



Article

Data Reliability in a Citizen Science Protocol for Monitoring Stingless Bees Flight Activity

Jailson N. Leocadio ^{1,*}, Natalia P. Ghilardi-Lopes ², Sheina Koffler ³, Celso Barbiéri ⁴, Tiago M. Franco ⁴, Bruno Albertini ¹ and Antonio M. Saraiva ^{1,3}

- ¹ Escola Politécnica, University of São Paulo, Av. Prof. Luciano Gualberto 158, Tv. 3, São Paulo 05508-010, SP, Brazil; balbertini@usp.br (B.A.); saraiva@usp.br (A.M.S.)
² Centro de Ciências Naturais e Humanas, Federal University of ABC, R. Arcturus 3, São Bernardo do Campo 09606-070, SP, Brazil; natalia.lopes@ufabc.edu.br
³ Instituto de Estudos Avançados, University of São Paulo, R. Praça do Relógio 109, São Paulo 05508-970, SP, Brazil; sheina.koffler@usp.br
⁴ Escola de Artes, Ciências e Humanidades, University of São Paulo, R. Arlindo Bettio 1000, São Paulo 03828-000, SP, Brazil; celso.barbieri@usp.br (C.B.); tfrancoy@usp.br (T.M.F.)
* Correspondence: jailsonleocadio@usp.br

Simple Summary: This work aims to validate a citizen science protocol for monitoring the flight activity of stingless bees. The count of flight activity (entrance, exit, and entrance carrying pollen) filmed in 30 s videos was compared among three different groups: “original” citizen scientists (group that filmed and performed the count in their own videos), “replicator” citizen scientists (group of citizen scientists who performed flight activity counts on videos shot by other citizen scientists), and experts (researchers who work with bees and who performed the counts on videos shot by citizen scientists). The analysis was divided into two levels: perception (detection of activity in videos) and counting. The results of this analysis revealed that citizen scientists and experts have similar perception and count of bee entrance and exit activity, as no statistical differences were found in these two items. However, replicator citizen scientists noticed more bees carrying pollen than original citizen scientists and experts. Despite this, considering only the videos in which the groups agreed on the presence of pollen, the count was similar for both. These results enabled the validation of the protocol and indicated high quality of data produced by individuals who participate in scientific practices following a citizen science approach.



Citation: Leocadio, J.N.; Ghilardi-Lopes, N.P.; Koffler, S.; Barbiéri, C.; Franco, T.M.; Albertini, B.; Saraiva, A.M. Data Reliability in a Citizen Science Protocol for Monitoring Stingless Bees Flight Activity. *Insects* **2021**, *12*, 766. <https://doi.org/10.3390/insects12090766>

Academic Editors: Alessandro Campanaro and Cristina Castracani

Received: 1 July 2021

Accepted: 20 August 2021

Published: 27 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Although the quality of citizen science (CS) data is often a concern, evidence for high-quality CS data increases in the scientific literature. This study aimed to assess the data reliability of a structured CS protocol for monitoring stingless bees' flight activity. We tested (1) data accuracy for replication among volunteers and for expert validation and (2) precision, comparing dispersion between citizen scientists and expert data. Two distinct activity dimensions were considered: (a) perception of flight activity and (b) flight activity counts (entrances, exits, and pollen load). No significant differences were found among groups regarding entrances and exits. However, replicator citizen scientists presented a higher chance of perceiving pollen than original data collectors and experts, likely a false positive. For those videos in which there was an agreement about pollen presence, the effective pollen counts were similar (with higher dispersion for citizen scientists), indicating the reliability of CS-collected data. The quality of the videos, a potential source of variance, did not influence the results. Increasing practical training could be an alternative to improve pollen data quality. Our study shows that CS provides reliable data for monitoring bee activity and highlights the relevance of a multi-dimensional approach for assessing CS data quality.

Keywords: biodiversity monitoring; data quality; meliponini; protocol validation; volunteer participation

1. Introduction

Citizen science (CS) is a scientific approach that allows members of the general public to contribute to the scientific process, usually as data collectors but desirably in other scientific inquiry steps [1,2]. As data contributors, the participants can provide large amounts of data that, otherwise, would require great time availability, as well as substantial financial resources [3,4]. This scattered and diverse information can reduce data scarcity problems, which are common in some study fields (e.g., species populations distribution, water quality monitoring). Albeit still underutilized, CS can be a valuable resource for global change research and UN Sustainable Development Goals achievement [3,5].

Data quality (DQ), which can be considered a multi-dimensional issue [6], is a concern for several researchers involved with CS activities [7,8]). Studies report on different dimensions of data quality, such as standardized sampling (e.g., [9]), spatial and temporal representativeness and bias (e.g., [10]), data accuracy and precision [8,11,12], sample size (e.g., [13–16]), volunteers' proper training (e.g., [17–19]), and the experience and ability levels of participants [20,21].

In this sense, the work developed by Wiggins and Crowston [2] analyzed the mechanisms for data quality assurance in 128 CS projects and concluded that the topic is a concern in most of them. The most common procedures to ensure data quality observed were: expert review, photo submissions, paper data sheets submitted along with online entry, replication or rating by multiple participants, quality assurance/quality control (QA/QC), training programs, and automatic filtering of unusual reports. Pilot-testing of citizen science protocols is also a strategy commonly used to improve data quality and reliability. The feedback from testing participants is essential to redesign the protocols and build appropriate materials for the project [22].

Strategies for DQ control may be applied during data collection, data classification, or data analyses through statistical and modeling tools [8]. Data quality control during classification, for instance, may include replication among volunteers (distinct volunteers performing the same task and reaching some consensus—cross-comparisons) and expert validation (comparing citizen scientists' data with professional scientists' data) [23]. However, few studies have systematically tested these control mechanisms (e.g., [16,24–26]). Aceves-Bueno et al. [27], for instance, analyzed 63 citizen science papers that reported 1363 observations of expert validation and found that 73% of the abstracts described the contributions of citizen science as positive (accurate, reliable, comparable, statistically similar, or valuable) and only 13% assessed citizen scientists' (cs) performance negatively (no significant correlations, overestimated, or contradictions). In addition, validated CS data was reported to be more cost-effective than traditional methods [15].

In a review of citizen science initiatives with bees, Koffler et al. [28] reported the use of various strategies related to data quality assessment and control, mainly digital vouchers (photographs submitted by citizen scientists) (43.2%), expert review of data (40.9%), use of structured protocols (40.9%), and training of participants (29.6%). The same initiative used up to five different strategies, indicating that data reliability was a major concern for the projects' teams. For instance, data quality was stated as the primary objective by 13.6% of the 88 studies analyzed, with protocols mainly related to sampling effort and species identification. While bumblebees and honey bees were the most investigated groups, only three works studied stingless bees, despite the increasing interest in this group due to their importance as pollinators [29] and the global expansion of meliponiculture activities [30,31]. Stingless bees comprise a diverse group with more than 500 recognized species in tropical and subtropical regions [32] and stingless beekeeping may be an important tool for sustainable rural development and conservation [33]. However, the lack of basic knowledge of stingless bee ecology and management still hampers the practice [31,34]. In this context, beekeepers may act as important partners in CS projects with stingless bees, following successful ongoing initiatives with honey bees [35]. Monitoring stingless bees' flight activity, for instance, can help us understand several factors that affect colony performance, such as responses of foraging bees to intra-colony stimuli and meteorological

conditions. Flight activity data also serve as an economic evaluation of the colony since the number of foraging trips is directly linked to colony production and pollination services [36]. Therefore, good quality data production is essential to subsidize management strategies.

