



Article

How Expert Is the Crowd? Insights into Crowd Opinions on the Severity of Earthquake Damage

Motti Zohar ^{1,*}, **Amos Salomon** ², and **Carmit Rapaport** ^{1,3}¹ Department of Geography and Environmental Studies, University of Haifa, Haifa 3498838, Israel; carmit.rapaport@gmail.com² The Geological Survey of Israel, Jerusalem 95501, Israel; salamon@gsi.gov.il³ NIRED—Institute for Regulation of Emergency and Disaster, College of Law and Business, Bnei Brak 511080, Israel

* Correspondence: motti.zohar@univ.haifa.ac.il

Abstract: The evaluation of earthquake damage is central to assessing its severity and damage characteristics. However, the methods of assessment encounter difficulties concerning the subjective judgments and interpretation of the evaluators. Thus, it is mainly geologists, seismologists, and engineers who perform this exhausting task. Here, we explore whether an evaluation made by semiskilled people and by the crowd is equivalent to the experts' opinions and, thus, can be harnessed as part of the process. Therefore, we conducted surveys in which a cohort of graduate students studying natural hazards ($n = 44$) and an online crowd ($n = 610$) were asked to evaluate the level of severity of earthquake damage. The two outcome datasets were then compared with the evaluation made by two of the present authors, who are considered experts in the field. Interestingly, the evaluations of both the semiskilled cohort and the crowd were found to be fairly similar to those of the experts, thus suggesting that they can provide an interpretation close enough to an expert's opinion on the severity level of earthquake damage. Such an understanding may indicate that although our analysis is preliminary and requires more case studies for this to be verified, there is vast potential encapsulated in crowd-sourced opinion on simple earthquake-related damage, especially if a large amount of data is to be handled.



Citation: Zohar, M.; Salomon, A.; Rapaport, C. How Expert Is the Crowd? Insights into Crowd Opinions on the Severity of Earthquake Damage. *Data* **2023**, *8*, 108. <https://doi.org/10.3390/data8060108>

Academic Editor: Vladimir Sreckovic

Received: 11 April 2023

Revised: 29 May 2023

Accepted: 2 June 2023

Published: 14 June 2023



Copyright: © 2023 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/>).

1. Introduction

Although occasionally incomplete and inaccurate, damage descriptions of earthquakes are of great importance in reconstructing the impact of past events, as well as in simulating scenarios of future events by means of modern seismological language and seismotectonic assessment. The common approach is to first interpret the given reports and assess their reliability, then determine the severity of the damage, and finally assign the appropriate seismic intensity degrees, which range from "not felt" to "complete destruction". The most notable seismic intensity scales are the American MMI [1], the European MSK [2], the Italian MCS [3], and the Japanese JMA (see explanation in [4]). Toward the end of the 20th century, a new European macroseismic scale was developed [5]. This 12-degree intensity scale (denoted, as with most of the other scales, by the Roman numerals I–XII), EMS-98, is designed to evaluate the degree of intensity of earthquake damage caused to modern structures and is also suitable for evaluating the level of damage caused to historical monuments.

Interpreting reports and, accordingly, determining the severity level of the damage caused to a given structure is not a straightforward process and faces several difficulties. First and foremost are the inherent inaccuracies and uncertainties involved in the interpretation of various reports (e.g., [6–8]) and the nature of the process, which is based on personal and subjective judgments [9]. Second, the reliability, quality, and quantity of the

reports may significantly affect the ability to make a decisive interpretation, particularly in the case of historical accounts. Third, there are differences among the various macroseismic scales that are in use worldwide [10], stemming mainly from a diversity in the style, design, type and quality of construction, the use of materials, and culture, all of which vary from one place to another. Given these complexities, the process has traditionally been the domain of experts and has as yet been implemented mainly by geologists, seismologists, and engineers.

2. Applying Crowd Wisdom in Earthquake Assessments

Over the last decade, the use of crowdsourcing, accessing the “wisdom of the crowd” or collective intelligence for resolving various issues, has gradually increased in many scientific disciplines. Kankamnge et al. [11] found that crowdsourcing definitions were mainly based on three basic features: outsourcing, crowd power, and voluntary participation and that it involves information shared by many people to create new knowledge. Crowdsourcing enables the collection of data from various sources, places, and perspectives that are eventually merged or combined into new knowledge [12]. Specifically, this is evident in spatial areas [13,14] and big data issues (for example, “Digital Earth”; see [15]). Crowd wisdom is created when groups of individuals who are not necessarily connected perform certain tasks or provide judgments, which, when aggregated, may enhance the quality and quantity of the knowledge gained while reducing biases [16]. This participation is subject to various motivations and channels, such as collaboration, competition, and knowledge-sharing (see [17] for additional examples). Such crowdsourcing or the collective intelligence of groups and crowds has been successfully used to address issues and concerns such as stock exchange market activities [18] and involvement in community services management [19], as well as in scientific assignments such as the volunteered geographic information (VGI) project and various life-sciences assignments [20].

