SA Journal of Industrial Psychology
ISSN: (Online) 2071-0763, (Print) 0258-5200

Derivation of individualised item interrelatedness indices of careless responding

Author:
Gert K. Huysamen[1]

Affiliation:
[1]Department of Psychology, Faculty of Humanities, University of the Free State, Bloemfontein, South Africa

Corresponding author:
Gert Huysamen,
gerthuysamen@mweb.co.za

Dates:
Received: 24 Nov. 2022
Accepted: 23 Oct. 2023
Published: 26 Dec. 2023

How to cite this article:
Huysamen, G.K. (2023). Derivation of individualised item interrelatedness indices of careless responding. *SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde, 49*(0), a2070. https://doi.org/10.4102/sajip.v49i0.2070

Orientation: Presently, carelessly responding (CR) individuals are omitted in terms of several individual indices, including consistency-type indices (that compare performance on only a limited number of matched item pairs), and subsequently, the effectiveness of such screening is evaluated in terms of, among others, the group mean item interrelatedness (IIR) (based on all $J[J-1]$ item pairs).

Research purpose: This research aims to develop individualised versions of the group IIR measures to render them applicable during the screening phase as substitutes for the presently used consistency-type indices.

Motivation for the study: Such individual consistency indices may be used together with other CR indices to jointly determine the eventual evaluation results.

Research approach/design and method: To develop the intended CR indices mathematical-statistical principles were applied on the product moment correlation and coefficient alpha formulae.

Main findings: Three individual IIR indices have been developed which show individual respondents' respective contributions to the mean item inter-correlation and to coefficient alpha, as measures of group mean IIR.

Practical/managerial implications: These indices may be used during screening in lieu of the existing restrictive consistency indices.

Contribution/value-add: Carelessly responding respondents who previously may have survived screening because of less-inclusive consistency-type IIR indices and consequently may have negatively affected the eventual evaluation results, are now screened out.

Keywords: self-report inventories; Likert items; survey research; careless responding; coefficient alpha.

Introduction

Since the beginning of this century a topic in self-report measurement, variously referred to as careless responding (CR) (e.g. Meade & Craig, 2012), inattentive responding (e.g. Johnson, 2005), and insufficient effort responding (e.g. Huang et al., 2012), has received increased attention. Such uncooperative responding behaviour is generally but not exclusively associated with online-administered surveys. It manifests in a variety of ways, from haphazardly endorsing options to choosing the same-numbered option, or a pattern of such options, on several consecutive items. Because of measurement reliability and validity concerns, indices have been developed to detect CR individuals with a view to screening them out as part of a data cleaning exercise (cf. Wilkinson & The Task Force on Statistical Inferences, 1999). As CR, unlike response styles such as socially desirable and acquiescent responding, is characterised by inadequate attention to item content, self-report CR inventories are unlikely to adequately capture such behaviour because they would be subject to this behaviour as well. Different strategies to devise individual CR detection indices have been published (e.g. Curran, 2016; DeSimone & Harms, 2018; Dunn et al., 2018; Hong et al., 2020; Huang et al., 2012; Johnson, 2005; Meade & Craig, 2012; Steedle et al., 2019). These strategies are applied on the responses obtained on scales that are regularly used in practice and not on a scale specifically designed to measure CR (as such an operation would be an exercise in futility given the nature of CR). Arthur et al. (2021) comprehensively reviewed both CR and socially desirable responding in terms of their respective definitions, prevention, and the optimal uses of different indices to detect them.

Because of its protean character, it is unlikely that a single detection method would be able to identify all of its different manifestations satisfactorily. Typically, a combination of CR indices is recommended to cover all the CR bases, so to speak (e.g. Dunn et al., 2018). In the *infrequency* method the endorsement of several options that are very unlikely to be true (e.g. 'I'm not aware of anyone who has contracted COVID-19') is taken as evidence of CR. The *response-pattern* class of indices is directed at identifying individuals who have persevered with the same-numbered option, or with the same sequence of options (such as *1*, *2*, 3, *4*, *1*, *2*, *3*, *4*, or *2*, *3*, *2*, *3*, *2*, *3*, in four-point scales) throughout a questionnaire. To identify such respondents, the *Long-string* index (Meade & Craig, 2012) uses a computer algorithm to determine the number of times the same-numbered option is chosen within a certain number of items. In the case of *outlier analyses* respondents whose scores deviate considerably from those of the rest, are flagged. *Mahalanobis distance* measure (Meade & Craig, 2012) is a multivariate extension of such outlier analysis. The *standardised log-likelihood* l_z *statistic* (Conijn et al., 2019) is the log-likelihood of a respondent's response pattern in terms of the item-response theory.