The present study aimed to assess the data reliability of a structured citizen science protocol for monitoring stingless bee flight activity. Our initial hypothesis is that there are no statistical differences between the data produced by citizen scientists in comparison to data produced by experts, although the dispersion of the data produced by citizen scientists may be greater. Participants of an outreach course related to citizen science and meliponiculture produced the data. Original data gathered by citizen scientists was first replicated by a group of citizen scientists, who also participated in the project, and then validated by a group of experts. We tested data accuracy for replication (comparing original and replicated data) and for expert validation (comparing citizen scientists and expert data). Moreover, precision was analyzed for data validation, comparing dispersion between replicator citizen scientists and expert data. Since flight monitoring is a task with considerable difficulties for untrained personnel, two distinct activity levels were considered: (a) perception of flight activity (whether the activity was detected or not) and (b) flight activity (bee counts when activity was detected). Our approach provides a multi-dimensional assessment (accuracy and precision in perception and counts) of reliability in citizen science data for a non-model insect organism.

2. Material and Methods

During an outreach course on meliponiculture and citizen science held in July 2020, participants were invited and trained to perform and pilot-test a structured protocol aimed at the monitoring of flight activity for *Tetragonisca angustula* (Latreille, 1811), a stingless bee widely distributed in Brazil and commonly reared by beekeepers across the country. The protocol stated that citizen scientists had to film the nest's entrance for 30 s within different time intervals (between 7:00 a.m. and 9:00 a.m.; 11:00 a.m. and 1:00 p.m.; 3:00 p.m. and 5:00 p.m.). Data collection consisted of watching the videos and counting how many bees entered (entrance), how many left the nest (exit), and how many came in carrying pollen (pollen) in that period (Figure 1A). As stingless bees carry pollen attached to the pollen baskets in the hind legs (Figure 1B), pollen loads were visible and could be identified in video recordings. Data submission was carried on a web system developed exclusively for this purpose (<https://beekeep.pcs.usp.br>, in Portuguese, University of São Paulo, São Paulo, Brazil—access date: 23 August 2021), also collecting other relevant variables for further studies.

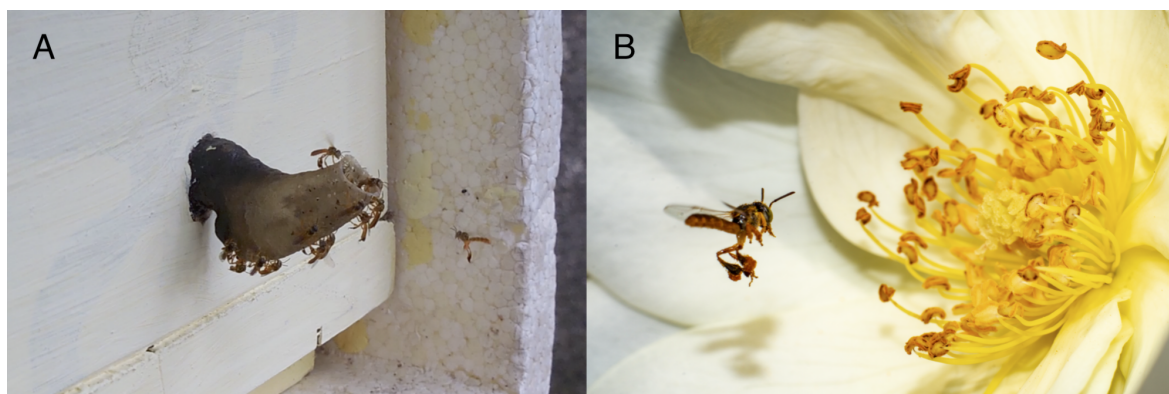


Figure 1. (A) A frame of one of the received videos showing a bee approaching the nest entrance tube and some guard bees. (B) An image of a bee carrying pollen attached to its hind legs. Photo by André Matos.

Of more than 400 submitted videos, 42 were randomly selected for this study purposes, along with the counts (Supplementary Material 1) provided by citizen scientists at the time of video submission (these participants are hereafter called “cs original”). The videos were divided into seven groups of six videos each. For the replication analyses among citizen

scientists, each group of videos was assessed by at least 11 participants from a total set of 101 citizen scientists (from now on called “cs replicators”), none of them included in the cs original group. Group size varied from 11 to 19, none of them analyzed videos of more than one group. For the validation process, a set of five experts in stingless bee behavior analyzed all videos, including three authors of this study. Thus, at the end of the quality control process, there were 2574 countings (considering 858 countings both on entrances, exits, and pollen), 126 of which were from the cs original (42 individuals × 1 video per individual × 3 countings per video), 1818 from the cs replicators (101 individuals × 6 videos per individual × 3 countings per video), and 630 from experts (5 individuals × 42 videos per individual × 3 countings per video) (Figure 2).

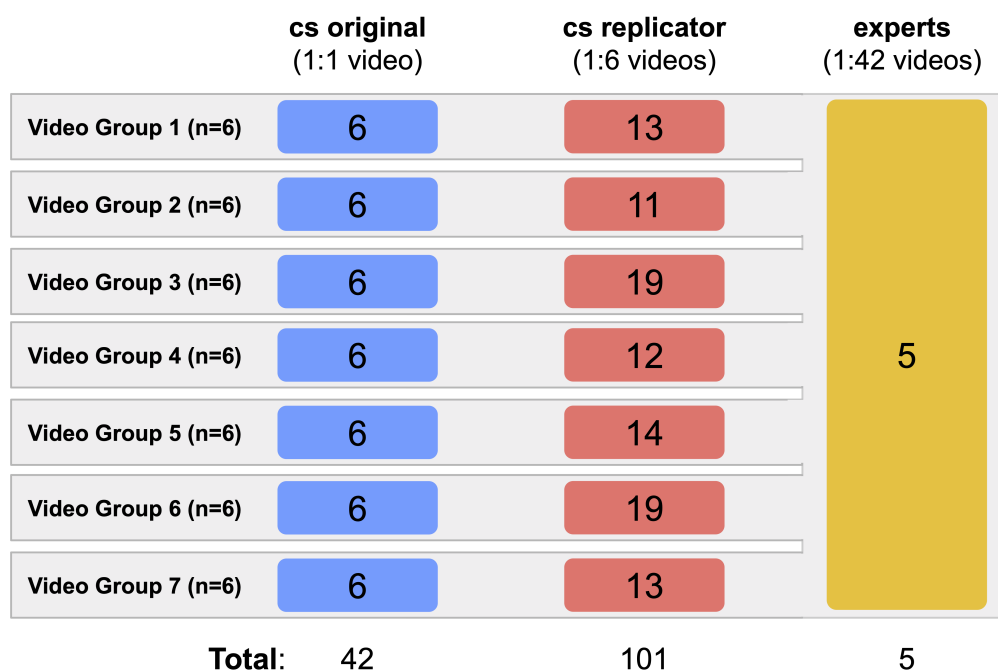


Figure 2. Distribution of the 42 videos for each group: cs original ($n = 42$ individuals—1:1 video), cs replicator ($n = 101$ individuals—1:6 videos) and experts ($n = 5$ individuals—1:42 videos).

