a simple task; especially challenging is the proper application of statistical analysis tools [3,5–7]. Scales are composed of items investigating specific aspects and are generally coded as multiple-choice questions. How combining such items to derive a suitable score is the matter of our discussion. There are two ways to merge results from different items performing an overall score. On the one hand, the simpler and most common approach is to sum the scores (SS) from the single items. On the other, the overall score could be computed via a confirmatory factor analysis or by similar approaches aimed to derive a latent factor score of the items (LFS).

The two methods are not theoretically interchangeable [8,9]. On the one hand, Mc Neish and Wolf [8] stated that sum scoring has a negative effect on the validity, reliability and qualitative classification from sum-score cut-off. On the other hand, Widaman and Revelle [9] reciprocated that sum scores are valid and reliable, especially when the dimensional structure of a set of items is well established. Furthermore, Widaman and Revelle highlight that the sum score is a more general solution since it does not depend on the analytical sample, while the factor score does.

While evidence from both viewpoints supports choosing one method over the other, some practical aspects still require further investigation. Firstly, it is of interest to understand in which condition the two approaches are equivalent and how this situation can be identified. Afterwards, we used real data from a rehabilitation registry, the "Smart&Touch-ID registry" [10], in which vertain tools were used to evaluate

DOI: 10.54103/2282-0930/29512
Accepted: 9th December 2025

the usability of digital rehabilitation solutions (System Usability Scale) and the efficacy on clinical outcome (the Yoni task). Moreover, we showed regression efficiency and showed how to improve the validity of linear model assumptions.

METHODS

Theoretical background

It is convenient to highlight that sum of items' scores (SS) and the latent factor score (LFS) are analytically similar for they are both a linear combination of the items. The following equation represents a linear combination of X_p elements with w_p being weights.

$$Y = w_1 X_1 + w_2 X_2 + \dots + w_p X_p \qquad (1)$$

In other words, the element Y is a score computed by p elements X, not all the p elements equally contribute to Y so the w_i (with $0 \leq i \leq 1$) element represent the contribution (generally referred as weight) of X_i on Y. Here the first difference between the SS and the LFS. In the SS, the weights are equal and assume the value of 1. In the LFS these weights differ, depending on the item and by its correlation with the latent factor.

The variance of Y can be represented by the following equation.

$$\sigma_Y^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + \dots + w_p^2\sigma_p^2 + 2\, w_1 w_2 r_{1,2}\sigma_1\sigma_2 + 2\, w_1 w_3 r_{1,3}\sigma_1\sigma_3 + \dots + 2\, w_{p-1} w_p r_{p-1,p}\sigma_{p-1}\sigma_p \qquad (2)$$

Where σ^2_i represents the variance of a given item i and $r_{jk}\sigma_j\sigma_k$ represents the covariance between the items j and k. This formulation is common to overall scores computed as the sum of items (here referred as the X) and to overall scores computed using latent factor analysis. Specifically, in the SS the weight element corresponds to 1 for all the items. For SS the variance is given by the sum of the items' variance plus items' covariance.

For the LFS, the expected value and its variance are thus similar to the one of SS. In summary, even if performed by different procedures, the SS and the LFS are similarly interpreted for they are both linear combination of the items. The difference between the two is the weight for a single item. The LFS is computed considering different weights for different items. These weights, often referred as factor loadings, are proportional to the correlation between the variables and the latent factor. There is also a difference between the expected value and the variance performed by these two methods. This numerical difference between the sum score and the factor score is again due to the numerical value of the weight.

Simulation studies

The simulation studies were conducted using the R software vers. 4.0.2. Briefly, we firstly simulated one latent factor (SLF) following the Z~N (0;1) distribution using the MVRnorm function from the MASS package [11]. Afterwards, five, nine and 17 items following the Z~N(0;1) distribution were generated having a given correlation pattern with the above mentioned latent factor using the rnorm_pre function of the faux package [12]. We simulated items having a medium to strong correlation to the latent factor, namely a correlation coefficient of 0.45, 0.6 and 0.75. We then simulated different scenarios to investigate to what extent the correlation between the items and the latent factors may have influenced the analyses. Firstly, we simulated iso-correlated variables. In a second set of simulations, we used variables with a range of correlations but still with a central or median correlation value of 0.45, 0.6 and 0.75. Specifically, in the five-variable setting the correlations between the items and the latent factor ranged between 0.35 to 0.55, by steps of 0.05 for the setting with a median correlation of 0.45. Similarly, for the settings with a median correlation of 0.6 (r = 0.5 to 0.7 by 0.05) and for the setting with a median correlation of 0.75 (r = 0.65 to 0.85 by 0.05). In the simulations with nine and 17 variables we adopted the same ranges with steps of 0.025 and 0.0125, respectively. A second set of simulation was conducted to investigate the effects of having skewed variables defined as ordinal categories, according to most common operative situations occurring when analysing data from questionnaires. To this aim, we firstly simulated a continuous rightly skewed latent factor using the rs_norm function from the fGarch package [13]. Afterwards, a series of five level Likert variables were simulated using the formula $V[i] = LAF \cdot r[i] + LAF \cdot \sqrt{(1-r[i]^2)}$, where r[i] represents the correlation coefficient between the variable and the latent factor. The function parameters were set to obtain items with a skewness greater than 1, defining highly skewed items. Finally, each simulation described above was performed with a sample size of 100, 250, 500 and 1000 observations to investigate to what extent the sample size may have influenced the relation between the SS and the LFS. The simulated scenarios are reported on Table 1, the simulation programs for normally distributed and skewed data were reported on Supplementary Files.

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2026, Volume 21, Issue 1

Table 1. Representation of the not iso-correlated simulation scenarios

	$M_{correlation}$ = 0.45	$M_{correlation}$ = 0.60	$M_{correlation}$ = 0.75
Item number	Correlation between the items and the latent factor		
1	0.3500	0.5000	0.6500
2	0.4000	0.5500	0.7000
3	0.4500	0.6000	0.7500
4	0.5000	0.6500	0.8000
5	0.5500	0.7000	0.8500
Item number			
1	0.3500	0.5000	0.6500
2	0.3750	0.5250	0.6750
3	0.4000	0.5500	0.7000
4	0.4250	0.5750	0.7250
5	0.4500	0.6000	0.7500
6	0.4750	0.6250	0.7750
7	0.5000	0.6500	0.8000
8	0.5250	0.6750	0.8250
9	0.5500	0.7000	0.8500
Item number			
1	0.3500	0.5000	0.6500
2	0.3625	0.5125	0.6625
3	0.3750	0.5250	0.6750
4	0.3875	0.5375	0.6875
5	0.4000	0.5500	0.7000
6	0.4125	0.5625	0.7125
7	0.4250	0.5750	0.7250
8	0.4375	0.5875	0.7375
9	0.4500	0.6000	0.7500
10	0.4625	0.6125	0.7625
11	0.4750	0.6250	0.7750
12	0.4875	0.6375	0.7875
13	0.5000	0.6500	0.8000
14	0.5125	0.6625	0.8125
15	0.5250	0.6750	0.8250
16	0.5375	0.6875	0.8375
17	0.5500	0.7000	0.8500

$M_{correlation}$: represents the median or central value of the correlation series

A supplementary set of simulation has been performed where 10% of times where randomly assigned to a null correlation (Results reported on Supplementary Tables 1-4).