However, in the case of (particularly extreme) random responding, so-called consistency indices are considered to be more successful (e.g. Arthur et al., 2021). Indices in this category attempt to reflect the consistency with which respondents endorse similar content (in different items), or refrain from endorsing both of the items in the case of contradictory item pairs. Huang et al. (2012) described one index in this category as being based on the idea that 'items on the same scale are expected to correlate with each other for each individual' (p. 102). Such *intra*-person correlations arguably apply to most of the other consistency indices except for the Individual *Response Variability* (*IRV*) Index (Marjanovic et al., 2015).

Several of the consistency methods of examining CR divide the scale items into two groups in such a manner that for every item in the one group, there is a matched counterpart in the other group, and compute the intra-person correlation between the two subsets so formed. For example, the *Even-Odd Consistency* or *Individual Reliability* Index (Meade & Craig, 2012) divides the scale items into odd-numbered and even-numbered items (or into randomly split halves). As carefully responding individuals are expected to register comparable scores on the paired halves, a negative Spearman–Brown adjusted intra-person correlation is then interpreted as indicative of CR. In the case of the *psychometric antonyms* procedure (Johnson, 2005) a set of item pairs that was earlier shown to have the highest, negative correlations, is used. Alternatively, such item pairs may be selected in terms of those that are contradictory semantically, that is, in a dictionary sense. As attentive respondents are unlikely to endorse both members of such item pairs, a high, positive intra-person correlation between them is then taken as an indicator of CR. In the *psychometric synonyms* approach (Meade & Craig, 2012) such an interpretation is attached to a high, negative intra-person correlation because it identifies respondents who failed to endorse both members of item pairs with similar meaning

After respondents have been screened (in terms of the *individual* CR indices, such as the Even-Odd consistency-type index), the effectiveness of such screening is typically inspected in an evaluation phase in terms of the average item interrelatedness (IIR), statistical power, and factor analyses results obtained for the retained *group* (e.g. Hong et al., 2020; Maniaci & Rogge, 2014; Steedle et al., 2019). The objective of the present methodological note is to develop three *individual* IIR CR indices, each of which directly shows each respondent's contribution to *group* IIR and which, therefore, may be used during screening with a view to potentially benefitting the eventual evaluation results. This methodological presentation concludes with a fictional, numerical example of the application of the new IIR indices on prototypical response protocols that may not necessarily be found in empirically obtained data sets but are intended to demonstrate the potential advantages of these indices in CR screening.

Individual indices of item interrelatedness

As indicated before, the existing consistency indices involve intra-person correlations between two lists of matched item pairs, so that there are at most $J/2$ such item pairs to be formed among a total of J items. By contrast, the indices to be introduced here compare performance on every item with performance on every other item, thus yielding altogether $J(J-1)$ such comparisons. In the case of properly constructed scales an increase in the number of items benefits both consistency reliability and content validity. Similarly, it could be argued that an increase in the number of item pairs would be psychometrically beneficial to the measurement of whatever the resulting variable is intended to reflect. Also, whereas the deviation scores in the methods in the preceding section are taken from an *individual* respondent's means on (paired) collections of items, the deviation scores for the IIR indices to be developed here are taken from the *group* item means. However, as the individual respondents in the former approach form part of the group in the latter case, this should not be the cause of markedly contradictory results obtained for the new IIR and extant CR indices.