Data Analysis

The groups (cs original, cs replicators, and experts) and video quality were considered potential sources of observed differences. In order to assess video quality, four variables were assembled through a Principal Component Analysis (PCA): Mean Structural SIMilarity (MSSIM), Focus, Contrast, and Frames per Second (FPS). MSSIM was obtained using a custom script that relies on OpenCV 4.5.1 implementation of MSSIM as described by Wang et al. [37]. We also considered other traditional metrics such as MSE (Mean Square Error) and the correlated PSNR (Peak Signal-to-Noise Ratio), but MSSIM is a better metric when considering the human perception of the video quality [38]. Focus and contrast indicate qualitatively if the nest was in the foreground and if it was possible to differentiate the bees from the background, respectively. These two metrics were determined by a designer specialist, who watched the videos and checked both attributes in each. FPS indicated the number of frames per second and was extracted from video metadata or inferred using the file size and the video bitrate (uncompressed) when the metadata was missing or wrong (e.g., recorder used variable bitrate). Before performing the PCA, FPS data were scaled by subtracting each value from the mean and dividing it by the standard deviation. PCA Axis 1 explained 68% of the data variability and was used as a proxy for video quality in our analyses.

Generalized Linear Mixed-Effects Models (GLMER) were adjusted to analyze accuracy for replication and validation processes. A Boolean variable was created to indicate the presence (when greater than zero) or absence of activity and to assess the probability

of flight activity perception. The perception of flight activity was modeled as a binary response (presence—when the participant responded with a value greater than zero for the specific activity—or absence of activity), following a Bernoulli distribution. At the same time, the effective counts of activity were modeled using a Poisson distribution (using only data for which counts were greater than zero). The group was included in the model as a fixed effect (cs original compared to cs replicators for replication analyses; cs replicators compared to experts for validation analyses). The video quality was included in the models as a covariate. A random effect was set for the videos to account for dependencies in the data, with participant identity nested within each video. Initially, complete models with response variables for each bee activity were set (entrances, exits, and pollen). Then, reduced models were adjusted by removing the fixed effects, one by one, until the null model. Likelihood-ratio tests were employed to compare models and to select the best model in each analysis. Overdispersion was assessed by verifying the data's standard deviation (sd) against the sd of simulated data. The estimated coefficients were back-log-transformed using the exponential function to obtain the values of odds or odds ratio from the models.

The Median Absolute Deviation (MAD) was used as an indicator of counting dispersion for both the group of cs replicators and the group of experts. All the counts were considered in this analysis, even when no activity was found (zero values). Data from cs original could not be included in this analysis, as no replicates were performed, and hence no variation could be measured. Like other studies that analyzed count-based protocols, we chose the median-based metric because data were non-parametric, and the median minimizes the influence of some extreme countings [16,39]. The mean of the MAD of each video was calculated for each group and compared through a paired Wilcoxon signed-rank test to verify possible statistically significant differences between the groups. This non-parametric test was used because the paired differences in MAD values between groups were not approximately normally distributed. We also analyzed the correlation between the mean of MAD and the median of the activities counts to assess whether there was a relationship between bee activity rates and variation in counts in our protocol, using the Pearson correlation coefficient.

All analyses were performed in R (version 4.0.4) [40], employing the functions and respective packages: `scale` (base), `prcomp` (stats), `GLMER` (lme4), `anova` (stats), `mad` (stats), `wilcox.test` (stats), `cor.test` (stats) and `testDispersion` (DHARMA).

3. Results

3.1. Perception and Effective Countings

The perception of entrance activity from the videos ranged from 69% (cs original and experts) to 70% (cs replicators), while the perception of exit activity ranged from 76% (experts) to 79% (cs replicators). Perception of pollen was much lower than the previous activities and varied from 14% (cs original) to 34% (cs replicators) (Figure 3). For those videos with perceived activity, the median for entrance counts was 5 for all the groups; medians for exit counts were 6, 5, and 6, while medians for pollen counts were 1, 2, and 1 (for cs original, cs replicators, and experts, respectively) (Figure 4).

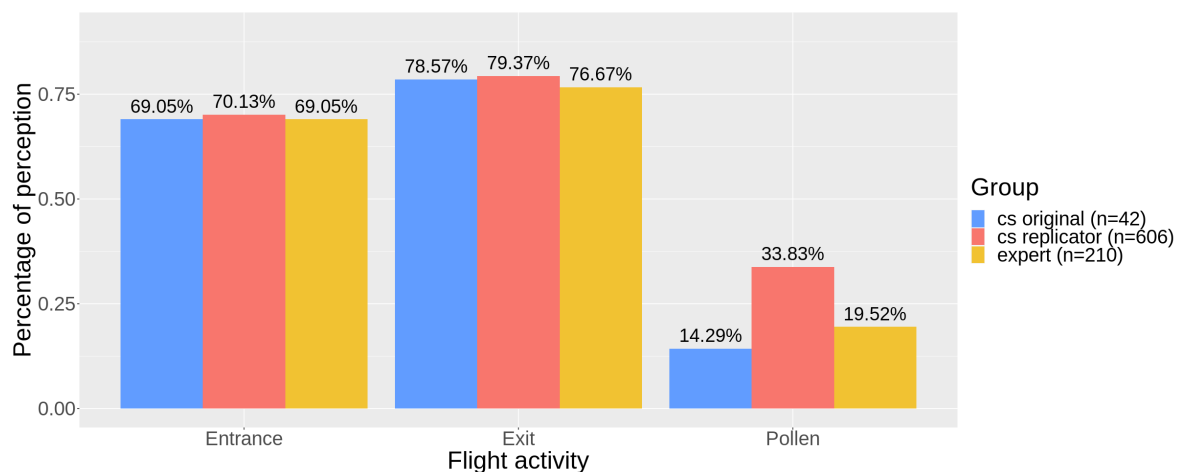


Figure 3. Absolute frequency of perception of activity (entrance, exit, pollen) by the groups (cs original—in blue —, cs replicator—in red—, expert—in yellow) (see also Figure 2).

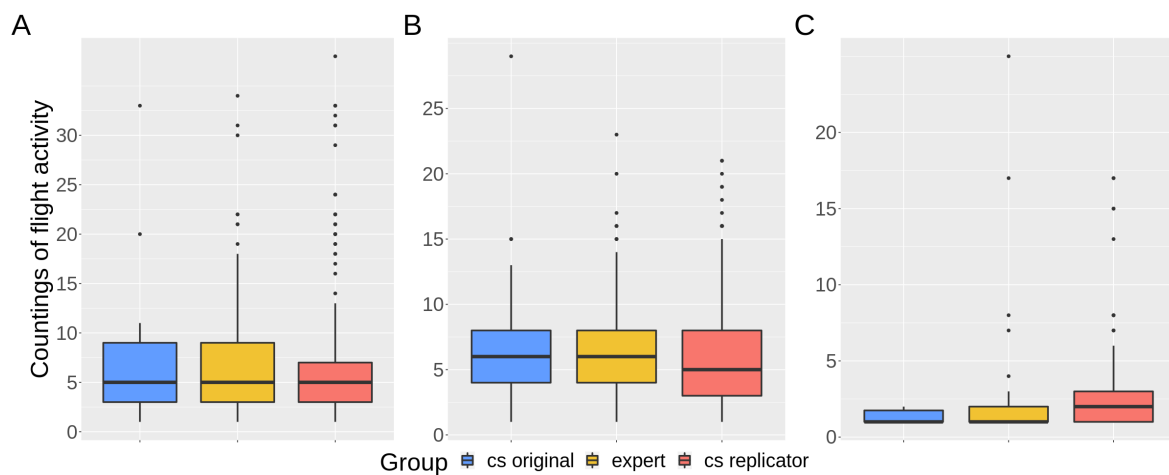


Figure 4. Medians and dispersion of effective countings for: (A) entrances; (B) exits; (C) pollen, performed by cs original (in blue), cs replicators (in red), and experts (in yellow). The dots are outliers, countings over or under 1.5 times the interquartile range.

