Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-024-02873-1 OPEN

Assessing scale reliability in citizen science motivational research: lessons learned from two case studies in Uganda

Mercy Gloria Ashepet[1,2,3]✉, Liesbet Vranken[2], Caroline Michellier[3], Olivier Dewitte[3], Rodgers Mutyebere[2,4], Clovis Kabaseke[5], Ronald Twongyirwe[6], Violet Kanyiginya[3,6,7], Grace Kagoro-Rugunda[8], Tine Huyse[1] & Liesbet Jacobs[9,10]

Citizen science (CS) is gaining global recognition for its potential to democratize and boost scientific research. As such, understanding why people contribute their time, energy, and skills to CS and why they (dis)continue their involvement is crucial. While several CS studies draw from existing theoretical frameworks in the psychology and volunteering fields to understand motivations, adapting these frameworks to CS research is still lagging and applications in the Global South remain limited. Here we investigated the reliability of two commonly applied psychometric tests, the Volunteer Functions Inventory (VFI) and the Theory of Planned Behaviour (TPB), to understand participant motivations and behaviour, in two CS networks in southwest Uganda, one addressing snail-borne diseases and another focused on natural hazards. Data was collected using a semi-structured questionnaire administered to the CS participants and a control group that consisted of candidate citizen scientists, under group and individual interview settings. Cronbach's alpha, as an a priori measure of reliability, indicated moderate to low reliability for the VFI and TPB factors per CS network per interview setting. With evidence of highly skewed distributions, non-unidimensional data, correlated errors and lack of tau-equivalence, alpha's underlying assumptions were often violated. More robust measures, McDonald's omega and Greatest lower bound, generally showed higher reliability but confirmed overall patterns with VFI factors systematically scoring higher, and some TPB factors—perceived behavioural control, intention, self-identity, and moral obligation—scoring lower. Metadata analysis revealed that most problematic items often had weak item–total correlations. We propose that alpha should not be reported blindly without paying heed to the nature of the test, the assumptions, and the items comprising it. Additionally, we recommend caution when adopting existing theoretical frameworks to CS research and propose the development and validation of context-specific psychometric tests tailored to the unique CS landscape, especially for the Global South.

A full list of author affiliations appears at the end of the paper.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Introduction

Citizen science (CS) is an approach where members of the general public contribute to scientific research in various ways including data collection, data analysis or even dissemination of research findings (Haklay, 2013; Pocock et al., 2019). This can be done in collaboration with scientists or institutions, with the aim of solving societal challenges (West, 2017). CS has several benefits including: (i) boosting research capacity by generating large volumes of data, (ii) diversification of knowledge by integrating both local and 'expert' knowledge, (iii) enhancement of social capacity in terms of providing 'informal' education (Ashepet et al., 2021; Hulbert, 2016; Pocock et al., 2019). As such, depending on the nature and goals of a project, both productivity objectives (focusing on scientific outputs) and democratization objectives (aligning scientific goals with the public interest and values) can be achieved with CS (Alender, 2016; Sauermann et al., 2020). However, in order to realize these benefits, participants not only need to be recruited but also retained especially if engagement over a long term is required.

Motivation is generally regarded as the driving force behind a given behaviour and this varies from person to person, place to place or even the kind of activity (Clary and Snyder, 1999; Fishbach and Ouré-Tillery, 2018). Therefore it is important to not only understand why people take part in given activities but also their specific social context (Beza et al., 2017; Pocock et al., 2019). Just like all volunteering activities, the success and sustainability of CS activities greatly depend on the citizen scientists who contribute their time, energy, and skills (Beza et al. 2017). As such, identifying the factors driving people into or out of action is particularly vital for those wishing to mobilize CS as this provides fundamental information for the design of recruitment and retainment strategies (Wright et al., 2015). Indeed, studies have shown that matching activities/tasks with participant motivation translates to participant satisfaction and increased retention (Alender, 2016; Finkelstein, 2008). However, literature on the motivation of CS participants and why they (dis)continue, is skewed towards cases in the Global North (Jeanmougin et al., 2017). In comparison, very few examples of studies on the factors that drive CS participants into or out of action in the Global South exist (Asingizwe et al., 2020; Beza et al., 2017; Jacobs et al., 2019; West et al., 2021; Wright et al., 2015). Therefore, we aimed to contribute to the literature by determining and assessing the motivation of the CS participants linked to three projects established in specific communities of southwest Uganda.

Several studies have explored participant motivations in the context of CS, classifying these motivational factors into different categories (Asingizwe et al., 2020). Over the years, CS motivational research has increasingly drawn inspiration from related fields such as volunteerism, psychology and sociology (Beza et al., 2017; Land-Zandstra et al., 2021; West et al., 2021). While most studies do not explicitly define a theoretical foundation that underpins their motivational categories, it is apparent that the majority are rooted in the functional approach to volunteering (Asingizwe et al., 2020; West et al., 2021). The functional approach stands as the most utilized framework to assess and understand motivations in general as well as in the CS field (Alender, 2016; West et al., 2021). Although criticized for not being exhaustive, the functional model describes six different psychological functions or reasons for volunteering (Clary and Snyder, 1999). These include (i) values—a concern for others, (ii) understanding—to gain new knowledge or skills, (iii) social—to create new or strengthen social relationships, (iv) career—to gain experience for future prospects, (v) protective—to address personal negative feelings, and vi) enhancement- to improve self and build one's self-esteem. To evaluate these motivations, Clary and colleagues (1992) established a framework known as the Volunteer Functions Inventory (VFI) built upon these six motivational categories.

In addition, the Theory of Planned Behaviour (TPB), from the field of psychology has been widely applied to predict an individual's intention to engage in a given behaviour/activity at a specific time and place. According to Ajzen, (1991), TPB proposes that behaviour is driven by intention- willingness to act—which is steered by three factors: (i) attitude—positive or negative judgements toward the behaviour, (ii) subjective norms—perceived social pressure from significant others to perform the behaviour or not, and (iii) perceived behavioural control (PBC)—perceived ease or difficulty in executing a given behaviour. Occasionally, additional variables are added to improve the predicting power of TPB. These usually include self-identity (how individuals view themselves) and moral obligation or personal norms (Chen, 2020).

The widespread use of these theoretical frameworks has been attributed to their proven and excellent psychometric properties (Chacón et al., 2017). The versatility of these frameworks has been tested across varied settings, languages and fields (Asghar, 2015; Maund et al., 2020; Niebuur et al., 2019; Wright et al., 2015; Xin et al., 2019). Given that most CS activities are often considered a form of volunteerism, these frameworks are therefore shaping the CS motivation research landscape (Agnello et al., 2022; West et al., 2021). However, tailoring of the instruments to the field of CS motivational research is still lagging; specifically, concerning the critical step of scale evaluation which is fundamental for psychometric measures (Bernardi, 1994). Scale evaluation encompasses assessing scale validity (the extent to which an instrument measures what it claims to measure rather than something else) and reliability (the extent to which an instrument gives the same measured outcome when measurements are repeated) of the instruments (Taber, 2018). To the best of our knowledge, there is currently a dearth of evidence to support scale validity and reliability in the context of CS. In this contribution, we discuss the reliability of the VFI and TPB factors when applied in the CS field.

Specifically, we aim to elaborate on the internal consistency reliability of the VFI and TPB factors when applied to a highly motivated group of people like citizen scientists. We first provide an overview of the conventional Cronbach's alpha, its applications, and limitations together with alternatives to Cronbach's alpha. We then introduce our case studies, the collected data and present the results: the calculation of Cronbach alpha without establishing the underlying data structure as often overlooked by studies (Bonett and Wright, 2015; Flora, 2020; McNeish, 2017). This is followed by assumption verification and an investigation of the effect of data transformation on alpha values. Lastly, results obtained using alternative internal consistency reliability measures—less constrained by underlying assumptions—are presented. Based on the lessons learned, we formulate recommendations for future research aiming to establish methods for analysing motivation in the CS domain.

Assessing reliability

Reliability refers to the degree of interrelatedness among item scores within a factor (Niebuur et al., 2019). For instance, consider a test where the same question is asked five times using different wording, the test is said to be reliable if the scores for all questions are nearly similar (Boslaugh and Watters, 2009, p. 357). The reliability index ranges between zero and one with values closer to one indicating a higher internal consistency and vice versa (Nimon et al., 2012). Three main approaches are used to determine the reliability of tests depending on the study context

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

and purpose: (i) test–retest reliability which assesses how consistently a test performs when administered on different occasions (repeated measurements), (ii) parallel-forms reliability which evaluates how different versions of a test measure the same concept, and lastly, (iii) internal consistency reliability focuses on the extent to which items measure the same concept i.e. how consistent are the test scores if the items/questions are slightly varied, as in multi-item tests (McNeish, 2017; Saad et al., 1999; Tang et al., 2014). Internal consistency reliability is particularly relevant when dealing with tests comprised of multiple items that are intended to measure the same unobserved concept and can thus be calculated from a single test administration (Tavakol and Dennick, 2011). In this contribution, we focus on internal consistency reliability for which several measurement indices exist but Cronbach's alpha is the most widely reported (Dunn et al., 2014).

Cronbach's alpha

Cronbach's alpha (α), hereafter referred to as alpha, measures the extent to which item responses or scores are correlated with each other by comparing the item correlations to the total variance of the items (McNeish, 2017). The general formula for α is

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k}\sigma_y^2}{\sigma_x^2}\right)$$

where k is the number of items in the test, $\sigma^2{}_y$ is the variance of each individual item i, $i = 1, \ldots, k$, and $\sigma^2{}_x$ is the variance for all items in the factor.

Therefore, alpha is a function of the number of items in a test, the item variances, and the variance of the total score. Benchmarks for acceptable values of alpha range between 0.7 and 0.9 depending on the objectives of the study (Nimon et al., 2012; Peters, 2014; Vaske et al., 2017). However, studies note that alpha values need to be interpreted cautiously (Bernardi, 1994; Bonett and Wright, 2015; Sijtsma, 2009; Taber, 2018; Tang et al., 2014). This is because what constitutes an acceptable alpha is dependent on the sample characteristics (Deng and Chan, 2016; Taber, 2018) and the seldom met underlying assumptions (Peters, 2014). Additionally, Sijtsma (2009) and others state that alpha does not reflect the internal structure of a measured factor. Dunn and colleagues (2014) thus recommend verifying the assumptions of alpha before estimating it. However, it is common for authors to only report the value of alpha without providing information on the underlying data structure (Bonett and Wright, 2015; McNeish, 2017).

Assumptions of alpha. Alpha relies on certain restrictive and often unrealistic conditions of (i) normally distributed data, (ii) unidimensionality, (iii) independent errors, and (iv) essential tau-equivalence (McNeish, 2017). An increasing number of studies have established that a transgression of these assumptions has unpredictable effects on alpha (Bernardi, 1994; Dunn et al., 2014; Flora, 2020; Sijtsma, 2009; Tavakol and Dennick, 2011; Trizano-Hermosilla and Alvarado, 2016).

Continuous and normally distributed data. Estimation of alpha assumes that the data is normally distributed given that alpha is a function of observed covariances or correlations between items (McNeish, 2017; Zumbo, 1999). The common practice for covariance estimation is to use the least-squares method often with a Pearson covariance matrix, which requires that the variables are continuous and normally distributed (McNeish, 2017; Zumbo, 1999). While most psychometric measurements rely on Likert scale scores, it has been suggested that scales with more than five levels can be treated as continuous (Flora, 2020; Trizano-Hermosilla and Alvarado, 2016). However, according to McNeish (2017) treating Likert scale scores as continuous data results in reduced covariance leading to a negative bias in alpha. This is even worsened by problematic skew which is inherent in Likert scales (Norris and Aroian, 2004; Tavakol and Dennick, 2011). Although the effect of nonnormal distribution on alpha is under-investigated, Sheng and Sheng (2012) demonstrated that alpha is a biased estimate of internal consistency reliability when the true scores of the measured items follow a non-normal distribution. Particularly they found that skewed leptokurtic distributions result in less precise estimates of alpha, an observation also reported by earlier studies (Green and Yang, 2009b).

Unidimensionality. Unidimensionality is a fundamental concept used to infer the internal structure of the data. It expresses whether the items used to measure a given concept are related to a single, common factor (Flora, 2020; McNeish, 2017; Ziegler and Hagemann, 2015). For unidimensionality to hold, the data must fit a one-factor model as this implies that the items measure only one latent variable. Therefore, differences in item responses—after discarding any random error—are due to differences stemming from a single latent variable (Ziegler and Hagemann, 2015). The dimensionality in the data is typically established by factor analytic models: confirmatory factor analysis (CFA) or exploratory factor analysis (EFA). When the relationships between items and the target construct are well established, then CFA is preferred over EFA which is utilized to uncover potential associations among the items (Flora, 2020; Ziegler and Hagemann, 2015). Dimensionality is then derived from the model-fit parameter estimates of the measurement model. Common statistical indices used to evaluate the goodness-of-fit of the measurement model include the root mean square error of approximation (RMSEA), comparative-fit index (CFI) and Tuckers–Lewis index (TLI). While lower RMSEA values (<0.08) indicate a good model fit, the reverse is true for the CFI and TLI values (>0.9). As such, poor model-fit statistics of a one-factor model suggest multidimensionality, a characteristic often associated with lower alpha values (Dunn et al., 2014; Flora, 2020; Trizano-Hermosilla and Alvarado, 2016).