Statistical framework for the development of item interrelatedness indices

For each and every individual respondent consider a separate $J \times J$ matrix such that in each of its J diagonal cells appears a *weighted squared deviation score*:

$$d_{ij}^{\,2} = (X_{ij} - M_j)^2 / N, \qquad \text{[Eqn 1]}$$

where:

M_j is the mean, over individuals, of item j, and every non-diagonal cell contains a *weighted deviation-score cross-product*,

$$dc_{ijk} = (X_{ij} - M_j)(X_{ik} - M_k)/N, \quad \text{[Eqn 2]}$$

where:

M_j and M_k are (sample) means of Items j and k, respectively.

For example, Respondent *H* (8th individual) in Table 1 has a score of 1 on Item *1* (mean = 3.3) and a (reversed) score of 4 on (negatively keyed) Item 2 (item mean = 3.0). In the first two diagonal cells of her matrix, the d_{ij}^2 scores of $(1 - 3.3)^2/10 = 0.529$ and $(4 - 3)^2/10 = 0.1$, respectively, are registered. In both the cell formed by the second row and first column and the one formed by the first row and second column $(-2.3 \times 1.0)/10 =) -0.23$ is entered. These $J(J - 1)$ weighted deviation-score cross-products for any particular respondent i sum to his or her *weighted deviation-score cross-product total*:

$$dct_i = \Sigma_j\Sigma_k dc_{ijk} = \Sigma_j\Sigma_k[(X_{ij} - M_j)(X_{ik} - M_k)/N], \quad \text{[Eqn 3]}$$

where:

$j \neq k$.

If the individual $J \times J$ matrix of d_{ij}^2 and dc_{ijk} values is aggregated across all N individuals, the familiar $J \times J$ item variance-covariance matrix for the total sample is obtained. In other words, the sum, over all N individuals, of d_{ij}^2, gives the sample variance of item j, s_j^2,

$$\Sigma_i d_{ij}^2 = \Sigma_i[(X_{ij} - M_j)^2/N] = s_j^2 \text{ (and } \Sigma_i d_{ik}^2 = \Sigma_i[(X_{ik} - M_k)^2/N] = s_k^2\text{; etc.)}, \quad \text{[Eqn 4]}$$

and the sum, over all N individuals, of dc_{ijk}, yields the sample covariance of items j and k,

$$s_{jk} = \Sigma_i dc_{ijk} = \Sigma_i[(X_{ij} - M_j)(X_{ik} - M_k)/N]. \quad \text{[Eqn 5]}$$

This sample item covariance, s_{jk}, is the numerator of the sample correlation between items j and k:

$$r_{jk} = s_{jk}/s_j s_k. \quad \text{[Eqn 6]}$$

If all the s_j^2 and s_{jk} entries in the (sample) item variance-covariance matrix are summed over all J items, the sample variance, s_X^2, of total test scores is obtained:

$$s_X^2 = \Sigma_j s_j^2 + \Sigma_j\Sigma_k s_{jk}, \quad \text{[Eqn 7]}$$

where:

s_j^2 is the variance of item j, and s_{jk} the covariance of items j and k.

The quantity $\Sigma_j\Sigma_k s_{jk}$ in the preceding equation is also equal to the sum, over all N individuals, of their deviation-score cross-product totals (Eqn 3):

$$\Sigma_i dct_i = \Sigma_i\Sigma_j\Sigma_k[(X_{ij} - M_j)(X_{ik} - M_k)/N] = \Sigma_j\Sigma_k s_{jk}, \quad \text{[Eqn 8]}$$

where:

$j \neq k$.

If individuals respond consistently to homogeneous item content, their scores (reversed where necessary) for any pair of items, j and k, are expected to be either both above or both below the group means of these items, so that the deviation scores, $(X_{ij} - M_j)$ and $(X_{ik} - M_k)$, and hence, the dc_{ijk} values (Eqn 2) are positive. Careless respondents, however, are likely to either fail to endorse both item pair members containing similar content (or to confirm both of contradictory item pair members). Such behaviour would result in deviation scores with opposite signs, and hence, negative dc_{ijk} values. If there are only a few small, negative dc_{ijk} values for a particular individual, their sum over item pairs may still be positive, but as their number and (absolute) sizes increase, their dct_i value (Eqn 3) will become negative, and its absolute value will increase.