3.2. Accuracy

Regarding replication analyses (comparison between cs original and cs replicator), the likelihood-ratio tests indicated the null models as the best ones for the perception of bees entering and leaving the nest (Table 1). However, for pollen perception, group effect was found. In this case, cs replicators presented odds increased by a factor of 4.63 compared to cs original (Table 2). The null models were the best also for the counts of entrance, exit, and pollen (Table 1). These results indicate that both groups performed countings quite similarly. Video quality did not present any effect in these analyses.

In the validation analyses (comparison between cs replicators and experts), null models were the best ones for the perception of entrance and exit and for the counts of entrance, exit, and pollen (Table 1). However, for pollen perception, the best model included the group variable, where the cs replicators had odds of perceiving pollen increased by a factor of 2.87 when compared to experts (Table 2). Here, the video quality index also did not add relevant information to explain the participants' data. Even though the model for pollen counts presented significant overdispersion despite adjustments (Supplementary Material Table S1), i.e., the residual variance was larger than expected under the fitted model, no effect of the tested variables was found.

Table 1. Likelihood-ratio tests for model selection in the replication (comparison between cs original and cs replicator) and validation (comparison between cs replicators and experts) analyses. χ^2 statistic, with the respective degrees of freedom (Df), and *p*-value for each test, are presented. Significant differences between models are followed by an asterisk.

	Response	Starting Model	Fixed Effect Removed	χ^2	Df	<i>p</i> -Value	
Replication	Perception	entrance	group + video quality	group	0.156	1	0.693
			video quality	video quality	0.045	1	0.832
		exit	group + video quality	group	0.356	1	0.551
	Count	pollen	video quality	video quality	0.065	1	0.799
			group + video quality	group	10.852	1	0.001 *
			group + video quality	video quality	0.064	1	0.801
		entrance	group	group	10.857	1	0.001 *
			group + video quality	group	1.674	1	0.196
			video quality	video quality	0.003	1	0.957
Validation	Perception	entrance	group + video quality	group	0.516	1	0.472
			video quality	video quality	0.056	1	0.812
		exit	group + video quality	group	0.592	1	0.442
	Count	pollen	video quality	video quality	0.003	1	0.958
			group + video quality	group	22.325	1	0.001 *
			group + video quality	video quality	0.077	1	0.781
		entrance	group	group	22.330	1	0.001 *
			group + video quality	group	0.039	1	0.843
			video quality	video quality	0.038	1	0.845
pollen	group + video quality	group	0.035	1	0.851		
	video quality	video quality	0.001	1	0.981		
	group + video quality	group	0.315	1	0.575		
video quality	video quality	0.219	1	0.640			

Table 2. Parameter estimates of the final models for pollen perception in replication and validation analyses.

Model	Comparison	Predictor	Estimate	SE	Pr(> z)	Odds/Odds Ratio
Pollen perception	Replication	cs original (intercept)	−2.7561	0.333	<0.001	0.06
		cs replicators	1.5332	0.508	0.003	4.63
	Validation	experts (intercept)	−2.3004	0.344	<0.001	0.1
		cs replicators	1.0529	0.229	<0.001	2.87

3.3. Precision

The MAD means were statistically equal between cs replicators and experts for entrance ($V = 24.5$, p -value 0.071) and exit ($V = 54$, p -value 0.052) counts. However, the MAD mean was higher for pollen counts for cs replicators in relation to experts ($V = 85$, p -value 0.005), who exhibited low dispersion for this task (Figure 5 and Supplementary Material Figure S1). These results are in accordance with the accuracy analyses, which presented significant differences for pollen perception between groups.

Data dispersion for the countings was positively correlated with the amount of bees in activity in the videos: entrance, cs replicators ($r = 0.8$, p -value < 0.001) and experts ($r = 0.78$, p -value < 0.001); exit, cs replicators ($r = 0.57$, p -value < 0.001) and experts ($r = 0.46$,

p -value 0.002). For entrance and exit, the pattern of increasing dispersion is similar for both groups, however, for pollen, the correlation is stronger for citizen scientists: cs replicators ($r = 0.94$, p -value < 0.001) and experts ($r = 0.43$, p -value 0.004) (Figure 6).

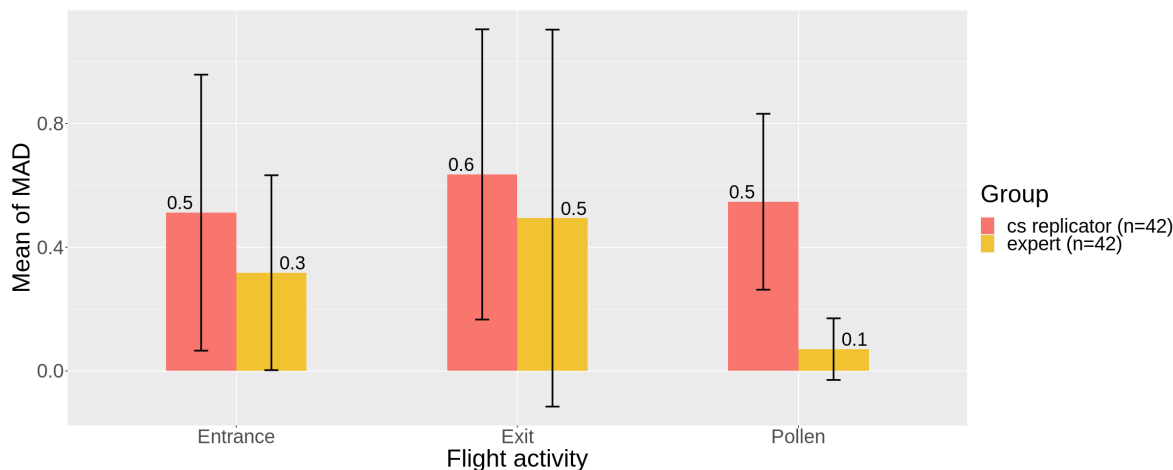


Figure 5. Mean of Median Absolute Deviation (MAD) of all videos for entrance, exit, and pollen counts, for cs replicators (in red) and experts (in yellow). Vertical bars indicate the Confidence Intervals (95%).