Empirical studies

Firstly, we investigated data obtained by the administration of the System Usability Scale (SUS), an instrument made by ten multiple choice items with response coded by a five points Likert scale (1=strongly disagree, 5=strongly agree). The SUS is a valid tool for the measure of usability of a wide range of technological systems or devices. The score of the SUS ranges from 0 to 100 and indicates the overall perceived usability of a technological system. In this study, the SUS overall scores were scored according to Brooke [14]. Notably, the SUS score computed according to Brooke is a SS-like score.

The SUS also considers two subscales or main aspects that affect the user's experience toward a technological system: the usability and the learnability subscales. The first indicates the ease of use perceived by the user during the interaction with the technology. Instead, the learnability subscale refers to the perceived ease of learning of using the technological system. Learnability and usability subscales were obtained in accordance to Lewis and Sauro's indications[15,16].

Secondly, the Yoni tasktool was investigated, we used the Yoni-48 which is the short version of the Italian Yoni task for the assessment of social cognition ability, and especially Theory of Mind (ToM)[17–19]. The Yoni-48 task includes 48 items in total, of those items 42 assesses ToM (mental items) and six are control items (physical items). The mental stimuli comprise 21 affective and 21 cognitive ToM items.

The task is digital and includes static visual cartoon-like stimuli, in which a face named "Yoni" appeared at the centre of the screen, surrounded by four coloured elements (for example fruit, animals, means of transport, or faces). Subjects are instructed to click as fast as they can on the elements Yoni refers to, based on the sentence reported at the top of the screen (e.g., "Yoni is thinking of ...", "Yoni likes ..."). For each item, only one answer is correct (score 0–1). Yoni total raw score, and sub-scores (such as affective and cognitive raw score), are computed by summing items.

Statistical analysis

The LFS and the SS were performed for all simulated scenarios and for the empirical sets of data. The comparison between the simulated latent factor, the LFS and SS was qualitatively evaluated by visual inspection of the scatterplots and quantitatively reported by the calculation of the correlation coefficients. In the simulation study, the bootstrap Pearson correlation coefficients with 95% confidence limits between the simulated latent factor, the latent factor from the confirmatory factor analysis and the score performed by the sum of items were reported using the ci_cor function for normally distributed and continuous data. The Spearman correlation coefficient was adopted when the items were simulated as skewed and ordinal categorical. The LAVAAN package [20] of the R software was used to perform the LFS according to the codes reported on Supplementary files. Briefly, the items were considered as continuous variables when analysing the data from the simulation study with items simulated as continuous variables and for the empirical study conducted using the SUS for it is a CTT instrument. In contrast, for the simulated settings with skewed and categorical variables and for the Yoni-48, the LFS was computed using a code considering the true/false dichotomous nature of the Yoni items using the ordered="TRUE" option of the CFA function of LAVAAN.

RESULTS

Simulation study. Comparison between the latent factor scores and the scores performed by items' sum.

In the first simulation setting we set the items' weight as constant being associated to the SLF factor with a fixed correlation coefficient between the items and the latent factor. In this first simulation setting we confirmed that the SS and the LFS are asymptotically equal in the iso-correlated scenarios when the items are continuous and normally distributed. The correlations between the SS and the LFS have very high values being above 0.95 also for small sample sizes and number of items (Table 2).

Table 2. Pearson correlation coefficients and 95% confidence limits obtained by bias corrected bootstrap with continuous iso-correlated items distributed as a standardized normal distribution