The (individual) proportional item inter-correlational *(PIC_i)* IIR index

Although during the evaluation stage, some CR researchers (e.g. Hong et al., 2020) used coefficient alpha (Cronbach, 1951)

TABLE 1: Fictional item data matrix, item totals, *ISD*, $\Sigma_j d_{ij}^2$, dct_i, PIC_i, PA_i and IA_i.

Persons	1	2	3	4	5	6	M_i	*ISD*†	$\Sigma_j d_{ij}^2$‡	dct_i§	PIC_i¶	PA_i††	IA_i‡‡
A	5	1 (5)	1 (5)	5	5	1 (5)	5.0	-	1.965	9.48	0.156	0.288	+0.080
B	1	5 (1)	5 (1)	1	1	5 (1)	1.0	0	3.005	14.68	0.246	0.446	+0.168
C	5	2 (4)	2 (4)	5	5	2 (4)	4.5	0.5	1.025	4.90	0.082	0.149	+0.034
D	3	1 (5)	1 (5)	3	3	1 (5)	4.0	1.0	1.365	0.84	0.016	0.025	-0.024
E	3	3	3	3	3	3	3.0	0	0.081	0.34	0.002	0.004	-0.001
F	4	4 (2)	4 (2)	4	4	4 (2)	3.0	1.0	0.345	-0.18	-0.003	-0.005	-0.009
G	5	5 (1)	5 (1)	5	5	5 (1)	3.0	2.0	1.805	-1.64	-0.028	-0.049	-0.059
H	1	2 (4)	3	4	5	4 (2)	3.2	1.344	0.905	-0.90	-0.014	-0.027	-0.030
I	2	3	2 (4)	3	1	2 (4)	2.8	1.067	1.125	-0.60	-0.008	-0.018	-0.032
J	4	4 (2)	4 (2)	3	4	5 (1)	2.7	1.105	0.625	0.46	0.008	0.014	-0.009
Total	**33**	**30**	**30**	**36**	**36**	**28**	-	-	**12.246**	**27.21**	**0.457**	**0.827**	-

Note: †, Individual inter-item standard deviation.
‡, Weighted deviation-score squares total.
§, Weighted deviation-score cross-products total.
¶, (Individual) Proportional inter-item correlational index.
††, (Individual) proportional alpha-related index.
‡‡, (Individual) Incremental alpha-related effect.

as a measure of consistency reliability, Huang et al. (2012) employed this coefficient as a measure of group IIR. However, the group average item inter-correlation, rather than coefficient alpha, first comes to mind as a measure of such IIR. The (individual) *Proportional Item Inter-Correlational* (PIC_i) Index is the average of a respondent's contributions to the $J(J-1)$ item inter-correlations among the J items of a scale. In terms of Eqns (5) and (6), the numerator of any item inter-correlation, s_{jk}/s_js_k, is the sum of all dc_{ijk} contributions of all N individuals (whereas the denominator of s_{jk}/s_js_k is a constant for all individuals). Therefore, it follows that for any particular individual, the ratio dc_{ijk}/s_js_k, represents his or her proportional contribution to that particular item inter-correlation. A similar statement applies to a respondent's contribution to each of the $J(J-1)$ item inter-correlations. The PIC_i Index:

$$PIC_i = \{\Sigma_j\Sigma_k[dc_{ijk}/(s_js_k)]\}/J(J-1), \qquad \text{[Eqn 9]}$$

where:

$j \neq k$,

gives the mean of all such proportional dc_{ijk}/s_js_k contributions across all $J(J-1)$ item inter-correlations due to a particular respondent, to $\bar{r}_{jk}$, the sample mean item inter-correlation: If this index is summed, over individuals, it gives $\bar{r}_{jk}$, the mean item inter-correlation for the total group. (Alternatively, a mean of the quotients involved may be determined by dividing the sum of the $J(J-1)$ dc_{ijk} values of the $J(J-1)$ mean item inter-correlations by the sum of their corresponding $J(J-1)$ s_js_k products.)

As correlations are involved, by definition, PIC_i cannot exceed unity (1.00). Consequently, for even relatively small samples it will have to be reported to several decimal places if finer distinctions among the values of individual respondents are required. If this presents a problem, respondents' PIC_i scores may be multiplied by N, to yield $PIC*N_i$. This multiplication operation has the same effect as replacing dc_{ijk} in Eqn (9), in terms of Eqn (5), by $(X_{ij} - M_j)(X_{ik} - M_k)$. It has no effect on the relative positions of respondents' PIC_i values or on the occurrence of negative signs, which is a critical feature of this index.