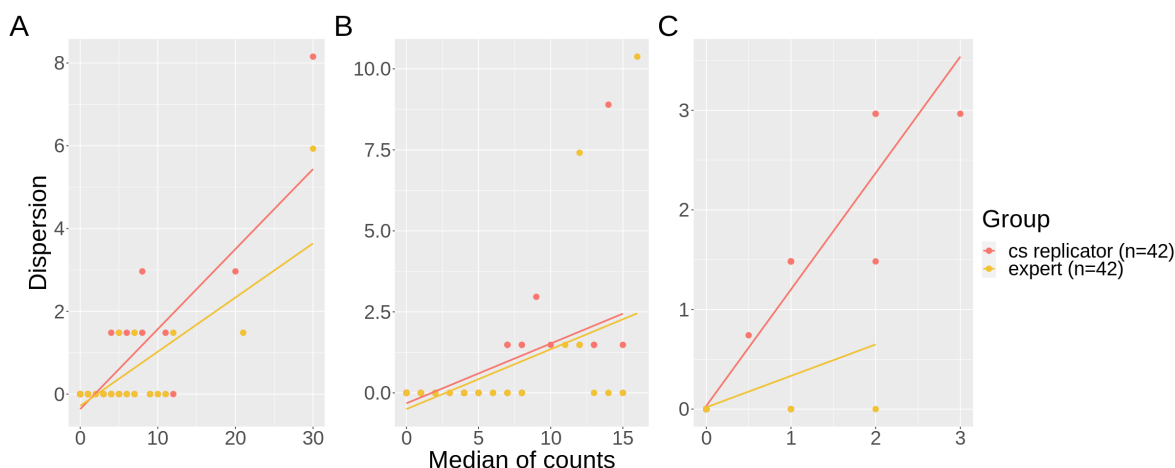


Figure 6. Correlation between the median of counts and dispersion (MAD) values of each video by group (cs replicator in red and experts in yellow), with tendency line (linear model line). (A) entrances; (B) exits; (C) pollen.

4. Discussion

Count-based activities are tasks with low to medium skill or training requirements [8]. In the present study, our results evidenced that, depending on the task, the accuracy of different groups can be affected. This was especially true when the perception of bees carrying pollen was considered. Thus, specific ability levels may be necessary to guarantee reliability in distinct contexts. For instance, perceiving a moving bee is significantly different from perceiving pollen present at a bee corbicula, which is a much smaller target and may require more volunteer training to reach the same level of quality of the experts. In the work of Bieluch et al. [41], CS program coordinators were interviewed about the contribution of volunteers in fish count-based protocols. They highlighted some aspects that can influence the counting accuracy, like the poor weather, high numbers of individuals passing at once, and the physical attributes of the counting site. Here, attributes of the video, which could be a proxy for context variation, did not affect countings. Target size may also influence perception, and large individuals and distinctive shapes can help to produce more accurate data [39]. It could pose as a barrier for data quality in our study since the size of the studied bees is small, varying between 4 mm and 5 mm, imposing

some difficulty for those who are analyzing the videos [42]. Finally, stingless bees normally speed up when near the nest entrance, which could also hinder the perception of bees carrying pollen by citizen scientists [43].

The observed divergence in pollen perception by cs replicators compared to the other groups cannot be fully explained in this study, where experts presented conservative counts, while cs replicators exhibited a permissive performance. We hypothesized that due to citizen scientists' eagerness to find some pollen and their lack of experience, they could overestimate the perception of this activity. If this is true, with more practical training effort in this protocol, we expect that cs replicators would present lower and less variation in the perception of pollen loads [44]. Some questionable research practices, such as falsification (wilful or unintended distortion of data or results), can negatively impact research [45], so they must be seriously considered when resulting from citizen science practices.

Generally, false negatives are a concern for researchers that deal with biodiversity occurrence data from citizen science (e.g., [46,47]), but false positives are generally considered unimportant, although they can lead to severe biases in conclusions about ecological systems [48]. Overestimation and underestimation of counts in citizen science protocols are reported in the literature. For instance, citizen scientists underestimated experts' countings of seals and sea lions in static images [39], countings of stomata and epidermal cells in static images [49], and the estimation of caterpillar density [50]. On the other hand, they overestimated fly and beetle density when applying a visual survey protocol in natural habitats [50]. These results varied according to the protocol and factors related to the individuals, such as experience and training [51]. However, empirical evidence suggests that citizen science data quality has often been sufficient for the projects' aims, and differences between citizen scientists' and professionals' data were not significant in most cases, e.g., 61.6% of studies showing no significant differences between citizen scientists and professionals in Aceves-Bueno et al. [27].

Considering the count precision in our study, intra-group dispersion was found in both groups, which is naturally expected in science, regardless of the individuals' expertise. For instance, in Swanson et al. [16], during an expert verification, precise counts of specimens were unresolvable in many of the images they were analyzing, and the specialists agreed on the number of individuals only 74% of the time. The authors concluded that multiple citizen scientists classifying an image could be more reliable if compared to a single person, even if this person is an expert. In our study, when more bees were in flight activity, more dispersion was found in all groups, which can reveal a greater difficulty for counting many bees at the same time. Other confounding factors could inflate countings, such as the presence of guards at the nest entrance. These guards are larger and heavier than the regular workers, hovering or standing next to the nest entrance tube (Figure 1A) [52], possibly being confused with bees in foraging activity when activity is high. We also found that CS data, specifically for pollen counts, showed higher dispersion than expert data. Likewise, data variability among citizen scientists was tested in Fehri et al. [53], in which the volunteers were engaged and trained on using rain gauge tools. Data dispersion was slightly higher in specific situations (high rainfall events), but in contrast, other events measurements (lower precipitation) showed more consistency for the group composed of citizen scientists. More straightforward tasks performed by volunteers tend to present less data dispersion when compared to more complex ones [53,54]. Additionally, volunteers generally improve their accuracy as they gain experience within a project [8]. The continuous execution of a task can promote personal learning and progress of the required skill [44] and improve data quality [27]. In our citizen science initiative, the participants of the outreach course learned to perform the protocol based on three video lessons and online guidance since any practical activity was not possible due to the restrictions imposed by the COVID-19 pandemic. Therefore, it is possible that an increased effort in the training of volunteers could lead to higher quality data in the present study, as has already been reported for other citizen science studies (e.g., [19,55,56]). Clear and objective protocols can also help in the rigorous collection of data [23,57].

Here, replication was used as a way to test for data quality, which is a practice reported to produce high rates of accuracy [16,39,58], since the combination of different and independent contributions decreases the errors observed individually in each one, according to the “Wisdom of Crowds” principle [59]. This quality assurance procedure can be used by default in the protocol design by proposing that all the collected data should be confirmed by a set of volunteers, or only in cases of unusual records, for example. A challenge with this approach is to define the value of agreement between individuals [39], sometimes with the support of expert checks. The level of accuracy needed will likely depend on the research question and the ability to perform post hoc statistical manipulation on these data [54]. In our case, replication provided reliable CS data for the counts of bees entering or leaving the nest. However, the perception of pollen was consistently different between groups (Figure 3, Tables 1 and 2) and countings were also highly variable for citizen scientists (Figures 5 and 6). Thus, although effective counts of pollen were reliable, both in replication and validation analyses (Figure 4), the observed variance in the replicators’ data can impose some data quality issues and compromise the application of these specific data.