	Bootstrap Pearson correlation coefficients with 95% confidence limits		
Simulation Scenarios	LAF with LFS	LAF with SS	LFS with SS
V = 5, N = 100, r = 0.45	0.6877 (0.5931; 0.7625)	0.7390 (0.6540; 0.8018)	0.9588 (0.9410; 0.9699)
V = 9, N = 100, r = 0.45	0.8071 (0.7403; 0.8584)	0.8267 (0.7640; 0.8745)	0.9912 (0.9874; 0.9936)
V = 17, N = 100, r = 0.45	0.8560 (0.8087; 0.8901)	0.8721 (0.8282; 0.9043)	0.9971 (0.9961; 0.9978)
V = 5, N = 250, r = 0.45	0.7114 (0.6623; 0.7537)	0.7158 (0.6668; 0.7579)	0.9973 (0.9966; 0.9978)
V = 9, N = 250, r = 0.45	0.8299 (0.7970; 0.8570)	0.8387 (0.8064; 0.8645)	0.9970 (0.9965; 0.9975)
V = 17, N = 250, r = 0.45	0.8829 (0.8607; 0.9015)	0.8877 (0.8659; 0.9059)	0.9992 (0.9990; 0.9993)
V = 5, N = 500, r = 0.45	0.7563 (0.7216; 0.7879)	0.7595 (0.7248; 0.7914)	0.9979 (0.9976; 0.9982)
V = 9, N = 500, r = 0.45	0.8264 (0.7997; 0.8496)	0.8291 (0.8022; 0.8523)	0.9990 (0.9988; 0.9991)
V = 17, N = 500, r = 0.45	0.8931 (0.8774; 0.9081)	0.8973 (0.8820; 0.9111)	0.9992 (0.9991; 0.9993)
V = 5, N = 1000, r = 0.45	0.7431 (0.7191; 0.7663)	0.7433 (0.7190; 0.7660)	0.9998 (0.9998; 0.9998)
V = 9, N = 1000, r = 0.45	0.8304 (0.8134; 0.8460)	0.8324 (0.8157; 0.8481)	0.9992 (0.9992; 0.9993)
V = 17, N = 1000, r = 0.45	0.8961 (0.8853; 0.9058)	0.8968 (0.8860; 0.9066)	0.9999 (0.9999; 0.9999)
V = 5, N = 100, r = 0.60	0.8350 (0.7846; 0.8726)	0.8528 (0.8064; 0.8877)	0.9904 (0.9864; 0.9928)
V = 9, N = 100, r = 0.60	0.9032 (0.8665; 0.9300)	0.9094 (0.8740; 0.9351)	0.9982 (0.9974; 0.9987)
V = 17, N = 100, r = 0.60	0.9299 (0.9052; 0.9471)	0.9357 (0.9122; 0.9522)	0.9993 (0.9991; 0.9995)
V = 5, N = 250, r = 0.60	0.8344 (0.8050; 0.8594)	0.8364 (0.8068; 0.8612)	0.9992 (0.9990; 0.9993)
V = 9, N = 250, r = 0.60	0.9136 (0.8958; 0.9276)	0.9165 (0.8994; 0.9302)	0.9993 (0.9992; 0.9994)
V = 17, N = 250, r = 0.60	0.9427 (0.9315; 0.9519)	0.9443 (0.9335; 0.9535)	0.9998 (0.9998; 0.9998)
V = 5, N = 500, r = 0.60	0.8656 (0.8453; 0.8840)	0.8667 (0.8464; 0.8848)	0.9995 (0.9994; 0.9996)
V = 9, N = 500, r = 0.60	0.9100 (0.8955; 0.9223)	0.9109 (0.8962; 0.9233)	0.9998 (0.9997; 0.9998)
V = 17, N = 500, r = 0.60	0.9481 (0.9404; 0.9554)	0.9495 (0.9418; 0.9563)	0.9998 (0.9998; 0.9999)
V = 5, N = 1000, r = 0.60	0.8557 (0.8412; 0.8696)	0.8558 (0.8411; 0.8693)	0.9998 (0.9998; 0.9999)
V = 9, N = 1000, r = 0.60	0.9122 (0.9029; 0.9204)	0.9129 (0.9039; 0.9212)	1 (1; 1)
V = 17, N = 1000, r = 0.60	0.9490 (0.9436; 0.9540)	0.9492 (0.9437; 0.9541)	1 (1; 1)
V = 5, N = 100, r = 0.75	0.9205 (0.8964; 0.9376)	0.9268 (0.9037; 0.9436)	0.9974 (0.9965; 0.9980)
V = 9, N = 100, r = 0.75	0.9550 (0.9374; 0.9677)	0.9572 (0.9398; 0.9694)	0.9996 (0.9994; 0.9997)
V = 17, N = 100, r = 0.75	0.9682 (0.9568; 0.9762)	0.9703 (0.9592; 0.9780)	0.9998 (0.9997; 0.9999)
V = 5, N = 250, r = 0.75	0.9167 (0.9016; 0.9297)	0.9175 (0.9020; 0.9303)	0.9997 (0.9997; 0.9998)
V = 9, N = 250, r = 0.75	0.9597 (0.9512; 0.9663)	0.9608 (0.9526; 0.9673)	0.9998 (0.9998; 0.9999)
V = 17, N = 250, r = 0.75	0.9739 (0.9687; 0.9781)	0.9744 (0.9694; 0.9787)	1 (0.9999; 1)
V = 5, N = 500, r = 0.75	0.9342 (0.9237; 0.9434)	0.9346 (0.9244; 0.9438)	0.9999 (0.9999; 0.9999)
V = 9, N = 500, r = 0.75	0.9576 (0.9507; 0.9634)	0.9579 (0.9509; 0.9638)	0.9999 (0.9999; 0.9999)
V = 17, N = 500, r = 0.75	0.9765 (0.9730; 0.9797)	0.9769 (0.9734; 0.9800)	1 (1; 1)
V = 5, N = 1000, r = 0.75	0.9285 (0.9210; 0.9355)	0.9285 (0.9209; 0.9354)	1 (1; 1)
V = 9, N = 1000, r = 0.75	0.9587 (0.9542; 0.9626)	0.9589 (0.9547; 0.9629)	1 (1; 1)
V = 17, N = 1000, r = 0.75	0.9767 (0.9742; 0.9790)	0.9768 (0.9742; 0.9791)	1 (1; 1)

LAF = Simulated latent factor; LFS = Latent factor score performed via confirmative factor analysis; SS: Score performed by the sum of items; V = Number of items; N = Sample size; r = Pearson correlation coefficient between the items

The correlations between the simulated latent factor with the SS and the LFS are not particularly high with correlations in the range of 0.7-0.8 for scenarios with small sample sizes and number of variables. We observed the same pattern when the items had different correlations with the simulated latent factors (Table 3).

Table 3. Pearson correlation coefficients and 95% confidence limits obtained by bias corrected bootstrap with continuous not iso-correlated items distributed as a standardized normal distribution