The (individual) proportional alpha-related *(PA_i)* IIR index

As said before, during the evaluation stage, Huang et al. (2012) used coefficient alpha as a measure of group IIR. In terms of this practice, a CR index of each individual's proportional contribution to coefficient alpha that could be used (during the screening phase), should be useful. The (Individual) *Proportional Alpha-related* (PA_i) Index:

$$PA_i = [J/(J-1)][(dct_i/s_X^2), \qquad \text{[Eqn 10]}$$

gives an individual's proportional contribution to coefficient alpha, because in terms of Eqns (7) and (8), if it is summed over individuals, coefficient alpha for the total group, is obtained:

$$\begin{aligned}\Sigma_i PA_i &= \Sigma_i\{[J/(J-1)][dct_i/s_X^2]\} = [J/(J-1)][\Sigma_i dct_i/s_X^2]\\ &= [J/(J-1)][1 - \Sigma_j s_j^2/s_X^2]\\ &= \textit{coefficient alpha.}\end{aligned} \qquad \text{[Eqn 11]}$$

As attentive (and, hence, consistent) responding is expected to benefit scores obtained on the formula for coefficient alpha (irrespective of whether it is used as a measure of consistency reliability or as an index of careful responding), it makes sense that an alpha-derived index could be used to reflect such consistent and, hence, attentive responding. In view of its relationship with the popular coefficient alpha, the PA_i Index may be interpreted in terms of the conventions that apply in interpreting coefficient alpha values. For example, values of at least in the lower 0.70s are typically regarded as acceptable. No similar frame of reference exists for interpreting the PIC_i indices.

As an IIR index, PA_i suffers from the same drawback as does PIC_i in that it will have to be reported to several decimal places in the case of even relatively small samples. Incorporation of the same remedy of multiplication by N, as in the case of PIC_i, to give PA_i*N, solves this problem.

The (individual) incremental alpha-related *(IA_i)* IIR index

The PA_i Index gives an individual respondent's proportional contribution to the group average IIR but does not directly convey the increase or decrease that the inclusion of any individual brings about in this quantity for those already in the group. A conceptually simple yet computationally cumbersome way of obtaining this information is by applying the coefficient alpha formula as many times as there are respondents, each time with a different respondent omitted. The (Individual) *Incremental Alpha-related* (IA_i) Index for a particular individual then is the result obtained for the total group minus the result obtained with that individual excluded. It shows the increment in the existing average IIR for a group brought about by the inclusion of an individual to that group. If IA_i is positive, it indicates the improvement in the average IIR for a group because of the addition of that individual to that group; if IA_i is negative, it indicates the decrease in this quantity because of him or her.

A numerical example of the item interrelatedness indices

The IIR indices introduced earlier will be demonstrated in terms of the fictional data set in Table 1, where the rows represent 10 respondents, *A* through *J*, who are displaying highly divergent CR behaviour on six five-point Likert-type items, *1* to *6*, of a unidimensional scale. Each of the positively keyed items, *1*, *4*, and *5*, is intended to reflect one pole of the construct involved, and the negatively keyed items, *2*, *3*, and *6*, to represent the opposite position. Notice that the terms 'positively keyed' and 'negatively keyed' do not necessarily refer to items that reflect positive and negative sentiments, respectively, regarding the attribute being measured. For example, a scale of dominance may include a (positively keyed) item 'In formal meetings, I enjoy being the chairperson', whereas a (negatively keyed) item, reflective of the opposite pole of the same attribute, would also involve a positive sentiment such as 'In formal meetings, being a regular member works best for me' (rather than a reworded version such as 'In formal meetings, I do not enjoy being the chairperson'.) The reversed scores on the negatively keyed items are indicated between brackets next to the original scores in the relevant columns. The respondents' PIC_i, PA_i, and IA_i index results are given in the last three columns, respectively. The bottom row of the column for PIC_i indicates that the sample mean inter-item correlation for the six items was 0.46. The bottom row for PA_i shows that the values for this index summed to 0.827, which was the value of the coefficient alpha formula for the total group.