As a consequence of these results obtained for pollen in the present pilot test of the protocol, a checkbox option labeled “pollen count was performed” was included in the platform, making it possible for citizen scientists to report an actual absence of pollen activity (“real” zeros) differently from the zeros that represent that they were unable to perform the task. Indeed, including an “I don’t know” option was shown to enhance data quality and contribute to the agreement among participants in a citizen science initiative [60]. In addition, we realized the importance of slowing down the video speed to make pollen loads easier to see. Thus we have also included this functionality in our platform as well. Future works may include developing an agreement algorithm to determine the consensual entrance, exit, and pollen activity in each video and the use of artificial intelligence to perform automatic counts.

5. Conclusions

Our results indicate that the flight activity protocol for stingless bees provides reliable data for bees entering and leaving the nest since original, replicated, and expert data were similar. These results are in accordance with our initial hypotheses that CS data do not differ statistically from those provided by specialists. However, a significant difference was found for the perception of pollen loads, with the cs replicator diverging from the experts and the cs original. Despite that, for those videos in which there was an agreement about pollen presence, i.e., in which both groups identified bees carrying pollen, the effective counts were very similar and confirmed the reliability of CS-collected data.

Data quality is a common concern in the era of data, and mechanisms to evaluate and improve quality are essential to ensure data applicability in its intended purpose. In CS, the problems are, in fact, comparable to those found in the traditional scientific exercise [3,8] and both models need to implement specific measures to guarantee the data quality and the other obtained products [54,61]. In CS, these quality requirements can be research questions of interest to participants, viable protocols, consistent training, evidence of observations, replication, expert review, among others. Several studies in different fields of knowledge, including the present one, report positive results and highlight characteristics that justify the investment in volunteers to act as scientists [12,27,62,63]. Depending on the particularities of the required task, adequate training and specific skill levels may be necessary, as in any other scientific approach. Our study highlights the importance of a multi-dimensional approach in CS data quality assessments to identify potential pitfalls and adequate protocols to improve data collection and use of CS data in research.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/insects12090766/s1>, Supplementary Material 1: Excel file with complete data-set, containing citizen scientists original (cs-original tab), replicated (cs-rep tab), and expert (exp tab) counts, along with video quality metric; (vq tab); Supplementary Material 2: Table S1: Dispersion of residual variance for each model. Significant dispersion models are highlighted in bold, Figure S1: Median

Absolute Deviation (MAD) value of each video for replicator citizen scientists and experts in different bee activities.

Author Contributions: Conceptualization, J.N.L., N.P.G.-L., S.K., C.B., T.M.F. and B.A.; methodology, J.N.L., N.P.G.-L., S.K., C.B. and B.A.; validation, J.N.L., N.P.G.-L., S.K. and C.B.; formal analysis, J.N.L., N.P.G.-L. and S.K.; investigation, J.N.L., N.P.G.-L., S.K. and C.B.; data curation, J.N.L., N.P.G.-L. and S.K.; writing—original draft preparation, J.N.L., N.P.G.-L., S.K., C.B., T.M.F. and B.A.; writing—review and editing, J.N.L., N.P.G.-L., S.K., C.B., T.M.F., B.A. and A.M.S.; visualization, J.N.L., N.P.G.-L. and S.K.; supervision, N.P.G.-L., S.K., T.M.F. and A.M.S.; project administration, N.P.G.-L., B.A., T.M.F. and A.M.S.; funding acquisition, A.M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Fundação de Amparo e Apoio à Pesquisa do Estado de São Paulo (FAPESP, grant numbers 2018/14994-1 and 2019/26760-8). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001; J.N.L grant number 88882.333367/2019-01; C.B. grant number 88882.377160/2019-01 and also by Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil (CNPq) A.M.S grant number 312.605/2018-8.

Institutional Review Board Statement: Ethical review and approval were waived for this study, as all analyzes were performed on counts that citizen scientists and experts performed on the videos. No collection of personal information or evaluation of personal data (e.g., insights or learning outcomes) was carried out with the citizen scientists. No handling or collection of bees was carried out.

Informed Consent Statement: Patient consent was waive as no collection of personal information or evaluation of personal data was carried out with the citizen scientists.

Data Availability Statement: The raw data used in this study are provided as Supplementary Material.

Acknowledgments: We thank all citizen scientists that participated in this study providing the videos and counts; the Provost of Culture and Extension at USP for the support to the outreach course on meliponiculture and citizen science; the experts Patrícia Nunes-Silva and Sergio Dias Hilário for their contribution to the validation stage, and Paula Marques Drago Oliveira for contributing to video quality index. We also would like to acknowledge the SURPASS2 project (Safeguarding pollination services in a changing world), FAPESP, and CAPES for their support.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CS	Citizen Science
cs	Citizen Scientist
Df	Degrees of Freedom
DQ	Data Quality
FPS	Frames per Second
GLMER	Generalized Linear Mixed-Effects Model
MAD	Median Absolute Deviation
MSE	Mean Square Error
MSSIM	Mean Structural SIMilarity
PCA	Principal Component Analysis
PSNR	Peak Signal-to-Noise Ratio
QA	Quality Assurance
QC	Quality Control
sd	Standard Deviation
SE	Standard Error