Uncorrelated errors. Related to unidimensionality is the assumption of uncorrelated errors. While unidimensionality indicates whether items measure only a single underlying factor, uncorrelated errors refer to the residual variances of the observed variables that are not explained by the underlying latent factor in the model (Flora, 2020; Ziegler and Hagemann, 2015). The linear associations between the items and the latent factor expressed by unidimensionality are rooted in the classical test theory (CTT) model. CTT states that an observed score on an item is the sum of the true score (for the underlying factor e.g., attitude) and the measurement error score, i.e. observed score (X) = true score (T) + error score (E) (Gu et al., 2013). As such, CTT assumes that the measurement error scores are uncorrelated with the true scores, the error scores for different items are uncorrelated, and the sum of error scores for all the items should equal zero (Zumbo, 1999). For instance, if a factor was measured using three items (X1, X2 and X3), the value of the error components (E1, E2 and E3) associated with the three observations should not be related to the value of their true scores (i.e., the error components should not systematically be larger if the true values are larger). Secondly, the error score should be independent and unrelated for any pair of items (Boslaugh and Watters, 2009, p. 7). However, due to measurement errors or factors such as the order of items, and unmodeled dimensions, there may be some variation in the observed variables that cannot be explained by the underlying latent factor (McNeish, 2017). Moreover, although

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

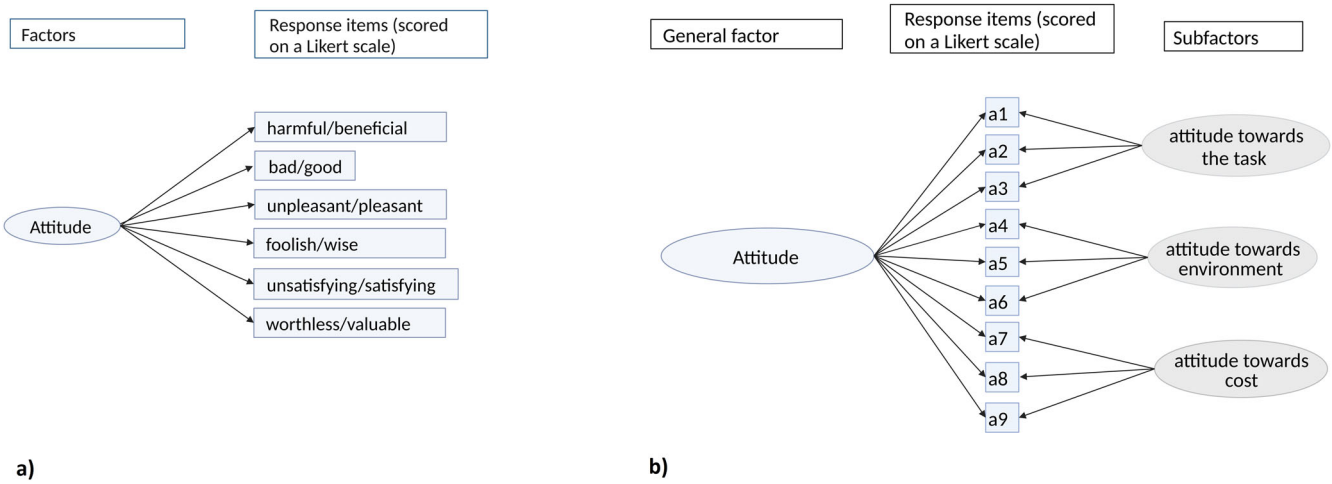

Fig. 1 Visual representation of two different models using path diagrams. In the first model (**a**), a one-factor model is illustrated for a test comprising six items, where all the items are influenced by a single underlying factor, attitude. In the second model (**b**), a bifactor model is depicted for the attitude test which includes a general attitude factor that influences all items. Additionally, the model includes three subfactors that capture the excess covariance among items with similar content.

measurement errors can arise from both random or chance factors and systematic factors that influence the measurement process consistently, derivations based on the CTT concept primarily focus on random errors as discussed in Bialocerkowski and Bragge (2008) and Henson (2001). Ideally, since random errors take no particular pattern, it is assumed to cancel out over repeated measurements. Therefore, the presence of correlated errors could highlight unexplained variations in a set of variables. Gu et al. (2013) found that uncorrelated errors inflated alpha, while McNeish (2017) noted that they generally led to unpredictable effects on alpha estimates.

Tau-equivalence. The CTT model, upon which alpha derivation is based, defines three models that measurement models can follow (Dunn et al., 2014). The models are described based on the units of measurement, degrees of precision, and/or error variances (Sheng and Sheng, 2012). These include: (i) the parallel model—the most severe model—assumes that the measurement units, precision and the error of the item scores are identical, (ii) the tau-equivalent model—similar to the parallel model but allows for differences in error variances and (iii) the congeneric model- the least restrictive and allows for variations in the measurement units, precision and error (Peters, 2014; Sheng and Sheng, 2012; Trizano-Hermosilla and Alvarado, 2016). To illustrate this, suppose we have a hypothetical six-item instrument to measure volunteer attitude towards reporting natural hazards all scored on a seven-point Likert scale. If a one-factor model (Fig. 1a) is true for the attitude factor, the standardized factor loadings of all items need to be equal for the parallel and tau-equivalent models to hold. In contrast, the congeneric model allows the standardized factor loadings among the six items to vary. In the context of alpha, alpha correctly represents the reliability of measurements that adhere to the parallel or the less restrictive tau-equivalent model (Bonett and Wright, 2015; Dunn et al., 2014; Flora, 2020). Severe underestimates of reliability using alpha have been noted when the assumption of tau-equivalence is violated and this is worsened when factors consist of fewer (<10) items which is quite often the case in practice (McNeish, 2017).

Alternative indices to estimate internal consistency

Overall, alpha is a true estimate of internal consistency reliability if the assumptions of unidimensionality, tau-equivalence, normal distribution and uncorrelated errors are not violated (Flora, 2020; Trizano-Hermosilla and Alvarado, 2016). However, in practice, many measurement factors rarely meet these assumptions, particularly the tau-equivalence assumption; some items often have strong associations with the target factor while others have weak associations (McNeish, 2017). This implies that in principle most measurement factors are congeneric or have items that measure the target factor with varying magnitudes (Dunn et al., 2014; Flora, 2020; McNeish, 2017). Several studies have noted greater negative biases of alpha, particularly when deviations from a tau-equivalent model occur (Fishbach and Ouré-Tillery, 2018; Green and Yang, 2009a; Tavakol and Dennick, 2011; Trizano-Hermosilla and Alvarado, 2016; Zhang and Yuan, 2016). In such situations, Peters (2014) cautions that "alpha is no longer a useful measure of reliability" and therefore alternative methods are recommended. While McNeish (2017) describes several alternative methods to assess reliability, we discuss two main indices fronted in literature when alpha assumptions cannot be met: omega (ω) coefficients and the greatest lower bound (GLB) (Trizano-Hermosilla and Alvarado, 2016).

Omega. Mcdonald's Omega, also known as omega total, is recommended as the most suitable estimate of internal consistency reliability when the observed variables have unequal factor loadings as in the congeneric model (Chakraborty, 2017; McNeish, 2017; Peters, 2014). Indeed, when item factor loadings are equal (tau-equivalent), omega estimates have been found to be equal to alpha but when this assumption is violated, studies have illustrated that omega is a more accurate estimate of reliability than alpha (Dunn et al., 2014; McNeish, 2017; Trizano-Hermosilla and Alvarado, 2016). However, according to Trizano-Hermosilla and Alvarado (2016), omega is also a reliable estimate when the conditions of unidimensionality and normality are met but is a better choice than alpha in the presence of skewed items. On the other hand, Flora (2020) found omega to be more robust when there is evidence for multidimensionality, item skew and when the sample size is small. Several variations of omega exist but omega hierarchical and omega total are most commonly applied. Omega hierarchy is suitable when the items used to measure a common factor form clusters leading to a multidimensional structure. In this case, a general factor influences all items along with sub-factors that capture the covariance of

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

clustered items with similar content. To demonstrate this, imagine a factor with multiple items representing a concept called "attitude towards solar panels", but some items are related to "tax", some to "environment", and others to "cost", a clustering effect might be observed due to the item-wording leading to multidimensionality with sub-factors representing "attitude towards tax", "attitude towards environment", "attitude towards cost" (Fig. 1b). A bifactor model is then usually preferred for factor analysis over the one-factor model whereby the general factor loadings are used as the parameter estimates (Flora, 2020). On the other hand, omega total is appropriate when there is evidence of a one-factor model (Fig. 1a) without subfactors (Deng and Chan, 2016; McNeish, 2017).

The greatest lower bound (GLB). Although seldom reported, GLB represents the lowest possible reliability a test can have, particularly when estimating reliability from a single test administration (Sijtsma, 2009). In the context of CTT, GLB is estimated from the covariance (Cov) matrix of the observed score (X) as the sum of the covariance of the true score (T) and the covariance of the error components (E) [i.e., $\mathrm{Cov}(X) = \mathrm{Cov}(T) + \mathrm{Cov}(E)$] (McNeish, 2017). Conceptually, GLB estimates reliability by first determining the highest possible error of the observed scores which is then used to estimate the smallest value of reliability possible (Bendermacher, 2017; McNeish, 2017). Consequently, GLB values represent interval reliability where the true reliability value of a latent factor lies between the GLB value and 1 (Sijtsma, 2009). GLB has been reported to not only exceed alpha even when all alpha assumptions are met but also to outperform omega particularly when data were skewed (McNeish, 2017; Trizano-Hermosilla and Alvarado, 2016). However, since the derivation of GLB relies on the CTT concept, it still assumes independent errors (Bendermacher, 2010). Additionally, some studies suggest that GLB overestimates reliability when the sample size is small although what constitutes a "small" sample size is not precisely defined (Bendermacher, 2017; McNeish, 2017; Trizano-Hermosilla and Alvarado, 2016).

In conclusion, the alternative reliability indices suggested are not without limitations and given that the violation of several conditions is common in practice, selecting the ideal index for estimating reliability is complex. Also, studies comparing the performance of alternative reliability indices with alpha mainly rely on simulated data with well-defined assumptions. This has led to varying conclusions and thus additional research is required to evaluate these findings, particularly when applied to real case studies (Trizano-Hermosilla and Alvarado, 2016). Therefore, Peters (2014) suggests reporting multiple reliability measures and investigating the internal structure of the data.

Data and methods

Study population. The study targeted two distinct groups: active citizen scientists participating in two CS networks and a control group composed of candidate citizen scientists. The CS networks were established in select communities in southwest Uganda as part of three projects namely:

1. Digital citizen science for community-based resilient environmental management (D-SIRe), which aimed to understand the risks posed by natural hazards. This project had two clusters of citizen scientists, one launched in 2017 in the Rwenzori region, and another launched in 2019 in the districts of Bushenyi and Buhweju.
2. Natural HAzards, RISks and Society in Africa (HARISSA) was launched in 2019, with the aim of reducing the incidence of natural hazards and associated risks in the Kigezi region. The D-SIRe and HARISSA projects together had 60 active citizen scientists also known as the 'geo-observers (GO)' and they collected data on seven natural hazards (landslides, floods, earthquakes, droughts, lightning, windstorms and hailstorms) (Jacobs et al., 2019; Kanyiginya et al., 2023; Sekajugo et al., 2022).
3. Action Towards Reducing Aquatic snail-borne Parasitic diseases (ATRAP), launched in 2020, was aimed at reducing schistosomiasis in Kagadi and Ntoroko districts and had a CS network of 25 active participants also called citizen researchers. The citizen researchers collected data on freshwater snails that transmit schistosomiasis weekly at fixed water contact sites to infer transmission hotspots and raise awareness regarding safe water practices.

Recruitment into the CS networks followed well-defined criteria carefully drafted in consultation with local stakeholders (i.e., community leaders and NGOs). The criteria included administrative requirements such as Ugandan nationality, residence in the community of interest, age over 18, proficiency in basic English, as well as flexible conditions such as gender balance, interest in community service, and past volunteering experience (Brees et al., 2021; Jacobs et al., 2019). The local leaders then nominated potential participants from which the project team selected the active citizen scientists (ATRAP = 25 and D-SIRE/HARISSA = 60), while the remaining candidate individuals included as a control group (ATRAP = 30 and D-SIRE/HARISSA = 60), had priority in case any active citizen scientist dropped out. To enable the citizen scientists to carry out their tasks smoothly, they were provided with resources, such as financial compensation to cater for costs incurred while collecting data (transport and mobile data), equipment like a smartphone for data collection, and gumboots for protection while collecting data, as well project identifiers like T-shirts and identity cards. They also received training on the topics of research, and how to navigate the open-source data collection application (KoboToolBox) before commencing the activities. Subsequently, refresher trainings to share progress, feedback and challenges were organized annually (Kanyiginya et al., 2023; Sekajugo et al., 2022). For this study, the respondents included both the citizen scientists (CSs) who are active participants in the CS networks and candidate citizen scientists or the control group (CG).

Theoretical background and measures. The study adopted a quantitative research design, primarily utilizing the well-established VFI and TPB frameworks as the theoretical foundation for our investigation into the motivations of CS participants in Uganda. The VFI and TPB items employed in the questionnaire were generated by reviewing existing literature to identify established items used in similar studies (Ajzen, 2006; Brayley et al., 2015; Clary and Snyder, 1999; Hagger and Chatzisarantis, 2006; Niebuur et al., 2019; Wright et al., 2015). These were then modified and formulated to suit the unique context of citizen science tasks described in the network overview provided above (Table 1, Supplementary information).