In recent decades, crowdsourcing has played a part in disaster management and disaster risk reduction (DRR) efforts, given technology developments with significantly broad distribution, such as smartphones. This technology allows emergency situation managers and researchers to utilize crowdsourcing and crowd wisdom to collect data on emergencies at times immediately after their occurrence [21]. With the development of social media platforms such as Facebook and Twitter, scientists and first respondents also assess posts and messages sent out by subscribers when studying and monitoring natural disasters [22–25].

Given that crowdsourced data are based on voluntarism and spontaneous information-sharing, it is important to examine the data’s quality and reliability. Several studies in various research fields have already shown that crowdsourced opinions, which are based on non-expert information, are similar in quality and usability to the experts’ outputs; thus, they are of practical use and can be used together with the experts’ work [26]. Over time, non-experts have also demonstrated a positive learning curve, implying the potential of using their evaluations in scenarios where they have been given basic training or information [27].

3. Crowdsourced Seismology

Lately, the involvement of the public in data collection and crowdsourcing through cellphones has increased, fostering the emergence of a new discipline—crowdsourced seismology, wherein citizens who have felt an earthquake will voluntarily broadcast, report, or relay information based on their personal experience [28]. Along with the technological development of seismic monitoring devices, systems, and sensors, crowdsourced information was also found to be valuable, allowing for new, previously unrecognized perspectives. For example, the “Did You Feel It?” application enables the evaluation of earthquake intensities and how they are attenuated according to the distance from the epicenter, in near-real time, by assessing the public’s reports on what they had just felt [29]. In some cases, such data

were found to be as accurate as seismological equipment [30,31] and have been used in leading national seismological institutions worldwide [32].

Crowd wisdom also allows researchers to achieve a more comprehensive understanding of earthquakes, as it may compensate for the failure or lack of sensors in remote locations [33]. Obviously, this notion is not trivial, since meaningful and credited contribution depends on population availability and reporting by a relatively large number of people. Yet the degree of accuracy of crowdsourced evaluations and crowd wisdom is not satisfactory in all cases. According to Quitoriano and Wald [34], in such cases, highly damaged buildings can only be assessed by experts, who have access to the specifications (e.g., construction quality) needed to determine the correct intensity. Furthermore, the analysis of such data has shown differences in the evaluations of the respondents, depending on their location, as well as their status at the time they felt the earthquake, whether resting, sleeping, or moving around [35].

4. Crowdsourcing for Assessing the Severity Level of Damage

Traditionally, the process of evaluating the level of severity of earthquake damage is performed by experts. Here, we examine whether crowdsourcing can assist experts in studying the damage from past earthquakes, as well as in evaluating damage reports during future events. We ask two questions that challenge the traditional practice of expert evaluation: (1) Is there any difference between professional and nonprofessional evaluations, or, in other words, how close is a crowdsourced evaluation to an expert's opinion? (2) Can crowdsourcing be engaged and to what extent should it be engaged in assessing the severity level of earthquake damage? To answer these questions, we examine the extent to which the severity level of damage evaluated by professional experts differs from the level evaluated by students who have graduated from natural hazards studies, whom we regard as semiskilled, and how it differs from those evaluated by the non-professional crowd. The underlying hypothesis is that averaging the evaluations of large crowds and semiskilled persons compensates for non-professionalism and may result in values similar to those assigned by experts. Verifying this hypothesis may suggest an additional approach for evaluating the severity levels of damage caused by historical and modern earthquakes via online platforms.

5. Materials and Methods

5.1. Choosing a Case Study: The 1927 M6.2 Jericho Earthquake

On 11 July 1927, at approximately 3:00 p.m. (local time), a strong earthquake affected Mandatory Palestine in its entirety, resulting in several hundred fatalities, about a thousand injured casualties, and considerable damage to cities and villages [36]. The magnitude (size) was determined to be M6.2 and the epicenter (focus) was located to the north of the Dead Sea [37,38] (Figure 1). The wealth of information regarding this event includes professional reports [39–42], popular sources and newspapers [43–48], letters, documents, accounts [49,50], and photographs [51,52] as well as scientific studies [53–60]. This earthquake is unique for being the first destructive seismic event in the region that was described in numerous historical accounts [36] while also being recorded by seismographs [57]. Thus, it suits our needs as a careful, well-documented case study.