References

1. Bonney, R.; Ballard, H.; Jordan, R.; McCallie, E.; Phillips, T.; Shirk, J.; Wilderman, C.C. *Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education*; A CAISE Inquiry Group Report. Online Submission; Center for Advancement of Informal Science Education (CAISE): Washington, DC, USA, 2009.
2. Wiggins, A.; Crowston, K. From Conservation to Crowdsourcing: A Typology of Citizen Science. In Proceedings of the 2011 44th Hawaii International Conference on System Sciences, Kauai, HI, USA, 4–7 January 2011; pp. 1–10. [[CrossRef](#)]
3. Theobald, E.J.; Ettinger, A.K.; Burgess, H.K.; DeBey, L.B.; Schmidt, N.R.; Froehlich, H.E.; Wagner, C.; HilleRisLambers, J.; Tewksbury, J.; Harsch, M.A.; et al. Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biol. Conserv.* **2015**, *181*, 236–244. [[CrossRef](#)]
4. Chandler, M.; See, L.; Copas, K.; Bonde, A.M.; López, B.C.; Danielsen, F.; Legind, J.K.; Masinde, S.; Miller-Rushing, A.J.; Newman, G.; et al. Contribution of citizen science towards international biodiversity monitoring. *Biol. Conserv.* **2017**, *213*, 280–294. [[CrossRef](#)]
5. Fritz, S.; See, L.; Carlson, T.; Haklay, M.M.; Oliver, J.L.; Fraisl, D.; Mondardini, R.; Brocklehurst, M.; Shanley, L.A.; Schade, S.; et al. Citizen science and the United Nations sustainable development goals. *Nat. Sustain.* **2019**, *2*, 922–930. [[CrossRef](#)]
6. Pipino, L.L.; Lee, Y.W.; Wang, R.Y. Data Quality Assessment. *Commun. ACM* **2002**, *45*, 211–218. [[CrossRef](#)]
7. United States Environmental Protection Agency. *Guidance for Data Quality Assessment: Practical Methods for Data Analysis EPA QA/G-9 (QA00 Update)*; United States Environmental Protection Agency: Washington, DC, USA, 2000.
8. Kosmala, M.; Wiggins, A.; Swanson, A.; Simmons, B. Assessing data quality in citizen science. *Front. Ecol. Environ.* **2016**, *14*, 551–560. [[CrossRef](#)]
9. Ottinger, G. Buckets of resistance: Standards and the effectiveness of citizen science. *Sci. Technol. Hum. Values* **2010**, *35*, 244–270. [[CrossRef](#)]
10. Zhang, G.; Zhu, A.X. The representativeness and spatial bias of volunteered geographic information: A review. *Ann. GIS* **2018**, *24*, 151–162. [[CrossRef](#)]
11. Fowler, A.; Whyatt, J.D.; Davies, G.; Ellis, R. How reliable are citizen-derived scientific data? Assessing the quality of contrail observations made by the general public. *Trans. GIS* **2013**, *17*, 488–506. [[CrossRef](#)]
12. Lewandowski, E.; Specht, H. Influence of volunteer and project characteristics on data quality of biological surveys. *Conserv. Biol.* **2015**, *29*, 713–723. [[CrossRef](#)]
13. Callaghan, C.; Lyons, M.; Martin, J.; Major, R.; Kingsford, R. Assessing the reliability of avian biodiversity measures of urban greenspaces using eBird citizen science data. *Avian Conserv. Ecol.* **2017**, *12*, 12. [[CrossRef](#)]
14. Steinke, J.; van Etten, J.; Zelan, P.M. The accuracy of farmer-generated data in an agricultural citizen science methodology. *Agron. Sustain. Dev.* **2017**, *37*, 32. [[CrossRef](#)]
15. Gardiner, M.M.; Allee, L.L.; Brown, P.M.; Losey, J.E.; Roy, H.E.; Smyth, R.R. Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. *Front. Ecol. Environ.* **2012**, *10*, 471–476. [[CrossRef](#)]
16. Swanson, A.; Kosmala, M.; Lintott, C.; Packer, C. A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. *Conserv. Biol.* **2016**, *30*, 520–531. [[CrossRef](#)]
17. Fuccillo, K.K.; Crimmins, T.M.; de Rivera, C.E.; Elder, T.S. Assessing accuracy in citizen science-based plant phenology monitoring. *Int. J. Biometeorol.* **2015**, *59*, 917–926. [[CrossRef](#)] [[PubMed](#)]
18. Ratnieks, F.L.; Schrell, F.; Sheppard, R.C.; Brown, E.; Bristow, O.E.; Garbuzov, M. Data reliability in citizen science: learning curve and the effects of training method, volunteer background and experience on identification accuracy of insects visiting ivy flowers. *Methods Ecol. Evol.* **2016**, *7*, 1226–1235. [[CrossRef](#)]
19. van der Velde, T.; Milton, D.A.; Lawson, T.; Wilcox, C.; Lansdell, M.; Davis, G.; Perkins, G.; Hardesty, B.D. Comparison of marine debris data collected by researchers and citizen scientists: Is citizen science data worth the effort? *Biol. Conserv.* **2017**, *208*, 127–138. [[CrossRef](#)]
20. Alabri, A.; Hunter, J. Enhancing the quality and trust of citizen science data. In Proceedings of the 2010 IEEE Sixth International Conference on E-Science, Brisbane, QLD, Australia, 7–10 December 2010; pp. 81–88.
21. Yu, J.; Wong, W.K.; Hutchinson, R.A. Modeling experts and novices in citizen science data for species distribution modeling. In Proceedings of the 2010 IEEE International Conference on Data Mining, Sydney, NSW, Australia, 13–17 December 2010; pp. 1157–1162.
22. Tweddle, J.C.; Robinson, L.D.; Pocock, M.; Roy, H.E. *Guide to Citizen Science: Developing, Implementing and Evaluating Citizen Science to Study Biodiversity and the Environment in the UK*; NERC/Centre for Ecology & Hydrology: Wallingford, UK, 2012.
23. Freitag, A.; Meyer, R.; Whiteman, L. Strategies employed by citizen science programs to increase the credibility of their data. *Citiz. Sci. Theory Pract.* **2016**, *1*, 2. [[CrossRef](#)]
24. Bonter, D.N.; Cooper, C.B. Data validation in citizen science: A case study from Project FeederWatch. *Front. Ecol. Environ.* **2012**, *10*, 305–307. [[CrossRef](#)]
25. Follett, R.; Strezov, V. An analysis of citizen science based research: usage and publication patterns. *PLoS ONE* **2015**, *10*, e0143687. [[CrossRef](#)] [[PubMed](#)]
26. Kelling, S.; Johnston, A.; Hochachka, W.M.; Iliff, M.; Fink, D.; Gerbracht, J.; Lagoze, C.; La Sorte, F.A.; Moore, T.; Wiggins, A.; et al. Can observation skills of citizen scientists be estimated using species accumulation curves? *PLoS ONE* **2015**, *10*, e0139600. [[CrossRef](#)] [[PubMed](#)]