An inspection of the PIC_i, PA_i, and IA_i IIR scores in the contrived example in Table 1 reveals that they have performed as expected: Individuals who have responded moderately to highly consistently, obtained positive index values, whereas those who have succumbed to CR, recorded negative values. The responses of *A* and *B* represent perfect consistency: *A* consistently endorsed the highest rating (option *e* or *5*) on the positively keyed items and consistently disapproved equally strongly of content reflective of the opposite (option *a* or *1*). As expected, because of their highly consistent responding behaviour, these respondents obtained positive IIR values, with the largest absolute values, on all of these indices. Individual *B* obtained somewhat higher PA_i and IA_i values than did *A*, because of *B*'s larger item deviations from the sample item means, and consequently higher dct_i value (cf. the column for dct_i). Respondent *C* was less consistent than *A*, and *D* was even less so, and this trend is reflected in their respective PIC_i, PA_i, and IA_i scores.

Respondents *E* through *H* were intended to represent respondents who have resorted to CR with abandon: *E*, *F*, and *G* used the same-numbered option throughout (but at differently numbered scale points), and *H*'s responses show a progressively increasing pattern (of *1*, *2*, *3*, *4*, *5*, *4*). Notice that after the reversal of the scores for items *2*, *3*, and *6*, the responses of *F*, *G*, and *H* have been 'scrambled' somewhat. However, this does not occur in the case of individual *E*, who consistently selected the middle-most option (*c* or *3* on a five-point scale) – a position that is understandably rather resistant to such score reversals. The responses of *I* and *J* were intended to mimic random responding.

When a respondent's item scores fluctuated with some being higher than the accompanying item mean and others being lower, as would be expected in the case of carelessly responding (CR) individuals, the dc_{ijk} values were, by definition, negative. This happened more often for *F*, *G*, *H*, and *I* than for *D*, *E*, and *J*. As a result, individuals *D*, *E*, and *J* returned smaller but still positive values for PIC_i and PA_i, but *F*, *G*, *H*, and *I* registered negative values for these indices. Individual *E*, who selected the middlemost position throughout, registered IIR values hovering around zero, which should be sufficient to cast doubt on any possible increase in mean IIR because of *E*. Both individuals *F* and *G*, who persevered with the same option (*4* and *5*, respectively) throughout, obtained negative values for these indices, but as *G*'s options were located relatively further away from the item scale midpoints, the resulting (negative) obtained values were higher in absolute value than those for *F*. Individual *H*, who resorted to a uniformly increasing score pattern, showed negative values on all these indices.

Typically, respondents have been eliminated in terms of CR index cutoff scores that have been developed through rational or empirical means (e.g. Huang et al., 2012). After CR screening has been concluded, its success has been evaluated in terms of, among others, the coefficient alpha formula as a measure of IIR (Huang et al., 2012), or as a measure of consistency reliability (e.g. Hong et al., 2020) for the retained group. However, the individual IIR indices developed here are intended to be used simultaneously with the other kinds of IIR indices in the screening procedure. For example, one could start by eliminating those with the poorest values and continuing up the scale until satisfactory mean values for these indices have been obtained for the retained group, or no more increases in them are observed. Obviously, Table 1 does not reflect the results of a real-world CR screening exercise but may nevertheless be useful for purposes of demonstration. For the entire group of 10 individuals, the mean PA_i equals 0.827. If respondent *G*, who has the poorest PIC_i and PA_i scores, is removed, the mean PA_i for the remaining group, increases from 0.83 to 0.89. Notice that this increase in PA_i is equivalent to the deviation of this individual's IA_i score from the mean PA_i value, as one would have expected in terms of her/his IA_i score. If individual *H*, the person with the next poorest scores, is also dropped, the mean PA_i further increases to 0.94. After the four individuals with negative PIC_i and IA_i scores (*F*, *G*, *H*, and *I*) are removed, the mean PA_i score becomes 0.95.