	Bootstrap Pearson correlation coefficients with 95% confidence limits		
Simulation Scenarios	LAF with LFS	LAF with SS	LFS with SS
V = 5, N = 100, r = 0.45	0.7025 (0.6088; 0.7741)	0.7395 (0.6552; 0.8024)	0.9495 (0.9291; 0.9638)
V = 9, N = 100, r = 0.45	0.8188 (0.7551; 0.8679)	0.8274 (0.7651; 0.8748)	0.9902 (0.9866; 0.9927)
V = 17, N = 100, r = 0.45	0.8631 (0.8175; 0.8954)	0.8720 (0.8283; 0.9043)	0.9948 (0.9928; 0.9962)
V = 5, N = 250, r = 0.45	0.7233 (0.6753; 0.7649)	0.7176 (0.6693; 0.7594)	0.9873 (0.9845; 0.9896)
V = 9, N = 250, r = 0.45	0.8361 (0.8045; 0.8618)	0.8390 (0.8069; 0.8650)	0.9901 (0.9879; 0.9918)
V = 17, N = 250, r = 0.45	0.8886 (0.8679; 0.9067)	0.8880 (0.8662; 0.9060)	0.9956 (0.9946; 0.9964)
V = 5, N = 500, r = 0.45	0.7692 (0.7359; 0.7996)	0.7600 (0.7257; 0.7919)	0.9920 (0.9907; 0.9932)
V = 9, N = 500, r = 0.45	0.8321 (0.8057; 0.8546)	0.8300 (0.8033; 0.8531)	0.9905 (0.9890; 0.9919)
V = 17, N = 500, r = 0.45	0.8948 (0.8793; 0.9095)	0.8982 (0.8832; 0.9120)	0.9941 (0.9931; 0.9950)
V = 5, N = 1000, r = 0.45	0.7536 (0.7304; 0.7759)	0.7445 (0.7202; 0.7670)	0.9903 (0.9892; 0.9913)
V = 9, N = 1000, r = 0.45	0.8354 (0.8188; 0.8502)	0.8336 (0.8168; 0.8491)	0.9907 (0.9897; 0.9916)
V = 17, N = 1000, r = 0.45	0.9010 (0.8907; 0.9103)	0.8970 (0.8862; 0.9070)	0.9964 (0.9960; 0.9968)
V = 5, N = 100, r = 0.60	0.8451 (0.7966; 0.8808)	0.8531 (0.8073; 0.8878)	0.9821 (0.9754; 0.9870)
V = 9, N = 100, r = 0.60	0.9099 (0.8758; 0.9351)	0.9100 (0.8745; 0.9352)	0.9959 (0.9945; 0.9969)
V = 17, N = 100, r = 0.60	0.9344 (0.9112; 0.9507)	0.9355 (0.9121; 0.9520)	0.9976 (0.9967; 0.9982)
V = 5, N = 250, r = 0.60	0.8435 (0.8156; 0.8678)	0.8382 (0.8094; 0.8629)	0.9917 (0.9899; 0.9931)
V = 9, N = 250, r = 0.60	0.9179 (0.9010; 0.9311)	0.9167 (0.8995; 0.9304)	0.9951 (0.9940; 0.9960)
V = 17, N = 250, r = 0.60	0.9459 (0.9356; 0.9547)	0.9445 (0.9336; 0.9536)	0.9975 (0.9969; 0.9980)
V = 5, N = 500, r = 0.60	0.8748 (0.8553; 0.8920)	0.8670 (0.8469; 0.8851)	0.9937 (0.9927; 0.9946)
V = 9, N = 500, r = 0.60	0.9140 (0.8999; 0.9258)	0.9116 (0.8970; 0.9238)	0.9949 (0.9940; 0.9956)
V = 17, N = 500, r = 0.60	0.9494 (0.9417; 0.9564)	0.9502 (0.9427; 0.9570)	0.9972 (0.9967; 0.9976)
V = 5, N = 1000, r = 0.60	0.8638 (0.8501; 0.8768)	0.8568 (0.8424; 0.8703)	0.9930 (0.9922; 0.9937)
V = 9, N = 1000, r = 0.60	0.9157 (0.9069; 0.9234)	0.9138 (0.9048; 0.9221)	0.9950 (0.9944; 0.9955)
V = 17, N = 1000, r = 0.60	0.9519 (0.9468; 0.9566)	0.9493 (0.9438; 0.9543)	0.9979 (0.9976; 0.9981)
V = 5, N = 100, r = 0.75	0.9306 (0.9087; 0.9460)	0.9271 (0.9049; 0.9435)	0.9891 (0.9854; 0.9919)
V = 9, N = 100, r = 0.75	0.9604 (0.9450; 0.9716)	0.9576 (0.9402; 0.9696)	0.9967 (0.9956; 0.9976)
V = 17, N = 100, r = 0.75	0.9719 (0.9617; 0.9790)	0.9701 (0.9591; 0.9778)	0.9982 (0.9975; 0.9986)
V = 5, N = 250, r = 0.75	0.9261 (0.9127; 0.9378)	0.9193 (0.9040; 0.9319)	0.9927 (0.9911; 0.9939)
V = 9, N = 250, r = 0.75	0.9638 (0.9560; 0.9697)	0.9608 (0.9527; 0.9673)	0.9960 (0.9951; 0.9967)
V = 17, N = 250, r = 0.75	0.9765 (0.9719; 0.9803)	0.9745 (0.9694; 0.9788)	0.9978 (0.9973; 0.9982)
V = 5, N = 500, r = 0.75	0.9426 (0.9334; 0.9506)	0.9346 (0.9245; 0.9438)	0.9931 (0.9920; 0.9941)
V = 9, N = 500, r = 0.75	0.9615 (0.9552; 0.9668)	0.9585 (0.9516; 0.9642)	0.9957 (0.9950; 0.9963)
V = 17, N = 500, r = 0.75	0.9779 (0.9746; 0.9810)	0.9775 (0.9742; 0.9805)	0.9978 (0.9975; 0.9981)
V = 5, N = 1000, r = 0.75	0.9368 (0.9301; 0.9431)	0.9294 (0.9220; 0.9363)	0.9931 (0.9922; 0.9938)
V = 9, N = 1000, r = 0.75	0.9623 (0.9583; 0.9657)	0.9597 (0.9555; 0.9636)	0.9959 (0.9954; 0.9963)
V = 17, N = 1000, r = 0.75	0.9791 (0.9768; 0.9811)	0.9769 (0.9744; 0.9792)	0.9981 (0.9979; 0.9983)

LAF = Simulated latent factor; LFS = Latent factor score performed via confirmative factor analysis; SS: Score performed by the sum of items; V = Number of items; N = Sample size; r = Pearson correlation coefficient between the items

Most notably, having skewed ordinal categorical data does not substantially change the results reported above (Tables 4 and 5).

Table 4. Spearman correlation coefficients and 95% confidence limits obtained by bias corrected bootstrap with five level Likert iso-correlated items with positive skewness