The VFI was assessed using the standard inventory items suggested by Clary et al (1992). Each of the VFI categories was assessed using five statements, totalling 30 items. Items related to the values function reflected the importance of helping others and showing compassion (e.g., I collect and report data on—citizen science activity—because I feel it is important to help others) whereas those measuring the protective function were oriented towards alleviating guilt and personal issues (e.g., collecting and reporting data on—citizen science activity—is a good escape from my own problems). Statements for the understanding function emphasized the pursuit of knowledge (e.g., collecting and reporting data on—citizen science activity—allows me to gain a

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Table 1 Data collection periods for the different citizen science networks and interview methods.

Group	Number of participants	Date of interview	Interview setting
Geoobservers (GO)	51	August & Nov 2019	Group
	59	Jan & Feb 2021	Individual
Control Group- GO	56	Nov 2019	Group
	41	Jan & Feb 2021	Individual
ATRAP	23	Jan 2021	Individual
	24	Oct 2021	Group
Control group-ATRAP	30	Jan 2021	Individual
	34	Nov 2021	Group

ATRAP stands for the snail monitoring CS network and GO stands for Geo-Observer which is the CS network monitoring natural hazards.

new perspective on things). Items measuring the enhancement function highlighted motives oriented towards enhancing self-esteem (e.g., collecting and reporting data—citizen science activity— increases my self-esteem) while the social function addressed motives related to societal influence (e.g., people who I'm close to want me to collect and report data on—citizen science activity). Lastly, the career function included items such as 'collecting and reporting data on—citizen science activity—allows me to explore different career options. Participants rated the VFI items on a seven-point Likert scale ranging from one (extremely inaccurate/unimportant) to seven (extremely accurate/important), indicating the alignment of their perception of the CS tasks with the provided statements.

On the other hand, items assessing the TPB factors (attitudes, subjective norms, PBC, and intention) varied in number according to Ajzen (2006). Six semantic differential scales, prompted by the common stem 'I think collecting and reporting data on—citizen science activity—'were used to assess attitudes. Response choices (e.g., bad/good, foolish/wise) were indicated on a seven-point Likert scale, with higher scores suggesting a more positive attitude. Six items were used to assess subjective norms, five for PBC and three for intention, similarly responded to on a seven-point Likert scale, with choices ranging from one (strongly disagree) to seven (strongly agree). To increase the predicting power of the TPB, self-identity and moral obligation were included in the model as in Hagger and Chatzisarantis (2006) and (Chen, 2020). Self-identity and moral obligation were assessed using three items each, also evaluated using a seven-point Likert scale ranging from one (strongly disagree) to seven (strongly agree). Altogether, the TPB items amounted to 26 items.

Data collection. To gather data, we designed a semi-structured questionnaire with three main sections: (1) the first section gathered personal information, such as gender, age, and education level; (2) the second section contained the VFI questions; and (3) the last section covered the extended TPB factors. To avoid bias, the VFI and TPB questions were randomly placed in the questionnaire. The questionnaire was administered in a face-to-face fashion providing an opportunity for immediate clarification of the questions, back-translation to minimize misinterpretation and capture nuanced information beyond the structured questions (Mukherjee et al., 2018; Young et al., 2018). Notes and expressions were thus recorded for additional context. The questions were posed in a conversational style and answered by the respondents individually under two settings: individual and group-based sessions, henceforth referred to as Individual (I) and Group (G) interviews respectively. While individual sessions provided a more private environment for participants to express their personal perspectives, group-based sessions were aimed at creating a supportive environment and fostering a sense of comfort and security (Milewski and Otto, 2017). The sessions, conducted by trained researchers within the CS projects, lasted about an hour and occurred at different moments between 2019 and 2021 (Table 1). The first round of data collection was followed by a second round after a refresher training to identify the potential effects of CS organizational designs. This generated four data sets based on the interview method and the CS network namely ATRAP_I ($n = 53$), ATRAP_G ($n = 58$), GO_G ($n = 107$) and GO_I ($n = 100$) whereby the division of the participants over control group and participants is specified in Table 1.

Data analysis. Data analysis was organized in three stages: (1) a priori analyses of reliability using Cronbach's alpha, (2) examination of assumptions related to the internal structure for each factor, and (3) reliability analysis using alternative indices. Stage one focused on reverse coding negatively worded items particularly in the TPB, descriptive statistics (i.e., mean, standard deviation, skewness, kurtosis) and internal consistency reliability using alpha for each factor in the Psych package (Revelle, 2015). Missing data, specifically occurring in the GO group interviews, were excluded from the analysis. For completeness, analyses were repeated for the respondent groups (CSs and CG) separately and results were provided in the supplementary information.

In the second stage, analyses to investigate the internal structure of the factors and the assumptions of alpha (i.e., normality, unidimensionality, uncorrelated errors and tau-equivalence) were conducted according to Flora (2020) and McNeish (2017). First, item analysis was conducted to evaluate the correlation between individual items and the total-item correlation per factor. Ideally, these correlations should surpass 0.2 (Boonyaratana et al., 2021; Niebuur et al., 2019). Next, item distribution was assessed using: (i) the Shapiro–Wilk test of normality and (ii) the degree of distortion from a normal distribution or skewness, considering values above ±2 as problematic, in line with Muzaffar (2016). Transformations were applied to address the identified skew following the recommendations of Tabachnick et al. (2013), and subsequently assessing alpha with the transformed data (Norris and Aroian, 2004). Then one-factor models based on our prior theoretical knowledge for each VFI and TPB factor were specified using confirmatory factor analysis within the *Lavaan* package (Rosseel, 2012). The CFA models were freely estimated using the robust maximum likelihood (MLR) to account for deviations from normal distributions inherent in Likert scale data (Flora, 2020). We assessed the unidimensionality assumption based on recommended model fit parameters, specifically CFI/ TLI ≥ 0.93 and RMSEA/SRMR ≤ 0.08 according to Leach et al. (2008). Additionally, the residual or observed correlation matrices for the items from the CFA models were analysed to investigate the presence of error correlations. We defined residual correlation between item pair values exceeding 0.1 as notable, given that smaller error

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Table 2 Demographics of the study participants.

Variable	CS network					
	ATRAP		Geoobservers		TOTAL	
	n	%	*n*	%	*n*	%
Gender						
Male	41	77.4	78	72.9	119	74.4
Female	12	22.6	29	27.1	41	25.6
Age						
20–29	7	13.2	43	40.3	50	31.2
30–39	27	50.9	46	43	73	45.6
40–49	16	30.2	11	10.3	27	16.9
+50	3	5.7	6	5.6	9	5.62
Education level						
Low (primary)	11	20.8	16	15.1	27	16.9
Lower secondary	19	35.8	24	22.4	43	26.9
Upper secondary	6	11.3	23	21.5	29	18.1
High (University & Tertiary)	17	32.1	44	41.1	61	38.1
Employment status						
Employed (salary)	8	15.1	17	15.9	25	15.6
Self-employed	40	75.5	76	71	116	72.5
Unemployed	5	9.4	14	13.1	19	11.9

correlation values are generally seen as indicative of a good fit between the model and the observed data (Flora, 2020; Gerbing and Anderson, 1984). Since the CFA models were freely estimated, the tau-equivalence assumption was established by assessing the standardized factor loadings of the items per factor. To elaborate on the tau-equivalence violation, corresponding CFA-constrained models were estimated and a model chi-square difference test also known as the likelihood ratio test conducted, comparing the freely estimated models with the constrained models following the recommendations of McNeish (2017). The models were compared using the ANOVA function from the *Lavaan* package, rejecting the null hypothesis if the likelihood ratio test results were significant ($p < 0.05$), implying that the constrained model fits the data significantly worse than the freely estimated model. In such cases, violation of tau-equivalence was confirmed and thus we proceeded with the freely estimated model (McNeish, 2017; Rosseel, 2012). Conversely, when the null hypothesis was accepted, model selection was based on descriptive model comparison, selecting the model with a lower AIC (Akaike Information Criterion) (Werner and Schermelleh-Engel, 2010). Furthermore, we employed the tau.test function from the *coefficientalpha* package by Zhang and Yuan (2016) to assess both the tau-equivalence and homogeneity of the items. The test utilizes the robust *F*-statistic which is particularly beneficial in cases of small sample sizes where chi-square tests may yield less stable results. The *F*-statistic is considered more reliable under such circumstances, contributing to the test's suitability for our study (McNeish, 2020). According to Zhang and Yuan (2016), outlying or extreme observations can be controlled by adding a down-weighting rate (e.g., varphi = 0.1) or assigning them lower weights. However, in this study, we opted not to downweigh any scores since our main aim was to estimate the tau-equivalence and homogeneity of items, thus using all the data for analysis.

Alternative reliability indices (omega total and GLB) were computed and reported alongside alpha as suggested by Revelle and Condon (2019). For meaningful comparison, we set the reliability threshold to 0.7 and above for all indices, since values closer to one are considered acceptable (Nimon et al., 2012). Omega total was computed from the freely estimated CFA models specified using the reliability function of the *semTools* package by Jorgensen et al. (2022) while GLB was estimated from the *glb* function of the *psych* package (Revelle and Condon (2019). Finally, we complement these quantitative observations with qualitative insights obtained from metadata or secondary data documented by the first author during the interviews. Specifically, we extract quotes from the notes to provide additional context to the quantitative data and to enhance the interpretability and relatability of the results.

All analyses were conducted in R version 4.2.2 and the criterion for establishing statistical significance was set at $p \leq 0.05$. The simplified R script with the code used for the analysis is provided (see the "Data availability" statement below).

Results

Participant characteristics. Table 2 shows that the respondents were predominantly male (74%), between 20 and 59 years of age (Mean 34, SD. 8 years). Although 38% of all participants had a high education level (completed university or tertiary training) most of them belonged to the Geoobserver network (44/61). Also, under 20% of the participants had a salaried job while the majority (72%) of the participants were self-employed in different sectors predominately agriculture.

Factor analysis and a priori Cronbach alphas. Altogether 48 factors were examined, encompassing six VFI and six TPB factors pooled from two different audiences with two different interview settings. Figure 2 presents the mean, standard deviation, and alpha values for these factors. The factors displayed high average scores ranging from 4.6 (PBC factor; ATRAP_G) to 6.7 (attitude factor; ATRAP_G). For the VFI, 'Understanding', 'Values', 'Career' and 'Enhancement' factors across both CS networks received the highest scores while the 'Social' and 'Protective' motivational factors were consistently scored low regardless of the interview method. Conversely, the 'Attitude' and 'Intention' factors of the TPB received the highest scores while 'PBC' consistently received low scores (Fig. 2).

The internal consistency reliability using alpha varied across factors ranging from $\alpha = 0.91$ (attitude factor; GO_G) to $\alpha = -0.03$ (PBC factor; GO_I) for the TPB factors and $\alpha = 0.83$ (protective factor; ATRAP_G) to $\alpha = 0.34$ (values factor; ATRAP_G) for the VFI factors. Overall, most VFI factors demonstrated relatively adequate alpha values, with 8 of the 24 analysed factors surpassing the recommended 0.70 criterion. In contrast, only two TPB factors (attitude and subjective norms) presented adequate alpha values while PBC, intention, moral obligation, and self-identity consistently exhibited low unacceptable alpha values. Notably, a persistent and consistent pattern in the average scores and alpha values was observed across the CS networks and interview settings, despite the interviews occurring several months apart. Note that these patterns generally do not change when splitting the data into the different respondent groups (i.e., active and control groups), as demonstrated in Fig. 1a, b, Supplementary information.

Internal structure and underlying alpha assumptions

Normality and skew. Table 3 shows the mean, item-total correlation, and standardized factor loadings for the analysed VFI and TPB items, while the specific item wording can be found in Table 1, Supplementary information. Generally, the average scores of the items were high, ranging from 3.5 (item PBC4: ATRAP_G) to 6.9 (item U2: ATRAP_I). Overall, the majority of the VFI items surpassed the item-total correlation of 0.2 while some items for most TPB factors didn't and at times had negative correlations. Regarding skew, most items were negatively skewed, with values ranging between −4.4 (item A3: ATRAP_G) and 0.05 (item PBC2: ATRAP_G), and only 6 items in the intention and PBC factors

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

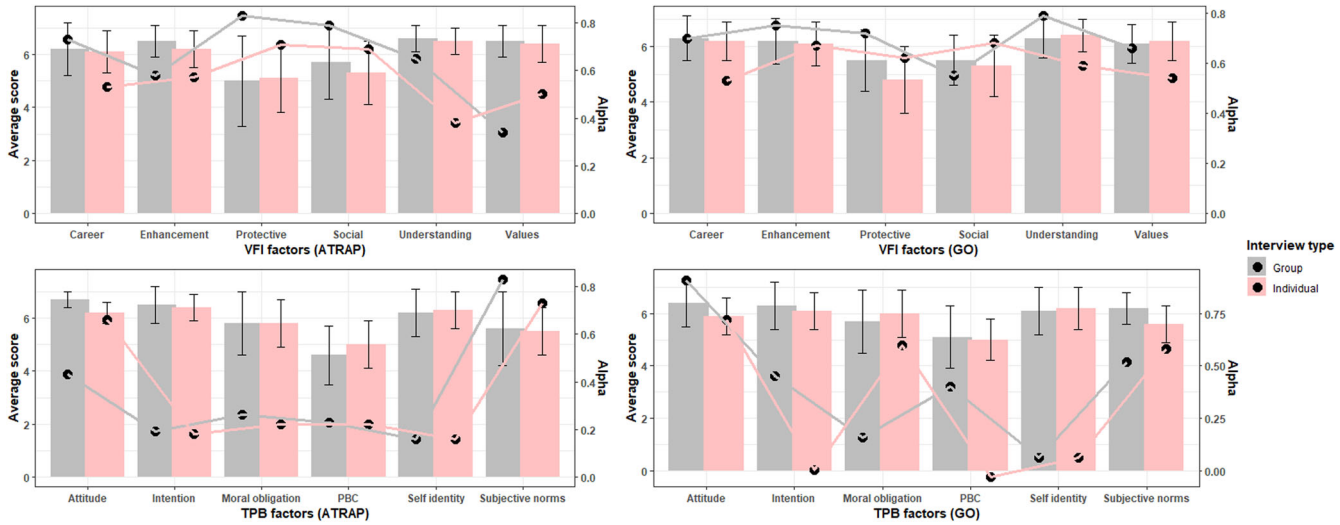

Fig. 2 Average scores (depicted with bars) with standard deviation for the VFI and TPB (top and bottom) factors and alpha values (depicted with points) represented per interview type (group or individual/face-to-face) and per CS network (ATRAP and GO) respectively where ATRAP is the CS network monitoring freshwater snails while GO is the CS network that monitors natural hazards. PBC is perceived behavioural control. The trend of alpha values is depicted by a line connecting the values of each factor.

showing positive skew (Fig. 3). Additionally, apart from items belonging to the social and the protective factor, all other factors had at least one or two items with skew exceeding the recommended threshold of >±2 for psychometric tests (Fig. 3) according to Muzaffar (2016). As such most factors analysed in this study suffered from problematic skew. Furthermore, the Shapiro–Wilk test for normality obtained significant results ($p < 0.001$) for all the items, rejecting the assumption of normality. The general negative skew persisted when splitting the data into control and active groups (Fig. 2a, b, supplementary information).