Discussion

As the formulae for both the PIC_i and PA_i indices are made up of the same item variances and covariances, they may be expected to be highly inter-correlated. (For the scores in Table 1, a product-moment correlation of 0.9999 has been computed.) The statistical correspondence between these indices implies that regardless of whether the PA_i index is interpreted as a measure of consistency reliability, or as the mean item-interrelatedness, like the PIC_i, at their core they are reflecting item inter-correlations. In view of this similarity, in practice, researchers are likely to give preference to the PA_i index because of its interpretability in terms of the familiar coefficient alpha, the most popular estimate of reliability (Cortina et al., 2020; Raykov & Marcoulides, 2019). Because only the highest (negative) IA_i value for a group of individuals is readily interpretable in the present context, this index is unlikely to be regularly used.

Obviously, if respondents have been screened in terms of the PA_i index, the coefficient alpha value for the eventually retained group would be known already, so that evaluating the consistency reliability of the scores for this group by means of this coefficient (at the evaluation phase) would be redundant. However, if researchers prefer nevertheless to perform a reliability analysis for the retained group, they may consider applying any of the other assessments of consistency reliability, such as those discussed by Cortina et al. (2020) and McNeissh (2018). Moreover, the eventual analyses that are typically performed on the retained group are not restricted to consistency reliability analyses but also include factor analyses that are likely to pick up serious remaining problems in consistency reliability as high coefficient alpha values do not, for example, necessarily reflect uni-dimensionality (e.g. Huang et al., 2012).

It should be pointed out that the merits of the proposed indices, just as in the case of extant consistency CR indices (cf. Huang et al., 2012), may be highly dependent on the inclusion of a balanced set of positively and negatively keyed items. Although this principle is usually incorporated in the development of standardised instruments (cf. Costa & McCrae, 1992), it is possibly less often adhered to in online administered questionnaires. At the same time a sufficient proficiency in the language in terms of which items are formulated is required to be able to coherently respond to such items.

While the removal of respondents to improve data quality is a legitimate option, caution should be exercised in rejecting sizable proportions of possibly carefully responding individuals for reasons other than a CR propensity. Human research participants constitute an indispensable part of psychological research and if the CR screening survival groups are biased in some or other way, the possibility of incorrect conclusions is a cause for concern (cf. Bowling et al., 2016). This is particularly relevant in a multilinguistic situation in which the home language of a considerable proportion of respondents may differ from the language in terms of which the scale items are presented. In such situations a comparable proportion of respondents may be screened out because of poor language proficiency rather than a tendency to indulge in CR behaviour. Of course, to prevent this from happening, a sufficiently large number of respondents should be available to begin with and care should be taken that the construct validity of the resulting measurement is not compromised through construct underrepresentation. Ultimately, greater effort should be directed at devising ways and means of preventing or reducing CR behaviour. (Electronic examination inventions at educational institutions, prompted by the COVID-19 pandemic, possibly may suggest safeguards to curtail excessive CR in the online administration of self-report surveys.)

Further empirical research may be directed at comparing how these indices fare in comparison with other extant consistency-type indices. Also, the relative effects of CR and linguistic ability on these indices may be investigated in a two-factor design in which facility with the language in which the measuring instrument is presented is completely crossed with a factor created by giving one randomly formed group instructions that strictly caution against CR behaviour and another randomly formed group for whom the instructions are maximally conducive to indulgence in CR behaviour (cf. Huang et al., 2012). In research with these indices, it should be born in mind that individuals' PIC_i, PA_i, and IA_i index scores are not experimentally independent in the sense that if these scores have been determined for $N - 1$ individuals, the score for the Nth individual would be fixed. (This also applies to the scores obtained if N individuals have been ranked from the 1st to the Nth and it has not proven to be an insurmountable barrier to research on this variable.)

Acknowledgements

The author would like to thank Prof. F de Kock (UCT) for providing him with access to several relevant journal articles.

Competing interests

The author declared that they have no financial or personal relationship(s) that may have inappropriately influenced them in writing this article.

Author's contributions

G.K.H. is the sole author of this article.

Ethical considerations

This article followed all ethical standards for research without direct contact with human or animal subjects.

Funding information

The author received no financial support for the research, authorship, and/or publication of this article.