	Bootstrap Pearson correlation coefficients with 95% confidence limits		
Simulation Scenarios	LAF with LFS	LAF with SS	LFS with SS
V = 5, N = 100, r = 0.45	0.5714 (0.4399; 0.6812)	0.6259 (0.5072; 0.7202)	0.9617 (0.9358; 0.9762)
V = 9, N = 100, r = 0.45	0.7340 (0.6461; 0.8064)	0.7460 (0.6550; 0.8174)	0.9630 (0.9508; 0.9753)
V = 17, N = 100, r = 0.45	0.7392 (0.6445; 0.8118)	0.7402 (0.6457; 0.8128)	0.9725 (0.9592; 0.9827)
V = 5, N = 250, r = 0.45	0.5683 (0.4882; 0.6397)	0.5879 (0.5116; 0.6548)	0.9420 (0.9243; 0.9553)
V = 9, N = 250, r = 0.45	0.7025 (0.6401; 0.7565)	0.6985 (0.6377; 0.7504)	0.9759 (0.9689; 0.9818)
V = 17, N = 250, r = 0.45	0.7893 (0.7369; 0.8327)	0.8005 (0.7507; 0.8406)	0.9836 (0.9783; 0.9881)
V = 5, N = 500, r = 0.45	0.6737 (0.6242; 0.7165)	0.6655 (0.6158; 0.7093)	0.9847 (0.9815; 0.9874)
V = 9, N = 500, r = 0.45	0.7027 (0.6588; 0.7414)	0.6948 (0.6504; 0.7337)	0.9901 (0.9877; 0.9921)
V = 17, N = 500, r = 0.45	0.8294 (0.8014; 0.8535)	0.8218 (0.7935; 0.8472)	0.9919 (0.9905; 0.9934)
V = 5, N = 1000, r = 0.45	0.6026 (0.5638; 0.6394)	0.5955 (0.5570; 0.6314)	0.9896 (0.9885; 0.9908)
V = 9, N = 1000, r = 0.45	0.7436 (0.7175; 0.7683)	0.7358 (0.7080; 0.7606)	0.9922 (0.9910; 0.9932)
V = 17, N = 1000, r = 0.45	0.7993 (0.7756; 0.8206)	0.7934 (0.7702; 0.8146)	0.9927 (0.9915; 0.9938)
V = 5, N = 100, r = 0.60	0.7262 (0.6375; 0.7988)	0.7470 (0.6611; 0.8146)	0.9767 (0.9659; 0.9849)
V = 9, N = 100, r = 0.60	0.8321 (0.7669; 0.8833)	0.8371 (0.7695; 0.8873)	0.9737 (0.9541; 0.9845)
V = 17, N = 100, r = 0.60	0.8562 (0.7927; 0.9017)	0.8507 (0.7840; 0.8959)	0.9908 (0.9855; 0.9947)
V = 5, N = 250, r = 0.60	0.7317 (0.6694; 0.7841)	0.7354 (0.6748; 0.7865)	0.9763 (0.9688; 0.9822)
V = 9, N = 250, r = 0.60	0.8127 (0.7648; 0.8517)	0.8082 (0.7618; 0.8470)	0.9922 (0.9903; 0.9941)
V = 17, N = 250, r = 0.60	0.8769 (0.8424; 0.9037)	0.8776 (0.8437; 0.9040)	0.9960 (0.9949; 0.9973)
V = 5, N = 500, r = 0.60	0.7907 (0.7540; 0.8227)	0.7820 (0.7439; 0.8139)	0.9926 (0.9912; 0.994)
V = 9, N = 500, r = 0.60	0.8188 (0.7900; 0.8447)	0.8128 (0.7835; 0.8387)	0.9945 (0.9934; 0.9955)
V = 17, N = 500, r = 0.60	0.9001 (0.8802; 0.9164)	0.8938 (0.8741; 0.9115)	0.9963 (0.9957; 0.9971)
V = 5, N = 1000, r = 0.60	0.7308 (0.7013; 0.7591)	0.7333 (0.7040; 0.7598)	0.9908 (0.9897; 0.9921)
V = 9, N = 1000, r = 0.60	0.8393 (0.8207; 0.8572)	0.8358 (0.8157; 0.8531)	0.9961 (0.9956; 0.9967)
V = 17, N = 1000, r = 0.60	0.8806 (0.8652; 0.8946)	0.8795 (0.8640; 0.8933)	0.9973 (0.9969; 0.9977)
V = 5, N = 100, r = 0.75	0.8092 (0.7312; 0.8683)	0.8269 (0.7548; 0.8798)	0.9779 (0.9658; 0.9866)
V = 9, N = 100, r = 0.75	0.9144 (0.8739; 0.9427)	0.9114 (0.8672; 0.9406)	0.9939 (0.9906; 0.9966)
V = 17, N = 100, r = 0.75	0.8989 (0.8538; 0.9329)	0.9022 (0.8545; 0.9344)	0.9935 (0.9883; 0.9965)
V = 5, N = 250, r = 0.75	0.8230 (0.7765; 0.8611)	0.8338 (0.7892; 0.8688)	0.9920 (0.9894; 0.9944)
V = 9, N = 250, r = 0.75	0.8762 (0.8393; 0.9059)	0.8794 (0.8439; 0.9072)	0.9958 (0.9943; 0.9971)
V = 17, N = 250, r = 0.75	0.9323 (0.9118; 0.9484)	0.9324 (0.9117; 0.9485)	0.9972 (0.9963; 0.9981)
V = 5, N = 500, r = 0.75	0.8689 (0.8445; 0.8898)	0.8701 (0.8462; 0.8905)	0.9933 (0.9915; 0.9949)
V = 9, N = 500, r = 0.75	0.8856 (0.8650; 0.9039)	0.8855 (0.8645; 0.9033)	0.9962 (0.9953; 0.997)
V = 17, N = 500, r = 0.75	0.9415 (0.9288; 0.9521)	0.9407 (0.9282; 0.9514)	0.9980 (0.9975; 0.9985)
V = 5, N = 1000, r = 0.75	0.8235 (0.8012; 0.8436)	0.8262 (0.8049; 0.8452)	0.9930 (0.9917; 0.9942)
V = 9, N = 1000, r = 0.75	0.9042 (0.8916; 0.9162)	0.9032 (0.8899; 0.9148)	0.9971 (0.9967; 0.9976)
V = 17, N = 1000, r = 0.75	0.9270 (0.9170; 0.9366)	0.9274 (0.9172; 0.9368)	0.9983 (0.9981; 0.9986)

LAF = Simulated latent factor; LFS = Latent factor score performed via confirmative factor analysis; SS: Score performed by the sum of items; V = Number of items; N = Sample size; r = Pearson correlation coefficient between the items

Table 5. Spearman correlation coefficients and 95% confidence limits obtained by bias corrected bootstrap with five level Likert not iso-correlated items with positive skewness