Since we detected negative skew in the items (average skew = −1.8), we performed log and inverse transformations to make the item distributions more symmetrical. This was done by first 'reflecting' the items and then applying the transformation methods. Reflecting entails first obtaining the largest score in the items and then creating a new variable by subtracting each score from the largest score plus one, i.e. $\max(x+1)-x$ (Tabachnick et al., 2013, p. 87). Following the log and inverse transformation, the skew was ameliorated, with most item skew values approaching the acceptable threshold of ±2 (Fig. 3). However, despite this improvement, when analysing the transformed data, all item distributions remained significantly non-normal ($p < 0.05$), thus failing to attain the assumption of normality. Given that the applied transformations did not resolve normality nor lead to higher alphas (Table 4), we proceeded with the raw data for the subsequent analysis.

Unidimensionality, correlated errors and tau-equivalence. As stated earlier, the results of the factor analysis allowed us to infer and verify the three assumptions of alpha for each of the factors analysed. Model fit indices for the freely estimated CFA models (Table 3) show that less than half of the analysed models (Fig. 4a) met the criteria of adequate model fit parameters discussed above. Frequently, the CFI and TLI values were below 0.93, while the RMSEA and SRMR values exceeded 0.08, especially for the TPB factors of PBC, intention, moral obligation, and self-identity (Table 2, Supplementary information). As such, the assumption of unidimensionality was rejected for more than half of the freely estimated CFA model fit parameters. However, the unidimensionality assumption based on the *F*-statistic using the tau.test function suggests that unidimensionality can be explicitly rejected for only three factors (Fig. 4b). Furthermore, the residual correlation matrix (Table 3, supplementary information) showed that most factors had at least a third of the item pairs exhibited substantial error correlation ($r > 0.1$), thereby violating the assumption of uncorrelated errors. Nevertheless, for some factors, particularly those with the lowest number of items (moral obligation, self-identity, and intention), almost all item pairs showed no violation of this assumption (Fig. 5).

Additionally, the standardized factor loadings of the freely estimated CFA models presented in Table 3, revealed dissimilar item loadings, often with a combination of weak (<0.4) and strong (>0.5) relationships between the items per factor. The wide range of the item factor loadings (e.g., the range for a career; GO_G was 0.05–0.93) provides evidence of non-tau-equivalence for all the factors. In contrast, the likelihood ratio test detected non-significant changes in the chi-square value for most compared models, indicating no significant difference between the constrained and freely estimated models for most factors. In such cases, we proceeded with the most parsimonious model, in which case, the freely estimated model often outperformed the constrained model (Fig. 4c). Conversely, the tau.test function provided evidence for tau-equivalence in most factors (Fig. 4d), with overlaps and agreements (e.g., understanding: GO_G) between the tau.test results and the chisq difference. However, it is important to note that the different tests applied here neither uniformly reject nor systematically confirm the assumption of tau-equivalence.

Alternative reliability indices. Given the evident violations of alpha assumptions, Table 4 presents alternative measures of reliability (omega total and GLB) alongside the a priori alpha values and alpha after transformation as recommended by various studies (McNeish, 2017; Peters, 2014; Sijtsma, 2009; Trizano-Hermosilla and Alvarado, 2016; Zhang and Yuan, 2016). Among all the reliability indices reported, alpha is systematically characterized by low estimates. Moreover, no big changes in alpha were observed with the transformed data regardless of the transformation method, except for minor differences, particularly for TPB factors. Notably, the TPB factors PBC, intention, moral obligation and self-identity, maintained values significantly below the recommended threshold even after transformation. Using omega total and GLB on the other hand, resulted in noticeably

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Table 3 Descriptive statistics (mean, item-total correlations, and factor loadings) of the items of each factor of the Volunteer Functions Inventory and Theory of Planned Behaviour.

CS network		ATRAP						GO					
Interview setting		Individual			Group			Individual			Group		
Factors	Items	Mean	Corr	λ	Mean	Corr	λ	Mean	Corr	λ	Mean	Corr	λ
Understanding	U1	6.4	0.1	−0.4	6.7	0.4	0.2	6.4	0.4	0.4	6.4	0.6	0.5
	U2	6.9	0.4	−0.3	6.8	0.5	0.3	6.7	0.1	0.1	6.6	0.6	0.5
	U3	6.3	0.1	0.0	6.4	0.5	0.8	6.6	0.4	0.4	6.5	0.6	0.7
	U4	6.1	0.4	−0.1	6.4	0.5	0.6	5.9	0.5	1.0	6.3	0.6	0.5
	U5	6.7	0.3	−0.2	6.8	0.3	0.2	6.3	0.4	0.5	6.4	0.6	0.9
Values	V1	6.3	0.0	0.1	6.6	0.0	0.1	6.5	0.1	0.1	6.5	0.4	0.3
	V2	6.1	0.4	−0.9	6.2	0.0	0.1	5.9	0.4	0.6	5.5	0.5	0.9
	V3	6.2	0.4	−0.8	6.2	0.4	1.5	6.3	0.5	0.8	6.2	0.4	0.6
	V4	6.5	0.0	0.0	6.8	0.1	0.1	6.0	0.3	0.6	6.0	0.5	0.7
	V5	6.6	0.6	−1.0	6.5	0.3	0.5	6.4	0.3	0.4	6.2	0.4	0.5
Career	C1	6.1	0.3	0.5	5.9	0.7	1.4	6.0	0.4	0.9	6.4	0.4	0.4
	C2	6.5	0.4	0.3	6.3	0.5	0.9	6.2	0.1	0.0	6.2	0.4	0.6
	C3	5.6	0.3	0.8	5.6	0.5	1.0	6.1	0.4	0.8	6.3	0.7	0.9
	C4	5.9	0.5	1.3	6.3	0.7	1.0	6.1	0.4	0.6	6.3	0.4	0.7
	C5	6.6	0.1	0.1	6.7	0.3	0.2	6.4	0.3	0.3	6.5	0.5	0.6
Social	S1	5.5	0.6	1.1	5.5	0.5	1.5	5.4	0.4	1.0	5.1	0.4	1.0
	S2	4.1	0.4	1.2	5.6	0.8	2.0	4.3	0.6	1.7	5.4	0.5	1.0
	S3	5.6	0.6	1.1	6.0	0.4	0.7	5.5	0.4	0.6	5.7	0.4	0.8
	S4	5.4	0.3	0.5	5.7	0.6	1.2	5.5	0.5	0.8	5.9	0.2	0.3
	S5	5.7	0.4	1.0	6.0	0.6	1.0	5.5	0.3	0.5	5.5	0.2	0.4
Enhancement	E1	6.1	0.1	0.1	6.7	0.4	0.5	6.1	0.3	0.4	5.9	0.5	0.8
	E2	6.1	0.4	0.7	6.6	0.4	0.2	6.0	0.6	1.1	6.1	0.6	0.9
	E3	6.3	0.3	0.3	6.7	0.3	0.5	6.2	0.5	0.7	6.2	0.5	0.6
	E4	6.3	0.5	0.8	6.3	0.4	0.5	6.2	0.4	0.6	6.5	0.6	0.7
	E5	6.0	0.4	0.8	6.3	0.3	0.4	6.2	0.4	0.4	6.5	0.4	0.5
Protective	P1	4.8	0.3	0.6	5.0	0.7	1.7	4.6	0.5	1.2	5.2	0.5	1.1
	P2	5.1	0.4	0.9	4.7	0.6	1.4	5.0	0.4	0.8	5.5	0.5	1.0
	P3	5.2	0.5	1.2	5.2	0.6	1.6	4.4	0.2	0.2	5.7	0.5	0.8
	P4	5.2	0.6	1.7	4.3	0.6	1.6	5.3	0.4	0.8	5.9	0.5	0.9
	P5	5.1	0.6	1.4	5.7	0.6	1.2	4.8	0.5	1.5	5.4	0.5	1.0
Attitude	A1	6.4	0.7	0.2	6.8	0.3	0.2	6.3	0.7	0.8	6.4	0.6	0.5
	A2	6.4	0.5	0.1	6.7	0.3	0.2	6.2	0.5	0.4	6.5	0.7	0.5
	A3	6.4	0.7	0.3	6.6	0.1	0.1	6.1	0.7	0.8	6.5	0.9	0.9
	A4	6.2	0.5	0.1	6.8	0.3	0.2	6.0	0.5	0.4	6.3	0.8	1.0
	A5	5.8	0.5	0.6	6.6	0.4	0.4	6.0	0.5	0.8	6.3	0.8	1.0
	A6	5.8	0.3	0.8	6.5	0.2	0.2	4.8	0.3	0.5	6.3	0.8	1.0
Subjective norms	SN1	6.5	0.3	0.3	6.1	0.4	0.4	6.3	0.3	0.3	6.3	0.3	0.6
	SN2	5.0	0.3	1.1	5.7	0.7	1.4	5.5	0.3	0.4	6.1	0.1	0.2
	SN3	6.3	0.1	0.5	5.8	0.7	1.7	6.1	0.1	0.1	6.5	0.5	0.8
	SN4	4.9	0.4	1.5	5.1	0.6	1.4	5.4	0.4	0.9	6.4	0.2	0.1
	SN5	4.6	0.4	1.5	5.2	0.6	1.2	4.7	0.4	1.1	5.8	0.3	0.5
	SN6	5.6	0.4	0.4	5.4	0.6	1.5	5.8	0.4	0.7	6.1	0.3	0.4
PBC	PBC1	6.0	0.0	−0.1	6.0	0.2	0.6	6.0	0.0	0.2	5.8	0.1	0.2
	PBC2b	4.3	0.0	−0.3	4.0	0.2	0.3	3.7	0.0	0.5	4.5	0.2	0.7
	PBC3	5.0	0.1	1.9	3.7	0.4	1.3	5.3	0.1	−0.8	4.9	0.5	2.3
	PBC4	3.8	−0.1	1.0	3.5	0.1	1.4	4.1	−0.1	−1.4	5.1	0.2	0.8
	PBC5b	6.0	−0.1	−0.1	5.7	−0.2	−0.7	5.9	−0.1	0.6	5.2	0.1	0.3
Intention	INT1	6.5	0.0	0.5	6.7	0.0	0.2	6.5	0.0	0.4	6.5	0.4	0.5
	INT2	6.5	0.0	0.1	6.7	0.5	0.6	6.3	0.0	0.1	6.7	0.3	0.2
	INT3b	6.1	0.0	0.2	6.0	0.1	0.6	5.7	0.0	−0.1	5.7	0.4	1.5
Moral obligation	MO1	5.1	0.0	0.4	5.1	0.2	0.4	5.7	0.0	0.1	4.9	0.0	0.2
	MO2	5.6	0.1	1.8	5.3	0.2	2.3	5.9	0.1	1.6	5.9	0.1	0.3
	MO3	6.6	0.0	0.0	6.9	0.0	0.0	6.5	0.0	0.0	6.4	0.2	1.5
Self-identity	SI1	6.6	0.5	0.1	6.4	0.4	0.9	6.4	0.5	0.4	6.5	0.1	0.8
	SI2b	6.1	0.4	0.2	6.1	−0.2	−0.4	6.0	0.4	0.6	5.4	0.0	0.1
	SI3	6.3	0.5	1.1	6.0	0.2	1.5	6.3	0.5	0.9	6.5	0.0	0.1

PBC is perceived behavioural control, corr refers to the item-total correlation, λ = factor loadings and items with b (e.g., PBC2b) were reverse-coded items.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Fig. 3 Skew values for the VFI and TPB items across the CS networks and interview type for both the original data without transformation and the transformed data (log and inverse transformation). Colour shading shows items with problematic skew above the recommended threshold (±2) for psychometric tests in red while values that meet this criterion are indicated in green.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Table 4 Internal consistency indices for the Volunteer Functions Inventory (VFI) and Theory of Planned Behaviour (TPB) factors using Cronbach's Alpha, McDonald's Omega, and Greatest Lower Bound (GLB).