5.2. Constructing a Questionnaire and an Online Survey

Of the vast number of accounts associated with the 1927 Jericho earthquake, we have selected 15 trial reports regarding damage. Two additional damage reports from other events were taken from the study by Grünthal [5] for verification purposes.¹ In all, the reports comprise thirteen written documents and accounts, including original newspaper articles and professional reports made by engineers, along with four photographs of damaged sites.² Information from these seventeen sources was extracted and structured into a detailed questionnaire in English (Supplementary Materials File S1). Overall, we have attempted to cover the cases of damage associated with Arab and Jewish structures

in Palestine, Egypt, and Jordan. Some cases are presented in their original form (scanned bitmaps) while others were transcribed for clarity. Question 6 was presented first in its original form and then, again for verification purposes, it was transcribed and presented as question 15, in 16-point bold font. Apart from question 3, all the reports included the original description, the location and name of the damaged structure, and the reporting source. Altogether, the questionnaire comprised 18 questions. For categorizing the severity level of damage (see Supplementary Materials File S1, page 2) we used the definitions obtained from the European macroseismic-scale EMS project that accompanies the formal EMS-98 scale [5].³ We note, however, that the present study focuses only on assessing the severity level of the damage and not on intensity evaluations.

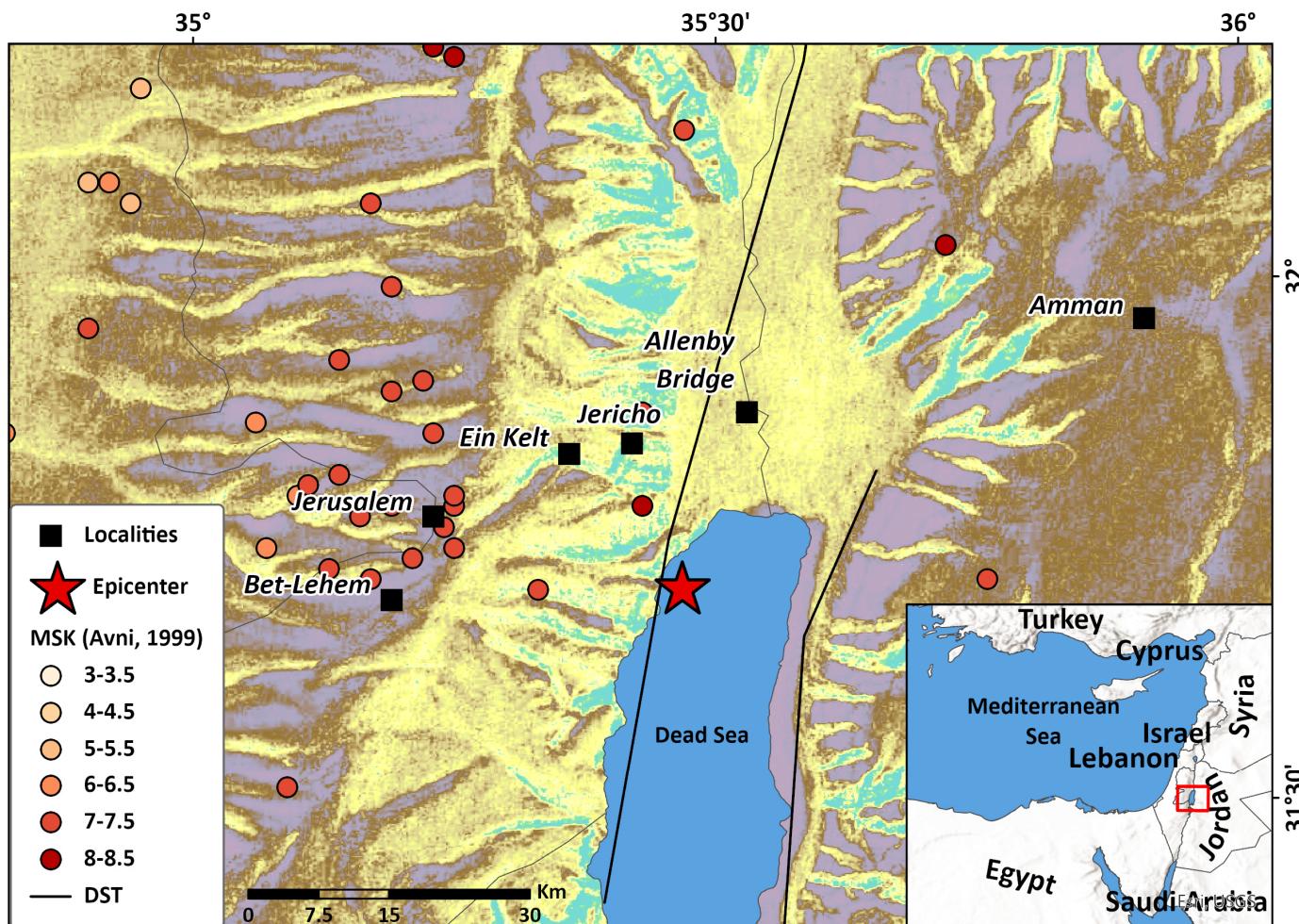


Figure 1. Damage distribution in central Mandatory Palestine (present-day Israel, Jordan, and the Palestinian Authority) as a result of the 1927 M6.2 Jericho earthquake. As a reference for our examination, the map presents the MSK seismic intensity values evaluated by [36]. The epicenter is denoted by a red star; black lines indicate the Dead Sea Transform faults; localities mentioned in the questionnaires are marked by black squares.