	Bootstrap Pearson correlation coefficients with 95% confidence limits		
Simulation Scenarios	LAF with LFS	LAF with SS	LFS with SS
V = 5, N = 100, r = 0.45	0.5423 (0.4058; 0.6575)	0.6021 (0.4765; 0.7039)	0.9057 (0.8589; 0.9385)
V = 9, N = 100, r = 0.45	0.7344 (0.6486; 0.8088)	0.751 (0.6624; 0.8227)	0.9541 (0.9360; 0.9698)
V = 17, N = 100, r = 0.45	0.7505 (0.6583; 0.8215)	0.7401 (0.6470; 0.8127)	0.9632 (0.9409; 0.9772)
V = 5, N = 250, r = 0.45	0.5820 (0.5015; 0.6530)	0.5900 (0.5164; 0.6570)	0.8719 (0.8390; 0.8982)
V = 9, N = 250, r = 0.45	0.7061 (0.6414; 0.7612)	0.6918 (0.6299; 0.7458)	0.9554 (0.9435; 0.9660)
V = 17, N = 250, r = 0.45	0.7945 (0.7448; 0.8366)	0.7974 (0.7463; 0.8376)	0.9783 (0.9716; 0.9839)
V = 5, N = 500, r = 0.45	0.6913 (0.6443; 0.7323)	0.6678 (0.6197; 0.7111)	0.9623 (0.9552; 0.9688)
V = 9, N = 500, r = 0.45	0.7179 (0.6769; 0.7551)	0.7015 (0.6595; 0.7400)	0.9752 (0.9698; 0.9798)
V = 17, N = 500, r = 0.45	0.8269 (0.7973; 0.8524)	0.8205 (0.7914; 0.8464)	0.9826 (0.9793; 0.9858)
V = 5, N = 1000, r = 0.45	0.6098 (0.5725; 0.6452)	0.5954 (0.5579; 0.6309)	0.9660 (0.9602; 0.9708)
V = 9, N = 1000, r = 0.45	0.7514 (0.7260; 0.7756)	0.7378 (0.7101; 0.7622)	0.9876 (0.9859; 0.9891)
V = 17, N = 1000, r = 0.45	0.8063 (0.7829; 0.8265)	0.7989 (0.7763; 0.8196)	0.9828 (0.9803; 0.9852)
V = 5, N = 100, r = 0.60	0.7456 (0.6585; 0.8153)	0.7450 (0.6540; 0.8144)	0.9568 (0.9368; 0.9724)
V = 9, N = 100, r = 0.60	0.8306 (0.7666; 0.8824)	0.8521 (0.7899; 0.8981)	0.9763 (0.9649; 0.9847)
V = 17, N = 100, r = 0.60	0.8634 (0.8053; 0.9059)	0.8521 (0.7871; 0.8977)	0.9837 (0.9737; 0.9907)
V = 5, N = 250, r = 0.60	0.7424 (0.6815; 0.7924)	0.7317 (0.6715; 0.7817)	0.9583 (0.9443; 0.9687)
V = 9, N = 250, r = 0.60	0.8237 (0.7792; 0.8609)	0.8082 (0.7628; 0.8466)	0.9848 (0.9794; 0.9888)
V = 17, N = 250, r = 0.60	0.8873 (0.8552; 0.9123)	0.8828 (0.8506; 0.9082)	0.9895 (0.9863; 0.9923)
V = 5, N = 500, r = 0.60	0.7983 (0.7635; 0.8289)	0.7745 (0.7367; 0.8066)	0.9779 (0.9730; 0.9824)
V = 9, N = 500, r = 0.60	0.8209 (0.7915; 0.8475)	0.8118 (0.7822; 0.8392)	0.9814 (0.9776; 0.9849)
V = 17, N = 500, r = 0.60	0.8971 (0.8774; 0.9135)	0.8897 (0.8694; 0.9076)	0.9906 (0.9887; 0.9925)
V = 5, N = 1000, r = 0.60	0.7496 (0.7217; 0.7760)	0.7435 (0.7157; 0.7691)	0.9757 (0.9718; 0.9791)
V = 9, N = 1000, r = 0.60	0.8452 (0.8271; 0.8627)	0.8353 (0.8152; 0.8527)	0.9935 (0.9927; 0.9944)
V = 17, N = 1000, r = 0.60	0.8816 (0.8656; 0.8954)	0.8780 (0.8623; 0.8921)	0.9916 (0.9903; 0.9928)
V = 5, N = 100, r = 0.75	0.8401 (0.7696; 0.8922)	0.8115 (0.7290; 0.8701)	0.9778 (0.9638; 0.9869)
V = 9, N = 100, r = 0.75	0.9126 (0.8688; 0.9422)	0.9139 (0.8695; 0.9437)	0.9849 (0.9788; 0.9909)
V = 17, N = 100, r = 0.75	0.9126 (0.8701; 0.9425)	0.9036 (0.8563; 0.9361)	0.9870 (0.9806; 0.9926)
V = 5, N = 250, r = 0.75	0.8295 (0.7830; 0.8666)	0.8210 (0.7722; 0.8596)	0.9747 (0.9616; 0.9828)
V = 9, N = 250, r = 0.75	0.8814 (0.8463; 0.9100)	0.8736 (0.8371; 0.9029)	0.9880 (0.9841; 0.9914)
V = 17, N = 250, r = 0.75	0.9364 (0.9159; 0.9518)	0.9302 (0.9079; 0.9471)	0.9926 (0.9904; 0.9948)
V = 5, N = 500, r = 0.75	0.8780 (0.8530; 0.8987)	0.8670 (0.8417; 0.8881)	0.9855 (0.9820; 0.9886)
V = 9, N = 500, r = 0.75	0.8914 (0.8712; 0.9089)	0.8834 (0.8625; 0.9019)	0.9898 (0.9879; 0.9917)
V = 17, N = 500, r = 0.75	0.9429 (0.9308; 0.9532)	0.9409 (0.9288; 0.9516)	0.9932 (0.9918; 0.9946)
V = 5, N = 1000, r = 0.75	0.8375 (0.8164; 0.8564)	0.8283 (0.8068; 0.8479)	0.9879 (0.9857; 0.9898)
V = 9, N = 1000, r = 0.75	0.9114 (0.8993; 0.9226)	0.9052 (0.8919; 0.9167)	0.9949 (0.9943; 0.9957)
V = 17, N = 1000, r = 0.75	0.9308 (0.9207; 0.9399)	0.9290 (0.9189; 0.9381)	0.9940 (0.9930; 0.9949)

LAF = Simulated latent factor; LFS = Latent factor score performed via confirmative factor analysis; SS: Score performed by the sum of items; V = Number of items; N = Sample size; r = Pearson correlation coefficient between the items

We observed that both LFSs and SSs are normally distributed and centred on 0, for normally distributed items while they are skewed when the scores are derived from skewed items.

The results above reported were confirmed by simulation conducted having 10% of items randomly assigned to a null correlation with the simulated latent factor. More specifically, in these simulation settings we only observed an overall reduction of the correlation between the simulated latent factor and the score computed from the variables.

Empirical studies

Analysis of the SUS

The analytical dataset derived from the application of the SUS in a clinical setting was composed by 208 participants of which 112 (53.8%) were women. The median age was 68 years (5th to 95th range = 45.4 to 81) with a median year of education of 13 years (5th to 95th range = 5 to 17). We observed a strong correlation between the LFSs and the SSs, for the overall total score and for both the subscales of learnability and usability. Specifically, we observed a Spearman correlation coefficient of 0.928 between the total LFS and the SS. The Spearman correlation coefficients between the subscales of learnability and usability were in the same range (R=0.964 and R=0.957 for the learnability and usability subscales, respectively). Notably, we observed a strong left skewness of both the LFS and the SS for the total and the subscales of the SUS. The LFS and SS distributions and correlations for the total SUS and the subscales were reported on Figure 1.