CS network/ interview setting	Reliability coefficients	VFI factors						TPB factors					
		Understanding	Values	Career	Social	Enhancement	Protective	Attitude	Subjective norms	PBC	Intention	Moral obligation	Self-identity
GO_I	Alpha raw	0.59	0.54	0.53	0.68	0.67	0.62	0.72	0.58	-0.03	0	0.05	0.6
	Alpha_Log	0.57	0.55	0.6	0.66	0.65	0.65	0.78	0.63	0.09	0.15	0.14	0.59
	Alpha_Inverse	0.55	0.54	0.62	0.61	0.61	0.68	0.81	0.65	0.24	0.31	0.33	0.57
	$Omega_t$	0.65	0.57	0.58	0.71	0.69	0.64	0.74	0.62	0.07	0.04	0.44	0.63
	GLB	0.72	0.62	0.66	0.72	0.77	0.75	0.85	0.71	0.36	0.15	0.34	0.67
GO_G	Alpha raw	0.79	0.66	0.7	0.55	0.75	0.71	0.91	0.52	0.4	0.45	0.16	0.05
	Alpha_Log	0.69	0.66	0.69	0.61	0.72	0.72	0.88	0.62	0.39	0.56	0.26	0.21
	Alpha_Inverse	0.7	0.69	0.67	0.62	0.71	0.68	0.86	0.65	0.35	0.59	0.34	0.34
	$Omega_t$	0.79	0.68	0.72	0.58	0.76	0.72	0.92	0.51	0.52	0.63	0.31	0.15
	GLB	0.86	0.76	0.79	0.66	0.84	0.79	0.96	0.68	0.62	0.56	0.43	0.39
ATRAP_I	Alpha raw	0.38	0.49	0.53	0.69	0.57	0.71	0.66	0.73	0.22	0.18	0.22	0.16
	Alpha_Log	0.54	0.49	0.56	0.72	0.57	0.68	0.66	0.73	0.11	0.35	0.38	0.29
	Alpha_Inverse	0.58	0.5	0.56	0.71	0.58	0.6	0.64	0.68	0.03	0.43	0.52	0.35
	$Omega_t$	0.27	0.58	0.58	0.7	0.6	0.72	0.7	0.78	0.37	0.26	0.52	0.5
	GLB	0.72	0.73	0.72	0.79	0.72	0.79	0.8	0.82	0.52	0.44	0.42	0.42
ATRAP_G	Alpha raw	0.65	0.34	0.73	0.79	0.58	0.83	0.43	0.83	0.23	0.19	0.26	0.16
	Alpha_Log	0.69	0.49	0.79	0.78	0.7	0.84	0.63	0.84	0.18	0.33	0.32	0.17
	Alpha_Inverse	0.7	0.56	0.8	0.76	0.73	0.82	0.69	0.85	0.17	0.43	0.42	0.2
	$Omega_t$	0.71	0.57	0.76	0.81	0.51	0.83	0.44	0.84	0.3	0.38	0.64	0.49
	GLB	0.76	0.48	0.83	0.84	0.77	0.88	0.67	0.89	0.49	0.55	0.45	0.55

Colour shading represents ranges of values from minimum (dark red) to acceptable values of 0.7 (dark green).
Alpha raw=alpha with original data, *Alpha_Log*= alpha when data was log-transformed, *Alpha_Inverse* =alpha when data was inverse-transformed and PBC is Perceived Behavioural Control.

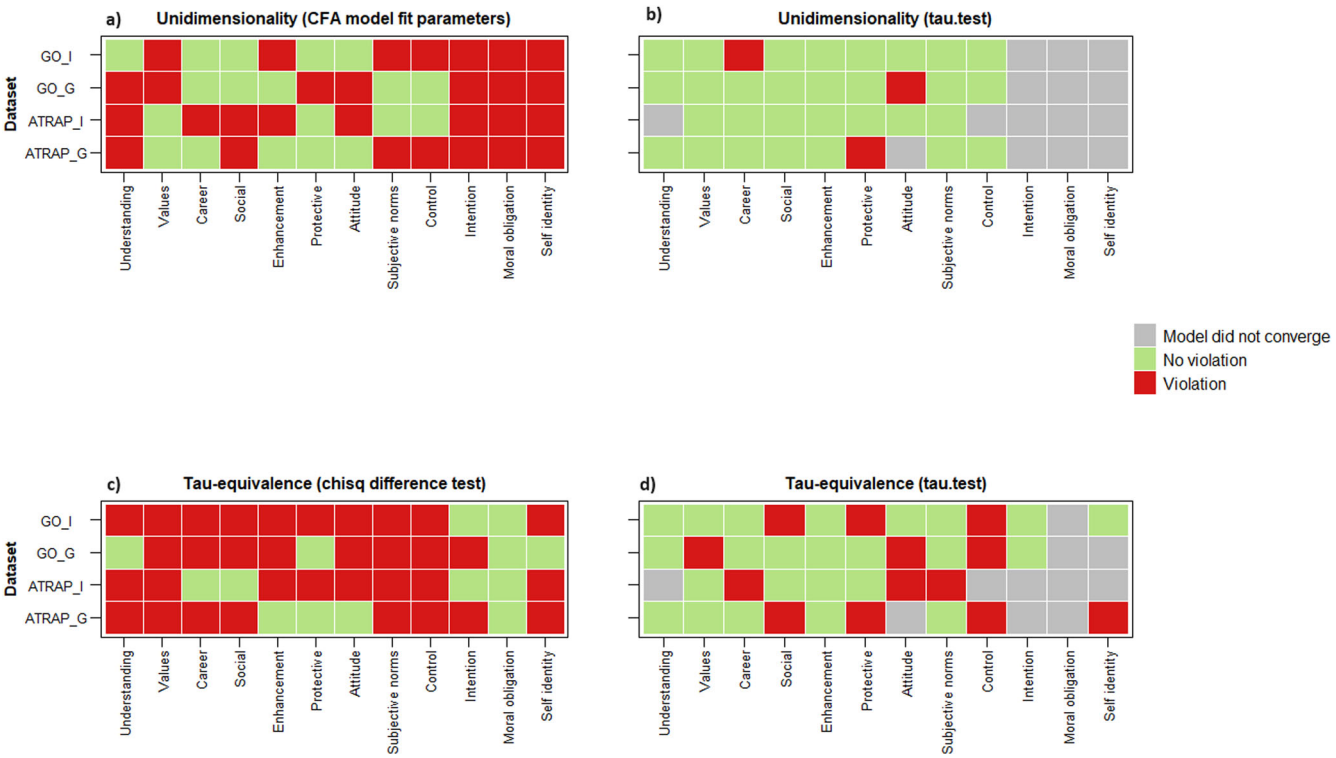

Fig. 4 The figure illustrates the degree of unidimensionality and tau equivalence violation by the VFI and TPB factors, using different methods. Unidimensionality assumption violation is identified when **a** CFI & TLI < 0.93 and RMSEA & SRMR < 0.08 based on freely estimated CFA models, and **b** $p < 0.05$ using the *F*-statistic tau.test function. Conversely, tau equivalence is rejected when **c** $p < 0.05$ for both model comparisons using the chisq difference test and **d** the *F*-statistic test using tau.test.

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

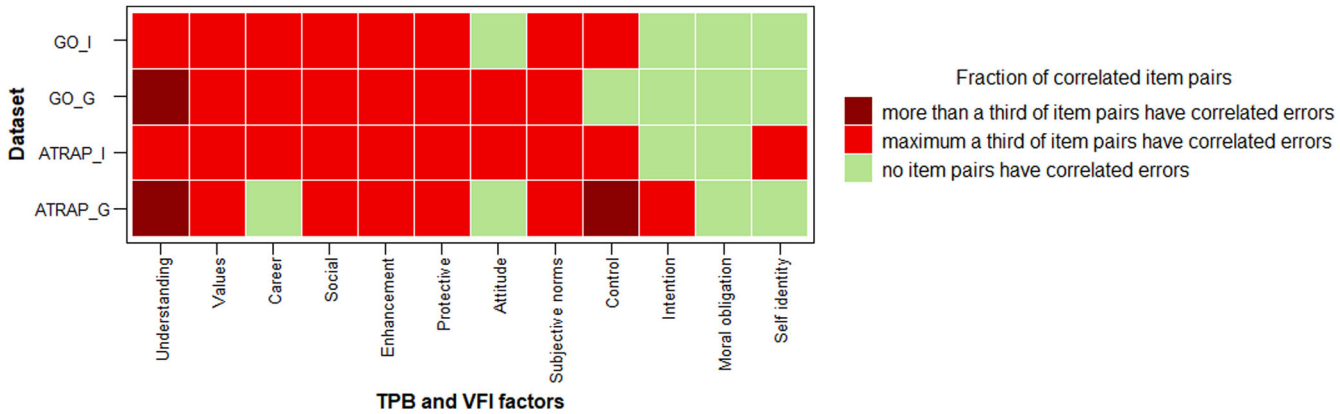

Fig. 5 An illustration of item pairs exceeding a residual error correlation threshold of 0.1 per factor, represented as a fraction derived from the total possible item pairs within each factor (colour shade represents level of violation i.e., Dark Red shade = high violation and green shade = no violation). ATRAP stands for the snail-borne focused CS network, GO is the natural hazard focused CS network, I is the individual interview setting while G represents the group interview settings.

higher indices: patterns remain consistent when considering the active and control groups (Table 4, supplementary information). We also notice small differences between omega and alpha particularly when a factor conformed to tau-equivalence (i.e., understanding: GO_G and protective: ATRAP_G). However, GLB outperformed both alpha and omega, with values often meeting the acceptable criteria (>0.7) for most VFI factors and two TPB factors (attitude and subjective norms). These results align with earlier detected patterns using alpha (Fig. 2), whereby the factors of PBC, intention, moral obligation, and self-identity consistently fell below the reliability threshold of 0.7. It is also important to note that these factors are characterized by a lower number of items that often-had weak item-total correlations, limiting the exclusion of such items (Boonyaratana et al., 2021). Lastly, although respecifying the models to address large error correlations improved model fit for almost all factors (Table 5, Supplementary information), the omega estimated from the respecified model is not reported. This is because the large residual correlations associated with these models are indicative of multidimensionality and thus reporting reliability estimates after accounting for error correlations requires appropriate justification of the causes of the error correlations (Flora, 2020; McNeish, 2017).

A qualitative interpretation of the VFI and TPB questions. The metadata documented during the interviews showed that participants faced challenges in understanding the meaning of several VFI and TPB questions (Table 5). These challenges were expressed through various participant reactions, ranging from confusion to laughter, indicating that the questions or the wording of these specific items seemed irrelevant or inapplicable to their specific context (Ajzen, 2006; Boonyaratana et al., 2021). To illustrate this, we consider and present 'problematic' items that received more than two reactions and corresponding quotes in Table 5, highlighting participants' struggles with these particular questions. A closer examination of these problematic items revealed interesting patterns (Tables 5 and 6, supplementary information), with the majority of such items exhibiting weak item-total correlations ($r < 0.3$). The exclusion of such items indeed led to notable improvements in the omega values (see Table 4, Supplementary information).

In summary, our results demonstrate that all factors violate at least one alpha assumption, data transformation has minimal impact on the alpha scores while some factors specifically in the TPB, systematically demonstrate lower reliability regardless of which reliability estimate is used. Furthermore, metadata from fieldnotes reveals participant challenges in understanding certain questions, as depicted by most of these items exhibiting weak item–total correlations.

Discussion

The volunteer functions inventory (VFI) and the theory of planned behaviour (TPB) stand as the most widely applied frameworks for understanding motivation and intentions underpinning participation in various activities. However, the credibility of measurements derived from these frameworks hinges mainly on their reliability (Zijlmans et al., 2019). The traditionally reported alpha is constrained by its strict, often violated assumptions (McNeish, 2017; Streiner, 2003). Moreover, what constitutes an acceptable reliability estimate depends on the type of application and purpose of the study according to Bonett and Wright (2015) and Vaske et al. (2017). In this study, we illustrate the roadblocks encountered when assessing the reliability of such psychometric tests (e.g., the VFI and TPB), solely using the commonly reported Cronbach's alpha within the context of citizen science. Additionally, we present evidence of deviations from alpha assumptions, explore alternative less restrictive indices, and reflect on the questions used for the study.

First, internal consistency reliability assessed using alpha fell below the recommended threshold (>0.7), particularly for the TPB factors: PBC, intention, moral obligation, and self-identity. With the latter three factors having the lowest number of items, Tavakol and Dennick (2011) suggest that adding more related items could enhance alpha. On a positive note, the relatively adequate estimates for the VFI factors in this study align with findings from similar studies investigating motivations in a CS context (Maund et al., 2020; Wright et al., 2015). Second, the factors analysed often violated one or more assumptions of alpha, rendering alpha an unsuitable measure of reliability (Bonett and Wright, 2015). In particular, violation of the tau-equivalence assumption, as rigorously demonstrated in this study, may falsely indicate lower reliability (Bonett and Wright, 2015; Flora, 2020; McNeish, 2017; Peters, 2014). When this assumption is violated, alpha is unable to differentiate between variations that are genuinely related to the underlying factor being measured from other sources of systematic variation such as the existence of multiple dimensions (Flora, 2020). However, McNeish (2017) cautions that low alphas could stem from assumption violation or indeed reflect the unreliability of the factors. The latter is illustrated by the weak item-total correlations and inadequate model fit of

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Table 5 Examination of the Volunteer Functions Inventory and Theory of Planned Behaviour questions: a recount of respondent reactions from field notes focusing on instances with more than two reactions (Items with weak item-total correlations ($r = 0.2$) are highlighted in red shade).