27. Aceves-Bueno, E.; Adeleye, A.M.S.; Feraud, M.; Huang, Y.; Tao, M.; Yang, Y.; Anderson, S.E. The accuracy of citizen science data: a quantitative review. *Bull. Ecol. Soc. Am.* **2017**, *98*, 278–290. [[CrossRef](#)]
28. Koffler, S.; Barbiéri, C.; Ghilardi-Lopes, N.P.; Leocadio, J.N.; Albertini, B.; Franco, T.M.; Saraiva, A.M. A buzz for sustainability and conservation: the growing potential of citizen science studies on bees. *Sustainability* **2021**, *13*, 959. [[CrossRef](#)]
29. Wolowski, M.; Agostini, K.; Rech, A.; Varassin, I.; Maués, M.; Freitas, L.; Carneiro, L.; BUENO, R.d.O.; Consolaro, H.; Carvalheiro, L.; et al. *Relatório Temático Sobre Polinização, Polinizadores e Produção de Alimentos No Brasil*; Editora Cubo: São Carlos, Brazil, 2019.
30. Cortopassi-Laurino, M.; Imperatriz-Fonseca, V.L.; Roubik, D.W.; Dollin, A.; Heard, T.; Aguilar, I.; Venturieri, G.C.; Eardley, C.; Nogueira-Neto, P. Global meliponiculture: challenges and opportunities. *Apidologie* **2006**, *37*, 275–292. [[CrossRef](#)]
31. Jaffé, R.; Pope, N.; Carvalho, A.T.; Maia, U.M.; Blochtein, B.; de Carvalho, C.A.L.; Carvalho-Zilse, G.A.; Freitas, B.M.; Menezes, C.; de Fátima Ribeiro, M.; et al. Bees for development: Brazilian survey reveals how to optimize stingless beekeeping. *PLoS ONE* **2015**, *10*, e0121157. [[CrossRef](#)] [[PubMed](#)]
32. Ascher, J.S.; Pickering, J. Discover Life Bee Species Guide and World Checklist (Hymenoptera: Apoidea: Anthophila). Available online: http://www.discoverlife.org/mp/20q?guide=Apoidea_species (accessed on 24 August 2021).
33. Lemelin, R.H. Entomotourism and the stingless bees of Mexico. *J. Ecotourism* **2020**, *19*, 168–175. [[CrossRef](#)]
34. Barbiéri, C.; Franco, T.M. Theoretical model for interdisciplinary analysis of human activities: Meliponiculture as an activity that promotes sustainability. *Ambiente Soc.* **2020**, *23*, 1–19. [[CrossRef](#)]
35. Gray, A.; Adjlane, N.; Arab, A.; Ballis, A.; Brusardis, V.; Charrière, J.D.; Chlebo, R.; Coffey, M.F.; Cornelissen, B.; Amaro da Costa, C.; et al. Honey bee colony winter loss rates for 35 countries participating in the COLOSS survey for winter 2018–2019, and the effects of a new queen on the risk of colony winter loss. *J. Apic. Res.* **2020**, *59*, 744–751. [[CrossRef](#)]
36. Gary, N.E. A method for evaluating honey bee flight activity at the hive entrance. *J. Econ. Entomol.* **1967**, *60*, 102–105. [[CrossRef](#)]
37. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]
38. Lin, D.C.C.; Chau, P.M. Objective human visual system based video quality assessment metric for low bit-rate video communication systems. In Proceedings of the 2006 IEEE Workshop on Multimedia Signal Processing, Victoria, BC, Canada, 3–6 October 2006; pp. 320–323.
39. Wood, S.A.; Robinson, P.W.; Costa, D.P.; Beltran, R.S. Accuracy and precision of citizen scientist animal counts from drone imagery. *PLoS ONE* **2021**, *16*, e0244040. [[CrossRef](#)]
40. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.
41. Bieluch, K.H.; Willis, T.; Smith, J.; Wilson, K.A. The complexities of counting fish: engaging citizen scientists in fish monitoring. *Maine Policy Rev.* **2017**, *26*, 9–18.
42. Wittmann, D. Aerial defense of the nest by workers of the stingless bee *Trigona (Tetragonisca) angustula* (Latreille) (Hymenoptera: Apidae). *Behav. Ecol. Sociobiol.* **1985**, *16*, 111–114. [[CrossRef](#)]
43. Tichit, P.; Alves-dos Santos, I.; Dacke, M.; Baird, E. Accelerated landing in a stingless bee and its unexpected benefits for traffic congestion. *Proc. R. Soc. B* **2020**, *287*, 20192720. [[CrossRef](#)] [[PubMed](#)]
44. Schmeller, D.S.; HENRY, P.Y.; Julliard, R.; Gruber, B.; Clobert, J.; Dziock, F.; Lengyel, S.; Nowicki, P.; Deri, E.; Budrys, E.; et al. Advantages of volunteer-based biodiversity monitoring in Europe. *Conserv. Biol.* **2009**, *23*, 307–316. [[CrossRef](#)] [[PubMed](#)]
45. Fanelli, D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE* **2009**, *4*, e5738. [[CrossRef](#)]
46. Sewell, D.; Beebe, T.J.; Griffiths, R.A. Optimising biodiversity assessments by volunteers: the application of occupancy modelling to large-scale amphibian surveys. *Biol. Conserv.* **2010**, *143*, 2102–2110. [[CrossRef](#)]
47. Hochachka, W.M.; Fink, D.; Hutchinson, R.A.; Sheldon, D.; Wong, W.K.; Kelling, S. Data-intensive science applied to broad-scale citizen science. *Trends Ecol. Evol.* **2012**, *27*, 130–137. [[CrossRef](#)]
48. Miller, D.A.; Weir, L.A.; McClintock, B.T.; Grant, E.H.C.; Bailey, L.L.; Simons, T.R. Experimental investigation of false positive errors in auditory species occurrence surveys. *Ecol. Appl.* **2012**, *22*, 1665–1674. [[CrossRef](#)] [[PubMed](#)]
49. Soul, L.C.; Barclay, R.S.; Bolton, A.; Wing, S.L. Fossil Atmospheres: a case study of citizen science in question-driven palaeontological research. *Philos. Trans. R. Soc. B* **2019**, *374*, 20170388. [[CrossRef](#)]
50. Hurlbert, A.H.; Hayes, T.E.; McKinnon, T.N.; Goforth, C.L. Caterpillars Count! A citizen science project for monitoring foliage arthropod abundance and phenology. *Citiz. Sci. Theory Pract.* **2019**, *4*, 1–12. [[CrossRef](#)]
51. Elliott, K.C.; Rosenberg, J. Philosophical foundations for citizen science. *Citiz. Sci. Theory Pract.* **2019**, *4*, 9. [[CrossRef](#)]
52. Grüter, C.; Menezes, C.; Imperatriz-Fonseca, V.L.; Ratnieks, F.L. A morphologically specialized soldier caste improves colony defense in a neotropical eusocial bee. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1182–1186. [[CrossRef](#)]
53. Fehri, R.; Khlifi, S.; Vanclooster, M. Testing a citizen science water monitoring approach in Tunisia. *Environ. Sci. Policy* **2020**, *104*, 67–72. [[CrossRef](#)]
54. Crall, A.W.; Newman, G.J.; Stohlgren, T.J.; Holfelder, K.A.; Graham, J.; Waller, D.M. Assessing citizen science data quality: An invasive species case study. *Conserv. Lett.* **2011**, *4*, 433–442. [[CrossRef](#)]
55. Foster-Smith, J.; Evans, S.M. The value of marine ecological data collected by volunteers. *Biol. Conserv.* **2003**, *113*, 199–213. [[CrossRef](#)]

56. Hunter, J.; Alabri, A.; van Ingen, C. Assessing the quality and trustworthiness of citizen science data. *Concurr. Comput. Pract. Exp.* **2013**, *25*, 454–466. [[CrossRef](#)]
57. Geoghegan, H.; Dyke, A.; Pateman, R.; West, S.; Everett, G. *Understanding Motivations for Citizen Science*; Final Report on Behalf of UKEOF, University of Reading, Stockholm Environment Institute (University of York) and University of the West of England; UKEOF: Wiltshire, UK, 2016.
58. Bain, R. Citizen science and statistics: Playing a part. *Significance* **2016**, *13*, 16–21. [[CrossRef](#)]
59. Surowiecki, J. *The Wisdom of Crowds*; Anchor: New York, NY, USA, 2005.
60. Torre, M.; Nakayama, S.; Tolbert, T.J.; Porfiri, M. Producing knowledge by admitting ignorance: Enhancing data quality through an “I don’t know” option in citizen science. *PLoS ONE* **2019**, *14*, e0211907. [[CrossRef](#)]
61. Bird, T.J.; Bates, A.E.; Lefcheck, J.S.; Hill, N.A.; Thomson, R.J.; Edgar, G.J.; Stuart-Smith, R.D.; Wotherspoon, S.; Krkosek, M.; Stuart-Smith, J.F.; et al. Statistical solutions for error and bias in global citizen science datasets. *Biol. Conserv.* **2014**, *173*, 144–154. [[CrossRef](#)]
62. Silvertown, J. A new dawn for citizen science. *Trends Ecol. Evol.* **2009**, *24*, 467–471. [[CrossRef](#)]
63. Krabbenhoft, C.A.; Kashian, D.R. Citizen science data are a reliable complement to quantitative ecological assessments in urban rivers. *Ecol. Indic.* **2020**, *116*, 106476. [[CrossRef](#)]