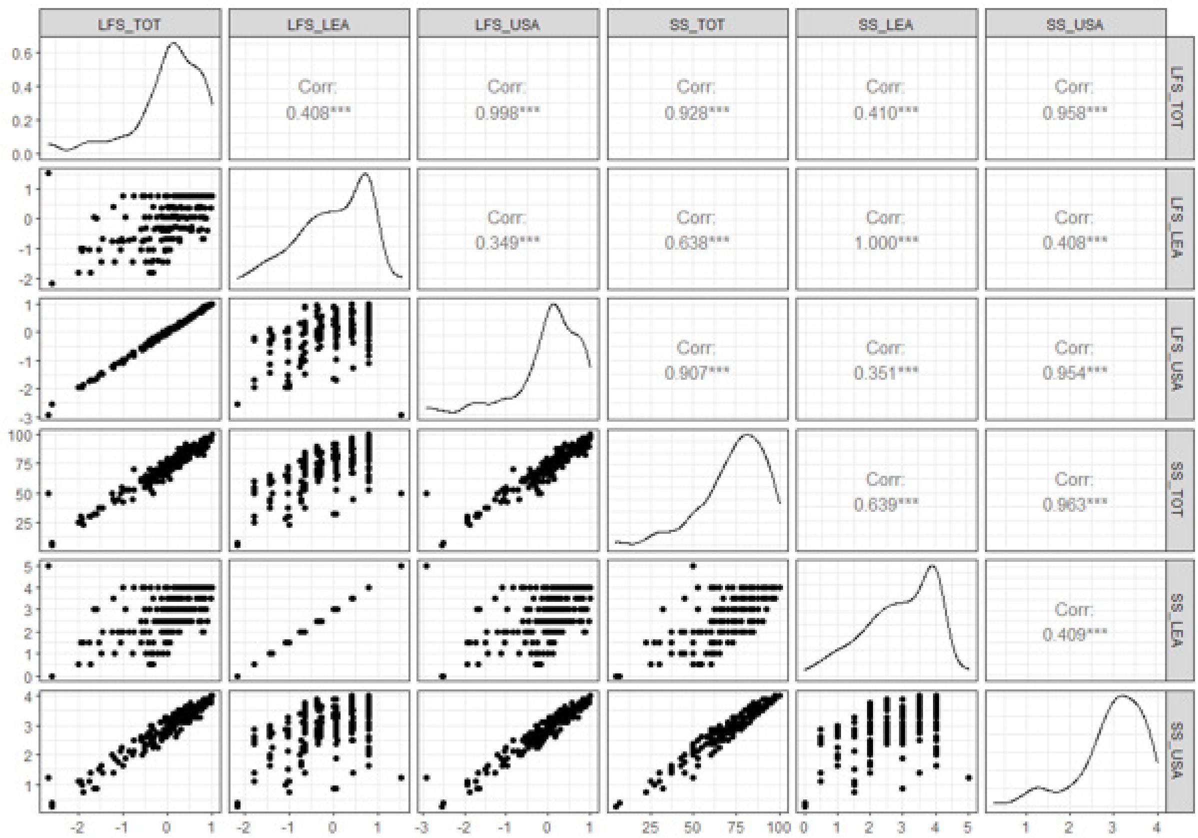
Analysis of the Yoni task

The Yoni task was administered to 235 participants (men=127 (54%)). The median age was 35 years (5th to 95th range = 19 to 54), the median education was 16 years (5th to 95th range =9 to 21). We reported a moderate to high Spearman correlation coefficient between the LFS and the SS for the total YONI scores (R=0.931) and for the subscales (R=0.942 and R=0.624, for the affective and cognitive subscales, respectively). Both of the LFS and the SS were positively skewed.

DISCUSSION

In the present work we showed some interesting aspects of the application of the SS and LFS approaches to derive scales' scores. Firstly, using numerical simulation we showed that the SS and the LFS are interchangeable when the items are iso-correlated to the latent factor. Basically, the fact that all the items have the same correlation with the latent factor corresponds with having the same weights in equation (1). Notably, those weights are not necessarily equal

Figure 1. Spearman correlations of latent factor scores and sum of items scores in the SUS



LFS_ = Latent factor score, SS_ = Sum of item's score, TOT = total score, LEA = Learnability score, USA = Usability score to one when the LFS is performed via a confirmatory factor analysis but have the same value when the SS is performed. This is the reason why the correlation between the LFS and the SS tends towards one when the correlation between the items and the latent factors are all equals. Therefore, the LFS and the SS are asymptotically equivalent when the items are equally correlated. Here a sample size above 500 and even a minimal number of items seems to be sufficient to guarantee the equivalence between the SS and the LFS. Moreover, according to our numerical simulation, it seems that the correlations between the LFSs and the SS are extremely high also in the presence of different strength of correlation between the items defying the latent factor and for ordinal categorical skewed items as well. This result points out the hypothesis that the use of LFS and SS are interchangeable in most common situations met when analysing scales. In fact, we may assume the syllogism that if the LFS is the best way to represent a latent factor (as we report by our simulation) and the LFS is equivalent to the SS then the SS also can be used to portray that given latent factor. The quantitative nature of the interchangeability between the LFS and the SS could be a matter of discussion. Here it could be suggested that the two are interchangeable given a certain level of correlation, for example above 0.9.

The interchangeability of the LFS and the SS is in agreement with the work from Widaman and Revelle [9] but not with the work of McNeish and Wolf [8]. The congeneric measurement model is a general approach where it is assumed that every item is differently related to a given construct [6,21]. It is intuitively correct as we can imagine that not all of the items in a scale share the same ability to represent a given construct. However, the different correlations between an item and its latent factor does not seem to influence the correlation between a score computed by the sum of items or the one derived by the application of a confirmatory factor analysis.

This result is truly conditioned to our simulation settings. However, our simulations may not represent the reality of scales of common use where the items are generally well intercorrelated. Notably, our results seems as quite robust being confirmed for sets of low and high correlations between the items and the latent factors. On the other hand, we cannot exclude that this evidence would not be confirmed in more extreme settings. Specifically, when both high and low correlated items are present.

In conclusion to our simulation studies aimed to investigate the relation between the LFS and the SS, we can state that the correlation structure between the items determine the similarity between the LFS and the SS. The similarity between the LFS and the SS is remarkable in the situations likely faced during analysis of scales. Here the use of a SS or a LFS to portray the latent factors are quite interchangeable. Moreover, there is a specific reason that justify the use of SS over the LFS, the so-called "indeterminacy of the factor scores" [22–24]. Briefly, the indeterminacy arises from the fact that, when the latent factor is not unique there is not a unique solution for the factor analysis. As a consequence, a virtually infinite number of solutions could be derived to define the relationship between the items and the latent factor [23,24].

It seems that the SS and the LFS are quite similar in most of the cases. However, there could be a number of situation when this is not the case. We can imagine that extreme correlation structures between the items may results in large differences between the SS and the LFS. For example, when there are relatively few items that are strongly correlated to the latent factor and a plethora of items with small or null correlation with the latent factor, acting as disturbance or white noise variable. Intuitively, here the LFS will be more accurate to represent the latent factor than the SS. Our work provides some suggestions to identify some of these situations as we showed the importance of the correlation structure between the items. In this perspective, it could be recommendable to perform an exploratory factor analysis before deciding which approach use to derive a score [25–27]. According to the results of an exploratory analysis, the researcher could even decide to exclude some items if they are not sufficiently correlated to the latent factors to define the confirmatory factor model [28].

We showed that the SSs and the LFs are strongly correlated when the correlation between the items and the latent factors are similar. This same approach could be used reversely, deriving a LFS and the SS and look at their correlation before to conduct the further analyses or to decide which approach to use. Here we showed that in case of a strong correlation between the LFS and the SS it seems reasonable to adopt any of the two approaches, even according to personal preferences, practical aspects or believe. On the contrary, the existence of a moderate or weak correlation between the LFS and the SS could represent a question mark that deserves a further investigation.