Factor	Item	item wording	Description of expressions and quotes
Understanding	U3	GO: - Reporting natural hazards in my community can help me learn how to deal with a variety of people ATRAP: - Reporting snail and environmental data weekly can help me learn how to deal with a variety of people	Most respondents agreed that being a CSs enabled them to build their interpersonal skills and communication abilities. *"You know that is true because now I interact with many people not just my neighbours and relatives, but also leaders, and even people from the university like you!"*
	U5	GO: - Reporting natural hazards in my community allows me to learn through direct "hands-on" experience ATRAP: - Reporting snail and environmental data weekly lets me learn through direct "hands-on" experience	The majority of the respondents were confused by the 'hands-on experience'. It was not immediately clear that it meant learning through direct involvement. *"Ehhh hand-on what?"*
Values	V2	GO: - I report natural hazards in my community because I am concerned about those who are less fortunate than myself ATRAP: - I report on snail and environmental data weekly because I am concerned about those less fortunate than myself	Almost all respondents questioned the meaning of 'those less fortunate than themselves'; often suggesting that they, themselves, were not doing any better than their community members. "*What do you mean by fortunate? Fortunate people live in Kampala! Who said that I am safe? Honestly, reporting hazards is more about safety for everyone*"
Career	C2	GO: - Reporting natural hazards in my community can help me get a foot in the door at a place where I'd like to work ATRAP: - Reporting snail and environmental data weekly can help me get a foot in the door at a place where I'd like to work	The expression 'Get a foot in the door at a place where I'd like to work' was confusing for some respondents who couldn't relate this directly to career factors. *"Do you mean that I can get a job because of volunteering with ATRAP?"*
	C3	GO: - Through reporting natural hazards in my community, I can make new contacts that could help my business ATRAP: - Through reporting snail and environmental data weekly, I can make new contacts that may help my business	Some CSs confirmed that being part of the CS networks has indeed opened doors for networking and visibility. *"Yes, collecting data has made me known to the community. Now people know my kabusiness and they come, they even send people to me."*
Social	S4	GO: - I report natural hazards because people I know share an interest in community services like reporting natural hazards in the community. ATRAP: - People I know share an interest in community service like reporting snail and environmental data weekly.	Questions S4 and 5 seemed vague as most respondents asked for the specific people implied by this text.
	S5	GO: - Reporting natural hazards in my community is an important activity for the people I know best ATRAP: - Reporting snail and environmental data weekly is an important activity for the people I know best	*"It depends, do you mean my friends, children or what? My friends don't care about snails!"*
Enhancement	E1	GO: - Reporting natural hazards in my community makes me feel needed	The question seemed incomplete for most respondents, and they asked:
		ATRAP: - Reporting snail and environmental data weekly makes me feel needed	*"Needed by who?"*
	E3	GO: - Reporting snail and environmental data weekly makes me feel important ATRAP: - Reporting natural hazards in my community makes me feel important	Some of the CSs confirmed this and gave examples of the benefits they have personally got. *"Working with people from the university is not easy gwe, I am part of a big team, and It makes me feel valuable in the community"*
Protective	P1	GO: - Reporting natural hazards in my community is a good escape from my own troubles ATRAP: - Reporting snail and environmental data weekly is a good escape from my own troubles	Most respondents asked for clarification after which they laughed. *"Reporting data can't make me forget my troubles. It is like an additional task even!"*
	P3	GO: - Reporting natural hazards in my community relieves me of some guilt of being more fortunate/blessed ATRAP: - Reporting snail and environmental data weekly work relieves me of some of the guilt over being more fortunate	The reactions here are similar to those in the values item V2 and most respondents were surprised, wondering how one can feel guilty because they are fortunate! They perceived being fortunate as prestigious and good. *"In fact, I am happy to be a citizen researcher because now the community knows me"*
Attitude		GO: - I think reporting natural hazards in my community whenever they occur in the next month is: ATRAP: - I think reporting snail and environmental data weekly is:	In general, some of the respondents laughed at most response options in the attitude factors as they thought that the tasks were not necessarily 'enjoyable' but rather contributed to community well-being.
	A3	[Harmful or beneficial]	*"Uhhhh we have, especially the community has not benefited from reporting hazards, that's why they don't call us anymore when hazards happen. They say that we just collect data and do nothing"*
	A6	[Unsatisfying or satisfying]	*"How can collecting snails be satisfying nawe?"*
Subjective norms	SN1	GO: - Those people most important to me think reporting natural hazards in my community each time they occur in the next month is [undesirable or desirable] ATRAP: - Those people most important to me think reporting snail and environmental data weekly is [undesirable or desirable]	Related to the question under the social factors, most respondents asked for the specific people implied in this question. *"Important, like who? My mother, wife…?"*
	SN2	GO: - How likely do you think it is that your fellow Geo-observers who are important to you, report natural hazards in my community each time they occur in the next month [unlikely/likely] ATRAP: - How likely do you think it is that your fellow citizen researchers who are important to you, report snail and environmental data weekly [unlikely/likely]	This brought mixed reactions as some CSs were aware of their peers who didn't fulfil their duties. During the group interviews, specific names were even brought up, often accompanied by laughter. *"Uhhh this one is tricky, but someone might be busy and not report the data and some people are just lazy hahaha!"*

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Table 5 (continued)

PBC	PBC1	GO: - How much control do you have over whether you can report environmental problems such as landslides or floods whenever such problems arise during the coming month: ATRAP: - How much control do you have over whether you can report snail and environmental data weekly:	Some respondents especially the CSs perceived the tasks as an obligation linked to the contracts that they freely had to sign at the start of the project to indicate their commitment to collecting and sharing the data, but also the possibility of termination at any time. *"You want to trick me? You know it is not a personal decision, I signed an MoU to collect data, so I have to do it"*
	PBC2b	GO: - Events or conditions outside my control will stop me from reporting natural hazards in my community each time they occur in the next month ATRAP: - Events or conditions outside my control will stop me from reporting snail and environmental data weekly	The majority of the respondents pin-pointed project-related conditions and not other factors at the community/individual level as hindering factors. *"Now like what unless the project stops now! But nothing can stop me. Even I signed a contract, so I have to do it!"*
	PBC3	GO: - If I wanted to, I can easily report natural hazards in my community each time they occur in the next month ATRAP: - If I wanted to, I can easily report snail and environmental data in my community weekly	For PBC3 and 4 questions, most respondents linked the decision to participate to the commitment stemming from the contracts signed. *"it's not just willingness, I also committed to doing so and this somehow also helps me to go when I feel like not going."*
	PBC4	GO: - It is mostly up to me whether I report snail and environmental data in my community weekly ATRAP: -It is mostly up to me whether I report natural hazards in my community each time they occur in the next month	
	PBC5b	GO: - It will be very difficult for me to report snail and environmental data in my community weekly ATRAP: - It will be very difficult for me to report natural hazards in my community each time they occur in the next month	Some respondents acknowledged the potential challenges that could make it difficult for them to perform the tasks assigned to them. *"Sometimes is difficult because we have to balance other responsibilities but also if the weather is not good then you can go out to collect data. Or sometimes we have travelled then that week you can't collect data."*
Intention	INT2	GO: - - I am determined to report natural hazards in my community each time they occur in the next month ATRAP: I am determined to report snail and environmental data in my community weekly	A few CSs indicated that this statement aligns with their commitments expressed through the signed contracts. *"If I wasn't, I would not have signed the contract again!"*
	NT3b	GO: - I do not intend to report natural hazards in my community each time they occur in the next month ATRAP: - I do not intend to report snail and environmental data in my community weekly	Among most respondents, this reverse-worded question raised suspicion and questions on the consequences if they responded in the affirmative! *"uhhh of course, I am fully prepared to collect data as long as the project is still here."*
Moral obligation	MO1	GO: - It would go against my principles not to report natural hazards in my community each time they occur in the next month ATRAP: - It would go against my principles not to report snail and environmental data in my community weekly	Some respondents didn't see the link between the tasks and their own principles but after explanation: they described their behaviour as volunteering spirit (mwoyo gwa gwanga)! *"It's mostly because of mwoyo gwa gwanga that I participate. We are not even paid so I do it for our community."*
	MO2	GO: - I will feel guilty if I do not report natural hazards in my community each time they occur in the next month ATRAP: - I will feel guilty if I do not report snail and environmental data in my community weekly	Most respondents thought that this question was 'funny' and out of place. *"Guilty? How? If I have a reason for not reporting, there is nothing I can do! But also, I report to the supervisor when I fail to go to the field."*
Self-identity	SI2b	GO: - I am not the type of person that can report natural hazards in my community each time they occur in the next month ATRAP: - I am not the type of person that can report snail and environmental data in my community weekly	This reverse-worded question needed to be read out more than three times in most cases for the majority of the respondents! *"ehh I see! Now I understand! But if I was not then I would NOT be here!"*
	SI1	GO: - I think of myself as someone who can report natural hazards in the community each time they occur in the next month ATRAP: - I think of myself as someone who can report snail and environmental data in the community weekly	Most of the respondents believed that they were competent enough to perform their tasks which is why they were trusted by the community leaders. *"If I couldn't do the task, the chairman would not have put my name!"*

nearly half of the CFA models. As such, the observed low alphas, particularly for some TPB factors, indicate that the relationship among such items is complex and cannot be represented by just one common factor or that the items are indeed not measuring the same latent concept (Hattie, 1985; Tavakol and Dennick, 2011; Vaske et al., 2017). Additionally, Streiner (2003) notes that a negative alpha value, as observed in the PBC factor of the GO_I, often indicates that the items measure different concepts. Furthermore, excluding items with weak correlations resulted in improved omega confirming Iacobucci and Duhachek's (2003) findings that alpha = 0 when items are not correlated ($r = 0$), regardless of the number of items. However, this is not ideal for factors with a low number of items, such as the problematic TPB factors (Iacobucci and Duhachek, 2003; Taber, 2018). Therefore, this opens up discussion on single-item measures by assessing the items using item-score reliability methods such as item–total correlation to identify the suitable item(s) to retain for a factor (Zijlmans et al., 2019). Third, despite improvements in skew after data transformation, we did not detect systematic improvements in the alpha estimates. This reflects findings by Norris and Aroian (2004) and confirms that data transformation does not always improve alpha estimates. This lack of improvement in alpha could be attributed to the inherent deviation from a normal distribution when responses cluster around high scores, resulting in many identical values that remain unaffected by data transformation (Childs et al., 2021). Clustered scores may arise due to the composition of the sample or characteristics of the respondents for instance, when participants share a similar background, thereby leading to uniform responses and a decrease in overall variability (Bademc, 2014; Bernardi, 1994; Deng and Chan, 2016; Dunn et al., 2014; Streiner, 2003). The consistency in scoring observed in this study could be explained by participant selection

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

bias, which possibly resulted in the nomination of community members who were already highly motivated. As such, the uniform patterns in scoring -high average scores and low alphas observed—across the CS networks, are an "artefact of the extremely homogenous sample" according to Bernardi (1994). Future research could therefore consider having a "true" control group consisting of a more diverse subset of the population, with experimental design setups to further assess the impact of the group composition on reliability (Sauermann et al., 2020).

When reliability was assessed using the less restrictive omega total and GLB indices, we observed: i) general increases in reliability estimates, and ii) a similar pattern of reliability where the VFI factors and 2 TPB factors demonstrated adequate reliabilities while the remaining TPB factors consistently exhibited low reliability across estimates. The lower alpha values compared to omega and GLB estimates are expected, given that the factors violate one or more assumptions of alpha (McNeish, 2017). Furthermore, slight differences observed between the omega and alpha estimates in this study, align with trends noted in other studies (Deng and Chan, 2016; Flora, 2020; Trizano-Hermosilla and Alvarado, 2016; Zhang and Yuan, 2016). However, this can be attributed to the violation of assumptions related to unidimensionality and uncorrelated errors by most factors, which are necessary for omega (Dunn et al., 2014; Flora, 2020; Tavakol and Dennick, 2011; Trizano-Hermosilla and Alvarado, 2016). The presence of uncorrelated errors is not trivial as these impact reliability estimates in an unpredictable way (McNeish, 2017; Trizano-Hermosilla and Alvarado, 2016). Additionally, Chakraborty (2017) notes that omega may underestimate reliability when the items have asymmetric or skewed distributions, as observed in this study. Conversely, GLB estimates generally outperformed both omega total and alpha, even when all alpha assumptions were violated, as reported by several other studies (Bendermacher, 2017; Chakraborty, 2017; Peters, 2014; Trizano-Hermosilla and Alvarado, 2016). Despite the robustness of GLB, it is sensitive to small sample sizes (i.e., less than 100) and weak item correlations (McNeish, 2017). Therefore caution is needed when interpreting the GLB estimates due to reported inflated estimates with small samples, although what constitutes a small sample size is debatable (McNeish, 2017; Trizano-Hermosilla and Alvarado, 2016).

Finally, the observation that most problematic items had weak item-total correlations (Table 5), highlights the relationship between reliability and item formulation or relevance to a particular context (Bademc, 2014; Boonyaratana et al., 2021; Deng and Chan, 2016). Moreover, the negative correlations observed for certain items, particularly PBC, signal issues in their construction according to Streiner (2003). This underscores the importance of verifying alpha assumptions and validating the utilized framework, ensuring that the framework accurately captures the intended concept and that its items are suitable within the specific study context (Boonyaratana et al., 2021; Chakraborty, 2017; Dunn et al., 2014; Flora, 2020; McNeish, 2017; Tavakol and Dennick, 2011; Trizano-Hermosilla and Alvarado, 2016; Zhang and Yuan, 2016; Zijlmans et al., 2019). Furthermore, the 'problematic' items observed in this study could be related to the nature and design of the CS projects. Specifically, participant motivations and behaviour may vary and be shaped by the expectations and responsibilities inherent in highly organized and formalized CS projects such as ATRAP, DSiRE and HARISSA (Lotfian et al., 2020). As previously noted, participants in this study were not self-selected and they also committed to the projects by signing official memorandums of understanding, which contrasts with typical CS projects that mostly rely on autonomous contributions from the general public (i.e., Lee et al., 2018). Additionally, the citizen scientists were facilitated with a smartphone for data collection and financial compensation to cover associated costs incurred (Brees et al., 2021; Jacobs et al., 2019; Sekajugo et al., 2022). Although the CSs voluntarily signed the contracts and could decide at any given moment to disengage, these factors may have influenced responses that align with socially desirable actions. Respondents might have been concerned about potential judgement or negative perceptions if they expressed deviations from their commitments (Milewski and Otto, 2017). As such, it is important to consider the nuanced dynamics of the project design when choosing a framework to apply.