Data availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Disclaimer

The views and opinions expressed in this article are those of the author and are the product of professional research. It does not necessarily reflect the official policy or position of any affiliated institution, funder, agency, or that of the publisher. The author are responsible for this article's results, findings, and content.

References

Arthur, W., Hagen, E., & George, G. (2021). The lazy and dishonest respondent: Detection and prevention. *Annual Review of Organizational and Organizational Behavior, 8*, 105–137. https://doi.org/10.1146/annurev-orgpsych-012420-055324

Bowling, N.A., Huang, J.L., Bragg, C.B., Khazon, S., Liu, M., & Blackmore, C.E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology, 111*(2), 218–229. https://doi.org/10.1037/pspp0000085

Conijn, J.M., Franz, G., Emons, W.H.M., De Beurs, E., & Carlier, I.V.E. (2019). The impact of careless responding in routine outcome monitoring within mental health care. *Multivariate Behavioral Research, 54*(4), 1–19. https://doi.org/10.1080/00273171.2018.1563520

Cortina, J.M., Sheng, Z., Keener, S.K., Keeler, K.R., Grubb, L.K., & Schmitt, N., Tonidandal, S., Summerville, K.M., Heggestad, E.D., & Banks, G.C. (2020). From Alpha to Omega and Beyond! A look at the past, present and (possible) future of psychometric soundness in the Journal of Applied Psychology. *Journal of Applied Psychology, 105*(12), 1351–1381. https://doi.org/10.1037/apl0000815

Costa, P.T., Jr., & McCrae, R.R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment, 4*(1), 5–13. https://doi.org/10.1037/1040-3590.4.1.5

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 2973–2944. https://doi.org/10.1007/BF02310555

Curran, P.G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66*, 4–19. https://doi.org/10.1016/j.jesp.2015.07.006

DeSimone, I.A., & Harms, P.D. (2018). Dirty data: The effects of screening respondents who provide low-quality data in survey research. *Journal of Business and Psychology, 33*(5), 559–577. https://doi.org/10.1007/s10869-017-9514-9

Dunn, A.M., Heggestad, E.D., Shanock, L.R., & Theilgard, N. (2018). Intra-individual response variability as an indicator of insufficient effort responding: Comparison to other indicators and relationships with individual differences. *Journal of Business and Psychology, 33*(1), 105–121. https://doi.org/10.1007/s10869-016-9479-0

Hong, M., Steedle, J.T., & Cheng, Y. (2020). Methods of detecting insufficient effort responding: Comparisons and practical recommendations. *Educational and Psychological Measurement, 80*(2), 312–345. https://doi.org/10.1177/0013164419865316

Huang, J.L., Curran, P.G., Keeney, J., Poposki, E.M., & DeShon, R.P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*(1), 99–114. https://doi.org/10.1007/s10869-011-9231-8

Johnson, J.A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality, 39*(1), 1031–29. https://doi.org/10.1016/j.jrp.2004.09.009

Maniaci, M.R., & Rogge, R.D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality, 48*, 61–83. https://doi.org/10.1016/j.jrp.2013.09.008

Marjanovic, Z., Holden, R., Struthers, W., Cribbie, R., & Greenglass, E. (2015). The inter-item standard deviation (ISD): An index that discriminates between conscientious and random responders. *Personality and Individual Differences, 84*, 79–83. https://doi.org/10.1016/j.paid.2014.08.021

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(4), 412–433. https://doi.org/10.1037/met0000144

Meade, A.W., & Craig, S.B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455. https://doi.org/10.1037/a0028085

Raykov, T., & Marcoulides, G.A. (2019). Thanks coefficient alpha, we still need you! *Educational and Psychological Measurement, 79*(1), 200–210. https://doi.org/10.1177/0013164417725127

Steedle, J.T., Hong, M., & Cheng, Y. (2019). The effects of inattentive responding on construct validity evidence when measuring social-emotional learning competencies. *Educational Measurement: Issues and Practice, 38*, 101–111. https://doi.org/10.1111/emip.12256

Wilkinson, L., & The Task Force on Statistical Inferences. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594–604. https://doi.org/10.1037/0003-066X.54.8.594