Some interesting evidence emerged by the analysis of real data here provided. Firstly, it seems that it is not uncommon that the LFS and the SS are strongly correlated in real settings; this is more evident for the SUS then for the Yoni task instrument. Notably, the difference between LFS and SS we observed in the SUS compared to the Yoni task seems to be more related to specific features of the tools and possibly by the study sample than to the CTT or IRT type tools. Specifically, on the Yoni task instrument we observed differences between the SS and the LFS for the cognitive score only while the correlations between the SS and the LFS appeared as satisfactory for other scores. However, we cannot fully exclude that the type of tool may have played a role. Theoretically, the IRT items are not parallel instruments for each of them have its own IRF, this should determine inaccuracy of the SS even if the ranking of subjects should not be affected by its use [29]. However, we observed a high correlation between the SS and the LFS of the Yoni total and affective scores. Both the LFS and the SS computed from the SUS and the Yoni task are skewed as the single items' score. This result is obvious when considering the loading of the single item should not improve the LFS and so the use of confirmatory factor analysis does not enhance the LFS distribution.

The present work has many strength points. Firstly, we showed when the two approaching of deriving scores from scales are equivalent. Here, we showed that in the most commonly situations the agreement between the two procedures is satisfactory, despite the presence of items with different correlations to the latent factors or even the presence of uncorrelated items that should determine a relevant difference between the two types of scores, at least theoretically. This is a relevant achievement because allows to use a simpler method to derive score's scales or even justify the use of sum of items scores, or relatively similar approaches, which is also intended to be used according to many scales' instruction manuals. Notably, our results are confirmed in different scenarios considering different sample sizes, correlation structures and items distributions.

Some limitations may have affected the present work. Firstly, the number of simulative scenarios may be limited by the number of items considered and the correlation range between the variables and the latent factors. Moreover, we simulated a very simple structure having only two latent factors while some scales may have a more complex structure. We did not observe any relevant differences when using different settings in some preliminary evaluation. However, many of the scales in use do not have such a complex structure and are composed with a limited number of items, for practical reasons of reducing the time to fill in the scale and because it seems that increasing the number of items does not necessarily improve the performances of the scale [30].

Finally, we explored strong and weak correlations between the items and the scale even if a moderate to strong correlation between the item and the scale should be assumed. Another possible limitation is the extensive use of simulations which have a limited capacity to represent the empirical complexity. However, we showed how simulative results are confirmed by real data.

ETHICAL STATEMENT

The project was approved by the ethic committee of the Don Gnocchi Foundation on 17.2.2021

REFERENCES

1. Jack B, Clarke AM. The purpose and use of questionnaires in research. Prof Nurse Lond Engl. 1998;14(3):176–9.
2. Lydeard S. The questionnaire as a research tool. Fam Pract. 1991;8(1):84–91.
3. Patten M. Questionnaire research: A practical guide. routledge; 2016.
4. Couper MP. Technology and the survey interview/questionnaire. Envisioning Surv Interview Future. 2008;58–76.
5. Humble S. Quantitative analysis of questionnaires: Techniques to explore structures and relationships. Routledge; 2020.
6. Jöreskog KG. Statistical analysis of sets of congeneric tests. Psychometrika. 1971;36(2):109–33.
7. Mellenbergh GJ. Tests and questionnaires Analysis. Advis Res Methods Consult Companion. 2008;235–68.
8. McNeish D, Wolf MG. Thinking twice about sum scores. Behav Res Methods. 2020 Dec;52(6):2287–305.
9. Widaman KF, Revelle W. Thinking thrice about sum scores, and then some more about measurement and analysis. Behav Res Methods. 2023 Feb;55(2):788–806.
10. Realdon O, Adorni R, Ginelli D, Micucci D, Blasi V, Bellavia D, et al. Embedding the Patient-Citizen Perspective into an Operational Framework for the Development and the Introduction of New Technologies in Rehabilitation Care: The Smart&Touch-ID Model. Healthcare. 2023 May 30;11(11):1604.
11. Ripley B, Venables B, Bates DM, Hornik K, Gebhardt A, Firth D, et al. Package 'mass'. Cran R. 2013;538:113–20.
12. DeBruine L. faux: Simulation for Factorial Designs [Internet]. Zenodo; 2023 [cited 2024 Jan 8]. Available from: https://zenodo.org/record/2669586
13. 13. Wuertz D, RUnit S, Chalabi MY. Package 'fGarch'. Technical report, working paper/manual, 09.11. 2009. URL http://cran. r ...; 2013.
14. Brooke J. Sus: a "quick and dirty'usability. Usability Eval Ind. 1996;189(3):189–94.
15. Lewis JR, Sauro J. The factor structure of the system usability scale. In Springer; 2009. p. 94–103.
16. Borsci S, Federici S, Lauriola M. On the dimensionality of the System Usability Scale: a test of alternative measurement models. Cogn Process. 2009 Aug;10(3):193–7.
17. Rossetto F, Castelli I, Baglio F, Massaro D, Alberoni M, Nemni R, et al. Cognitive and Affective Theory of Mind in Mild Cognitive Impairment and Parkinson's Disease: Preliminary Evidence from the Italian Version of the Yoni Task. Dev Neuropsychol. 2018;43(8):764–80.
18. sernia S, Rossetto F, Blasi V, Massaro D, Castelli I, Ricci C, et al. Measuring cognitive and affective theory of mind with the Italian Yoni task: normative data and short versions. Curr Psychol. 2023;42(27):23519–30.
19. Isernia S, Rossetto F, Shamay-Tsoory S, Marchetti A, Baglio F. Standardization and normative data of the 48-item Yoni short version for the assessment of theory of mind in typical and atypical conditions. Front Aging Neurosci. 2022;14:1048599.
20. Rosseel Y. lavaan: An R package for structural equation modeling. J Stat Softw. 2012;48:1–36.

21. Graham JM. Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. Educ Psychol Meas. 2006;66(6):930–44.
22. DiStefano C, Zhu M, Mindrila D. Understanding and using factor scores: Considerations for the applied researcher. Pract Assess Res Eval. 2009;14(1):20.
23. Gorsuch RL. Factor analysis: Classic edition. Routledge; 2014.
24. Grice JW. Computing and evaluating factor scores. Psychol Methods. 2001;6(4):430.
25. Bollen KA. Structural equations with latent variables. Vol. 210. John Wiley & Sons; 1989.
26. Bollen KA. Latent variables in psychology and the social sciences. Annu Rev Psychol. 2002;53(1):605–34.
27. Bollen K, Lennox R. Conventional wisdom on measurement: A structural equation perspective. Psychol Bull. 1991;110(2):305.
28. Brown TA. Confirmatory factor analysis for applied research. Guilford publications; 2015.
29. Sijtsma K, Hemker BT. A taxonomy of IRT models for ordering persons and items using simple sum scores. J Educ Behav Stat. 2000;25(4):391–415.
30. Jacobsen BK, Hasvold T, Høyer G, Hansen V. The General Health Questionnaire: how many items are really necessary in population surveys? Psychol Med. 1995 Sep;25(5):957–61.