Whilst the VFI and TPB have been extensively used, these frameworks were developed and predominantly applied in WEIRD (Western, Educated, Industrialized, Rich and Democratic) societies (Ajzen, 1991; Brayley et al., 2015; Clary et al., 1992; Jones, 2010). The limited utilization of these frameworks in diverse cultural and socioeconomic contexts, such as Uganda, prompts inquiries into the universality of these theoretical frameworks. The participant reactions revealed by the metadata highlight the influence of socioeconomic disparities and cultural mismatch, suggesting that some questions were designed with assumptions that may align more with the experiences and perspectives of some audiences and not others (e.g., 'less fortunate': stemming from volunteering literature where typically 'more fortunate' audiences are volunteering) (Asghar, 2015; Brayley et al., 2015; Güntert et al., 2016; Niebuur et al., 2019; Pocock et al., 2019; West et al., 2021; Wright et al., 2015; Zhang and Yuan, 2016). This mismatch therefore emphasizes the importance of considering cultural nuances in questionnaire design for more accurate and meaningful data collection (Niebuur et al., 2019). In addition, the psychometric properties of VFI and TPB could be evaluated by replicating the study across other CS projects in the Global South.

Limitations. Whilst the study provides useful insights into the challenges of estimating reliability for instruments developed in one field but applied in another new context, the study is not without limitations. Firstly, the primary purpose of the data collection was to understand the motivations and intentions of CS participants rather than to evaluate the psychometric properties of the frameworks. As such, other forms of reliability (e.g., test–retest) could not be assessed as the interviews were conducted several months apart. Secondly, the VFI and TPB items/questions were lengthy, and the Likert scales (accurate/important /agree) might have been complex and not intuitive. Related to this, the questionnaire was not translated into the local language, which could have influenced the interpretation and comprehension of the questions by the participants. Thirdly, besides having a small sample size, the control group characteristics were closely similar to the CSs, thereby resulting in minimal differences. Lastly, as in many studies utilizing interviews, the power dynamics between the interviewers and the CS participants could have led to socially desirable answers as the respondents anticipated staying in or joining the CS networks (Schwarz and Strack, 1999).

Recommendations. The VFI and TPB are important frameworks for assessing motives and behavioural decisions in various fields. However, the results of the present study show that caution is needed when applying these frameworks in contexts beyond their original application, particularly in fields like citizen science and across varied geographic regions. In moving forward, forthcoming research endeavours should consider verifying reliability assumptions, including the reporting of multiple reliability indices such as omega and GLB. Additionally, future studies could consider using a categorical-variable method for factor analysis since Likert scale scores are ideally ordered categorical scales (discrete integers) and not continuous data (Flora, 2020). Furthermore, in light of the participants' reactions and difficulties with some questions (see Table 5), it is recommended to thoroughly review and refine the wording of items through content or face validation of the proposed frameworks

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

before data collection (Boonyaratana et al., 2021). Specifically, qualitative research should precede and complement quantitative research, including pilot testing, to ensure the items are both reasonable measures of the underlying factor and relevant to the participants of that specific context (Ajzen, 2006). Lastly, it is crucial to acknowledge the diverse dynamics of CS projects when selecting theoretical frameworks. Recognizing that projects based on non-committal volunteering may exhibit distinct characteristics from formalized CS project designs where participants are part of a structured committed framework.

Conclusion

In this study, we investigate the internal consistency reliability of the VFI and TPB factors in a citizen science context in Uganda. We find that a priori calculations of Cronbach alpha tend to lead to low internal reliability. As such, Cronbach's alpha cannot blindly be applied to new applications as the underlying assumptions of estimating coefficient alpha might not be guaranteed. Also, we demonstrate deviations from alpha assumptions and outline our recommendations regarding the use of alternative reliability metrics. Finally, we stress the importance of investigating factors—even if well-defined and broadly applied—when applying them in new contexts, such as the field of citizen science motivational research in the Global South. This research contributes to the broader discussion on the reliability of frameworks used to assess participant motivations in CS projects and underscores the need for context-specific approaches to better understand participant motivations in this growing field.

Data availability

The datasets generated and analysed in the current study, along with a trial code demonstrating the data analysis procedure are freely available on the open science framework platform.

Received: 14 April 2023; Accepted: 22 February 2024;
Published online: 13 March 2024

References

Agnello G, Vercammen A, Knight AT (2022) Understanding citizen scientists' willingness to invest in, and advocate for, conservation. Biol Conserv 265:109422. https://doi.org/10.1016/j.biocon.2021.109422

Ajzen I (2006) Constructing a Theory of Planned Behavior Questionnaire: Conceptual and methodological considerations, vol 49(4). pp. 1–12

Ajzen Icek (1991) The theory of planned behavior. Organ Behav Hum Decis Process 50(2):179–211. https://doi.org/10.4135/9781446249215.n22

Alender B (2016) Understanding volunteer motivations to participate in citizen science projects: a deeper look at water quality monitoring. J Sci Commun 15(3):1–19. https://doi.org/10.22323/2.15030204

Asghar H (2015) The volunteer functions inventory: examination of dimension, scale reliability and correlates. Int J Innov Appl Res 3(4):52–64. http://www.journalijiar.com

Ashepet MG, Jacobs L, Van Oudheusden M, Huyse T (2021) Wicked solution for wicked problems: citizen science for vector-borne disease control in Africa. Trends Parasitol 37(2):93–96. https://doi.org/10.1016/j.pt.2020.10.004

Asingizwe D, Marijn Poortvliet P, Koenraadt CJM, van Vliet AJH, Ingabire CM, Mutesa L, Leeuwis C (2020) Why (not) participate in citizen science? Motivational factors and barriers to participate in a citizen science program for malaria control in Rwanda. PLoS ONE 15(8):e0237396. https://doi.org/10.1371/JOURNAL.PONE.0237396

Bademc V (2014) Cronbach's alpha is not a measure of unidimensionality or homogeneity. J Comput Educ Res 2(3):19–27

Bendermacher N (2010) Beyond alpha: lower bounds for the reliability of tests. J Mod Appl Stat Methods 9(1):95–102. https://doi.org/10.22237/jmasm/1272687000

Bendermacher N (2017) An unbiased estimator of the greatest lower bound. J Mod Appl Stat Methods 16(1):674–688. https://doi.org/10.22237/JMASM/1493598960

Bernardi A,R (1994) Validating research results when Cronbach's alpha is below. 70: a methodological procedure. Educ Psychol Meas 54(3):766–775

Beza E, Steinke J, Van Etten J, Reidsma P, Fadda C, Mittra S, Mathur P, Kooistra L (2017) What are the prospects for citizen science in agriculture? Evidence from three continents on motivation and mobile telephone use of resource-poor farmers. PLoS ONE 12(5):1–26. https://doi.org/10.1371/journal.pone.0175700

Bialocerkowski A, Bragge P (2008) Measurement error and reliability testing: application to rehabilitation. Int J Ther Rehabil 15(10):422–427

Bonett DG, Wright TA (2015) Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. J Organ Behav 36(1):3–15. https://doi.org/10.2307/26610966

Boonyaratana Y, Hansson EE, Granbom M, Schmidt SM, Kondo K, Ojima T (2021) The psychometric properties of the meaning of home and housing-related control beliefs scales among 67–70 year-olds in Sweden. https://doi.org/10.3390/ijerph18084273

Boslaugh S, Watters PA (2009) Statistics in a nutshell. In: Scientific computing, vol 26(1). Farnham O'Reilly

Brayley N, Obst PL, White KM, Lewis IM, Warburton J, Spencer NM (2015) Examining the predictive value of combining the theory of planned behaviour and the volunteer functions inventory. Aust J Psychol 67(3):149–156. https://doi.org/10.1111/AJPY.12078

Brees J, Huyse T, Tumusiime J, Kagoro-Rugunda G, Namirembe D, Mugabi F, Nyakato V, Anyolitho MK, Tolo CU, Jacobs L (2021) The potential of citizen-driven monitoring of freshwater snails in schistosomiasis research. Citiz Sci: Theory Pract 6(1):1–13. https://doi.org/10.5334/CSTP.388/METRICS/

Chacón F, Gutiérrez G, Sauto V, Vecina ML, Pérez A (2017) Volunteer Functions Inventory: a systematic review. Psicothema 29(3):306–316. https://doi.org/10.7334/psicothema2016.371

Chakraborty R (2017) Estimation of greatest lower bound reliability of academic delay of gratification scale. IOSR J Res Method Educ 7(2):75–79. https://doi.org/10.9790/7388-0702017579

Chen M (2020) The impacts of perceived moral obligation and sustainability self-identity on sustainability development: a theory of planned behavior purchase intention model of sustainability-labeled coffee and the moderating effect of climate change skepticism. Bus Strateg Environ 29(6):2404–2417. https://doi.org/10.1002/bse.2510

Childs D, Hindle B, Warren P (2021) APS 240: Data Analysis and Statistics with R. https://dzchilds.github.io/stats-for-bio/

Clary EG, Snyder M (1999) The motivations to volunteer: theoretical and practical considerations. Curr Dir Psychol Sci 8(5):156–159. https://doi.org/10.1111/1467-8721.00037

Clary EGil, Snyder M (1999) The motivations to volunteer: theoretical and practical considerations. Curr Dir Psychol Sci 8(5):156–159. https://doi.org/10.1111/1467-8721.00037

Clary EGiz, Snyder M, Ridge R (1992) Volunteers' motivations: retention of volunteers. Nonprofit Manag Leadersh 2(4):333–348

Deng L, Chan W (2016) Testing the difference between reliability coefficients alpha and omega. Educ Psychol Meas 77(2):185–203. https://doi.org/10.1177/0013164416658325

Dunn TJ, Baguley T, Brunsden V (2014) From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. Br J Psychol 105(3):399–412. https://doi.org/10.1111/bjop.12046

Finkelstein MA (2008) Volunteer satisfaction and volunteer action: a functional approach. Soc Behav Personal 36(1):9–18. https://doi.org/10.2224/sbp.2008.36.1.9

Fishbach A, Ouré-Tillery M (2018) Motives and goals. HKU PSYC2020. pp. 647–655. https://doi.org/10.1016/B978-0-12-809324-5.05637-6

Flora DB (2020) Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. Adv Methods Pract Psychol Sci 3(4):484–501. https://doi.org/10.1177/2515245920951747

Gerbing DW, Anderson JC (1984) On the meaning of within-factor correlated measurement errors. Source: J Consum Res 11(1):572–580. https://www.jstor.org/stable/2489144

Green SB, Yang Y (2009a) Commentary on coefficient alpha: a cautionary tale. Psychometrika 74(1):121–135. https://doi.org/10.1007/s11336-008-9098-4

Green SB, Yang Y (2009b) Reliability of summed item scores using structural equation modeling: an alternative to coefficient alpha. Psychometika 74(1):155–167. https://doi.org/10.1007/S11336-008-9099-3

Gu F, Little TD, Kingston NM (2013) Misestimation of reliability using coefficient alpha and structural equation modeling when assumptions of tau-equivalence and uncorrelated errors are violated. Methodology 9(1):30–40. https://doi.org/10.1027/1614-2241/A000052

Güntert ST, Strubel IT, Kals E, Wehner T (2016) The quality of volunteers motives: Integrating the functional approach and self-determination theory. J Soc Psychol 156(3):310–327. https://doi.org/10.1080/00224545.2015.1135864

Hagger MS, Chatzisarantis NLD (2006) Self-identity and the theory of planned behaviour: between-and within-participants analyses. Br J Soc Psychol 45(4):731–757. https://doi.org/10.1348/014466605X85654

Haklay M (2013) Citizen Science and Volunteered Geographic Information—overview and typology of participation. In: Sui DZ, Elwood S, Goodchild MF

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

(eds) Crowdsourcing Crowdsourcing geographic knowledge: volunteered geographic information (VGI) in theory and practice (Issue Elwood 2008). Springer, pp. 105–122. https://doi.org/10.1007/978-94-007-4587-2
Hattie J (1985) Methodology review: assessing unidimensionality of tests and ltems. Appl Psychol Meas 9(2). http://www.copyright.com/
Henson RK (2001) Understanding internal consistency reliability estimates: a conceptual primer on coefficient alpha. Meas Eval Couns Dev 34(3):177–189. https://doi.org/10.1080/07481756.2002.12069034
Hulbert J (2016) Citizen science tools available for ecological research in South Africa. South Afr J Sci 112(5–6):1–2
Iacobucci D, Duhachek A (2003) Advancing alpha: measuring reliability with confidence. J Consum Psychol 13(4):478–487. https://doi.org/10.1207/S15327663JCP1304_14
Jacobs L, Kabaseke C, Bwambale B, Katutu R, Dewitte O, Mertens K, Maes J, Kervyn M (2019) The geo-observer network: a proof of concept on participatory sensing of disasters in a remote setting. Sci Total Environ 670:245–261. https://doi.org/10.1016/j.scitotenv.2019.03.177
Jeanmougin M, Levontin L, Schwartz A (2017) Motivations for participation to citizen-science program: a meta-analysis purpose of the STSM. https://cs-eu.net/sites/default/files/media/2017/07/Jeanmougin-etal-2017-STSMReport-%0AMotivationParticipation.pdf
Jones D (2010) A weird view of human nature skews psychologists' studies. Science 328(5986):1627. https://doi.org/10.1126/SCIENCE.328.5986.1627
Jorgensen TD, Pornprasertmanit S, Schoemann AM, Rosseel Y (2022) semTools: useful tools for structural equation modeling. R Package Version 0.5–6. https://cran.r-project.org/web/packages/semTools/citation.html
Kanyiginya V, Twongyirwe R, Kagoro-Rugunda G, Mubiru D, Sekajugo J, Mutyebere R, Deijns AAJ, Kervyn M, Dewitte O (2023) Inventories of natural hazards in under-reported regions: a multi-method insight from a tropical mountainous landscape. Afr Geogr Rev 1–20. https://doi.org/10.1080/19376812.2023.2280589
Land-Zandstra A, Agnello G, Gültekin YS (2021) Participants in citizen science. in *The science of citizen science*. (eds Vohland, K., Anne Land-Zandstra, Luigi Ceccaroni, Rob Lemmens, Josep Perelló, Marisa Ponti, RoelandSamson, and Katherin Wagenknecht) Springer International Publishing
Leach CW, van Zomeren M, Zebel S, Vliek MLW, Pennekamp SF, Doosje B, Ouwerkerk JW, Spears R (2008) Group-level self-definition and self-investment: a hierarchical (multicomponent) model of in-group identification. J Personal Soc Psychol 95(1):144–165. https://doi.org/10.1037/0022-3514.95.1.144
Lee TK, Crowston K, Harandi M, Østerlund C, Miller G (2018) Appealing to different motivations in a message to recruit citizen scientists: results of a field experiment. JCOM, J Sci Commun 17(01), A02. https://doi.org/10.22323/2.17010202
Lotfian M, Ingensand J, Brovelli MA (2020) A framework for classifying participant motivation that considers the typology of citizen science projects. ISPRS Int J Geo-Inf 9(12). https://doi.org/10.3390/ijgi9120704
Maund PR, Irvine KN, Lawson B, Steadman J, Risely K, Cunningham AA, Davies ZG (2020) What motivates the masses: understanding why people contribute to conservation citizen science projects. Biol Conserv 246:108587. https://doi.org/10.1016/J.BIOCON.2020.108587
McNeish D (2017) Thanks coefficient alpha, we'll take it from here. Psychol Methods 23(3):412–433. https://doi.org/10.1037/met0000144
McNeish D (2020) Should we use *F*-tests for model fit instead of chi-square in overidentified structural equation models? Organ Res Methods 23(3):487–510. https://doi.org/10.1177/1094428118809495/FORMAT/EPUB
Milewski N, Otto D (2017) The presence of a third person in face-to-face interviews with immigrant descendants: patterns, determinants, and effects. Life Course Res Soc Polic 7:77–96. https://doi.org/10.1007/978-94-024-1141-6_4/TABLES/4
Mukherjee N, Zabala A, Huge J, Nyumba TO, Adem Esmail B, Sutherland WJ (2018) Comparison of techniques for eliciting views and judgements in decision-making. Methods Ecol Evol 9(1):54–63. https://doi.org/10.1111/2041-210X.12940
Muzaffar B (2016) The development and validation of a scale to measure training culture: the TC scale. J Cult Soc Dev 23:49–58. https://iiste.org/Journals/index.php/JCSD/article/download/33004/33900
Niebuur J, Liefbroer AC, Steverink N, Smidt N (2019) Translation and validation of the Volunteer Functions Inventory (VFI) among the general Dutch older population. https://doi.org/10.3390/ijerph16173106
Nimon K, Zientek LR, Henson RK (2012) The assumption of a reliable instrument and other pitfalls to avoid when considering the reliability of data Front Psychol 3:102. https://doi.org/10.3389/FPSYG.2012.00102/BIBTEX
Norris AE, Aroian KJ (2004) To transform or not transform skewed data for psychometric analysis that is the question! Nurs Res 53(1):67–71
Peters G-JY (2014) The alpha and the omega of scale reliability and validity. Eur Health Psychol 16(2):56–69
Pocock MJO, Roy HE, August T, Kuria A, Barasa F, Bett J, Githiru M, Kairo J, Kimani J, Kinuthia W, Kissui B, Madindou I, Mbogo K, Mirembe J, Mugo P, Muniale FM, Njoroge P, Njuguna EG, Olendo MI, Trevelyan R (2019) Developing the global potential of citizen science: assessing opportunities that benefit people, society and the environment in East Africa. J Appl Ecol 56(2):274–281. https://doi.org/10.1111/1365-2664.13279
Revelle W (2015) Package "psych"—procedures for psychological, psychometric and personality research. R Package. pp. 1–358. https://cran.r-project.org/web/packages/psych/citation.html
Revelle W, Condon DM (2019) Reliability from α to ω: a tutorial. Psychol Assess 31(12):1395–1411. https://doi.org/10.1037/pas0000754
Rosseel Y (2012). Lavaan: an R package for structural equation modeling. J Stat Softw 48(2). https://doi.org/10.18637/JSS.V048.I02
Saad S, Carter GW, Rothenberg M, Israelson E (1999) Testing and assessment: an employer's guide to good practices. For full text: http://wdr.doleta.gov/opr/FULLTEXT/99-testassess.pdf. http://wdr.doleta.gov/opr/FULLTEXT/99-
Sauermann H, Vohland K, Antoniou V, Balázs B, Göbel C, Karatzas K, Mooney P, Perelló J, Ponti M, Samson R, Winter S (2020) Citizen science and sustainability transitions. Res Policy 49(5):103978. https://doi.org/10.1016/J.RESPOL.2020.103978
Schwarz N, Strack F (1999) Reports of subjective well-being: Judgmental processes and their methodological implications. In Kahneman, D., Diener, E. & Schwarz N. (eds), Well-being: The foundations of hedonic psychology (pp. 61–84). Russell Sage Foundation
Sekajugo J, Kagoro-Rugunda G, Mutyebere R, Kabaseke C, Namara E, Dewitte O, Kervyn M, Jacobs L (2022) Can citizen scientists provide a reliable geo-hydrological hazard inventory? An analysis of biases, sensitivity and precision for the Rwenzori Mountains, Uganda. Environ Res Lett 17(4). https://doi.org/10.1088/1748-9326/ac5bb5
Sheng Y, Sheng Z (2012) Is coefficient alpha robust to non-normal data? Front Psychol 3. https://doi.org/10.3389/FPSYG.2012.00034/ABSTRACT
Sijtsma K (2009) On the use, the misuse, and the very limited usefulness of Cronbach's alpha. Psychometrika 74(1):107–120. https://doi.org/10.1007/s11336-008-9101-0
Streiner DL (2003) Starting at the beginning: an introduction to coefficient alpha and internal consistency. J Personal Assess 80(1):99–103. https://doi.org/10.1207/S15327752JPA8001_18
Tabachnick BG, Fidell LS, Ullman JB (2013) Review of using multivariate statistics (Pearson M (ed)) vol. 6(8). https://doi.org/10.1037/022267
Taber S,K (2018) The use of Cronbach's alpha when developing and reporting research instruments in science education. Res Sci Educ 48(6):1273–1296
Tang W, Cui Y, Babenko O (2014) Internal consistency: do we really know what it is and how to assess it? J Psychol Behav Sci 2(2):205–220
Tavakol M, Dennick R (2011) Making sense of Cronbach's alpha. Int J Med Educ 2:53–55. https://doi.org/10.5116/ijme.4dfb.8dfd
Trizano-Hermosilla I, Alvarado JM (2016) Best alternatives to Cronbach's alpha reliability in realistic conditions: congeneric and asymmetrical measurements Front Psychol 7:1–8. https://doi.org/10.3389/fpsyg.2016.00769
Vaske JJ, Beaman J, Sponarski CC (2017) Rethinking internal consistency in Cronbach's alpha. Leis Sci 39(2):163–173. https://doi.org/10.1080/01490400.2015.1127189
Werner C, Schermelleh-Engel K (2010) Deciding between competing models: chi-square difference tests. In: Introduction to structural equation modeling with LISREL1-3. pp. 1–3. https://www.researchgate.net/publication/241278052_Deciding_Between_Competing_Models_Chi-Square_Difference_Tests
West S (2017) How could citizen science support the Sustainable Development Goals? https://www.sei-international.org/mediamanager/documents/Publications/SEI-2017-PB-citizen-science-sdgs.pdf
West S, Dyke A, Pateman R (2021) Variations in the motivations of environmental citizen scientists. Citiz Sci: Theory Pract 6(1):1–18. https://doi.org/10.5334/CSTP.370
Wright DR, Underhill LG, Keene M, Knight AT (2015) Understanding the motivations and satisfactions of volunteers to improve the effectiveness of citizen science programs. Soc Nat Resour 28(9):1013–1029. https://doi.org/10.1080/08941920.2015.1054976
Xin Z, Liang M, Zhanyou W, Hua X (2019) Psychosocial factors influencing shared bicycle travel choices among Chinese: an application of theory planned behavior. PLoS ONE 14(1):1–17. https://doi.org/10.1371/journal.pone.0210964
Young JC, Rose DC, Mumby HS, Benitez-Capistros F, Derrick CJ, Finch T, Garcia C, Home C, Marwaha E, Morgans C, Parkinson S, Shah J, Wilson KA, Mukherjee N (2018) A methodological guide to using and reporting on interviews in conservation science research. Methods Ecol Evol 9(1):10–19. https://doi.org/10.1111/2041-210X.12828
Zhang Z, Yuan KH (2016) Robust coefficients alpha and omega and confidence intervals with outlying observations and missing data: methods and software. Educ Psychol Meas 76(3):387–411. https://doi.org/10.1177/0013164415594658
Ziegler M, Hagemann D (2015) Testing the unidimensionality of Items: pitfalls and loopholes. Eur J Psychol Assess 31(4):231–237. https://doi.org/10.1027/1015-5759/A000309

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Zijlmans EAO, Tijmstra J, van der Ark LA, Sijtsma K (2019) Item–score reliability as a selection tool in test construction Front Psychol 9:2298. https://doi.org/10.3389/FPSYG.2018.02298/BIBTEX

Zumbo BD (1999) A glance at coefficient alpha with an eye towards robustness studies: some mathematical notes and a simulation model. Edgeworth Ser Quant Behav Sci 99(1):1–13

Acknowledgements

We wish to acknowledge the support received from the ATRAP and HARISSA projects of the Development Cooperation programme of the Royal Museum for Central Africa which funded the research with support from the Belgian Directorate-General Development Cooperation and Humanitarian Aid. Also, we would like to express our gratitude to all the study respondents who availed their time to participate in the interviews and for their continued collaboration. Last but not least, we are grateful to the project team members; Matthieu Kervyn De Meerendre, Casim Umba Tolo, Viola Nyakato, Kewan Mertens, David Mubiru, Esther Namara, John Sekajugo, Julius Tumusiime, Maxson Anyolitho, Daisy Namirembe, Faith Mugabi and the village chiefs who played key roles in the setup of the CS networks and the data collection phase.

Author contributions

Research idea and study design: AMG, LV, TH, CM and LJ. Data collection: AMG, VK, RM, and OD. Data analysis and interpretation: AMG, LV and LJ. Manuscript draft: AMG and LJ. Administrative support: LV, TH, CM, GKR, CK, RT, and OD. Supervision: LV, TH, CM and LJ.

Competing interests

The authors declare no competing interests.

Ethical approval

The study was approved by the research ethics committees of the partnering institutions: Mbarara University of Science and Technology (reference number MUST-2021-56) and KU Leuven (reference number G-2019 11 1842). Additionally, the study was performed following the ethical guidelines on research with human participants in accordance with the Declaration of Helsinki.

Informed consent

All respondents voluntarily participated after the research aims and objectives were disclosed to them and written informed consent was obtained before the start of the interviews. The participants were also informed that their responses were to be used solely for academic purposes and anonymity during data analysis was guaranteed. Lastly, the participants received compensation for transport costs incurred.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-024-02873-1.

Correspondence and requests for materials should be addressed to Mercy Gloria Ashepet.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

© The Author(s) 2024

[1]Department of Biology, Royal Museum for Central Africa, Tervuren, Belgium. [2]Division of Bioeconomics, Department of Earth and Environmental Sciences, KU Leuven, Belgium. [3]Department of Earth Sciences, Royal Museum for Central Africa, Tervuren, Belgium. [4]Department of Agricultural and Resource Economics, Mountains of the Moon University, Fort Portal, Uganda. [5]School of Agriculture and Environmental Sciences, Mountains of the Moon University, Fort Portal, Uganda. [6]Department of Environment and Livelihood Support Systems, Faculty of Interdisciplinary Studies, Mbarara University of Science and Technology, Mbarara, Uganda. [7]Department of Geography, Vrije Universiteit Brussels, Brussels, Belgium. [8]Department of Biology, Mbarara University of Science and Technology, Mbarara, Uganda. [9]Ecosystem & Landscape Dynamics, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands. [10]Division of Geography and Tourism, Department of Earth and Environmental Sciences, KU Leuven, Belgium. ✉email: mercy.ashepet@africamuseum.be

Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH ("Springer Nature").
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users ("Users"), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use ("Terms"). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com