Next, two of the current authors (M.Z. and A.S.) evaluated the level of severity of the damage for each question and made an independent appraisal, relying on their professional expertise. The two evaluations were implemented individually and then averaged into a single dataset, referred to throughout this manuscript as “experts’ opinion”. The questionnaire (Supplementary Materials File S1) was then distributed among knowledgeable graduate students from the Department of Geography and Environmental Studies at the University of Haifa, who are well-versed in the study of natural hazards and historical earthquake damage. The students are referred to herein as “semiskilled”. Their answers

included voluntary personal information (age, gender, religiosity, education level, and whether they had previously experienced an earthquake) and the severity of the level of damage they determined for each question (see Supplementary Materials File S2). The survey was conducted in two phases (in two class groups) in January 2020 and January 2021. Altogether, this part of the analysis is based on the data from 44 respondents.

Based on the questionnaire developed for the graduate students, we formed another, shorter version that is more suitable for online crowd data collection, in order to keep the respondents focused throughout the survey. The online survey was distributed by iPanel (<https://www.ipanel.co.il/en/>, accessed on 10 April 2023) on a public survey website. iPanel is a survey company, with panel members representing the (entire) Israeli population according to parameters such as gender, age, religion, and education. To ensure that the questions, the instructions, and the questionnaire format were clear and comprehensible for use in an online setting, we first ran a pilot survey of 50 respondents. Having received no adverse comments, queries, or negative feedback on technical issues, we proceeded to the main sample. The latter dataset was collected in August 2021 and included an additional 560 respondents (610 respondents in total).⁴ The complete online questionnaire is given in Supplementary Materials File S3 (in Hebrew). The socio-demographic breakdown of the sample respondents (<https://figshare.com/s/282f9e0c035102985e1e>, partly in Hebrew) was 51% males and 49% females, the mean age was 43.6 (SD = 18.54), and 55% of the respondents had experienced at least one earthquake in the past, while the other 45% had not. Regarding their religious views, 53% answered that they were secular or nonreligious, 44% mentioned that they were religious or religious to some extent, and 3% mentioned they were very religious (Orthodox). Regarding education levels, 18% had a high school diploma, 33% had a bachelor's degree, 20% had a graduate degree, and 23% had reached a pre-academic education level.

6. Results

The general descriptive statistics of both the surveys (graduate students and the online crowd) are presented in Table 1. The survey of the graduate students included 18 questions (denoted as q1–q18; see Supplementary Materials File S1) while the online crowd survey was narrowed down to 10 of these 18 questions (q1–q4, q7, q9, q12, q14, q16, and q18; see Supplementary Materials File S3). Naturally, the damage levels assigned by the respondents are discrete, while the statistics can obviously be real numbers. Herein, we have noted the most important outcomes of the results.

At least one respondent of the graduate student group selected the highest available severity damage level (12) in response to seven of the questions (q3, q4, q7, q12, and q15–q17); the maximum (max) response for q18 (11.5) may also be added to this list. The lowest max response was given to questions q1 and q14, with the assignment of 8.5 and 8 levels only, respectively.

Notably, q1 is the only question based on a photograph that was answered with a max response lower than 11.5; the other three photograph-based questions (q7, q12, and q18) were replied to with a max response of 11.5 or 12. The mean response proposed in response to q1 is also relatively lower than that proposed for the other three questions.

There was a significant difference between the mean level returned for questions q6 (7.9) and q15 (8.9), which were based on the same report but were then presented differently to the respondents. The first (q6) used a scan of the source—an article that appeared in *The Times* newspaper from 14 July 1927, reporting on damage in Jerusalem ("Baghdadese Synagogue collapsed")—while the second (q15) is merely a font-enlarged transcript of the same report. This manipulation of the material was carried out in order to test whether the clarity of the text or the report affected the interpretation and the attribution of damage severity level.

Table 1. Statistics of the severity level of damage evaluated by the graduate student and online crowd survey: min, Q1 (first quartile), median, Q3 (3rd quartile), max, and mean values. Questions based on photographs are marked with (p); questions replied to with a maximal (12) or minimal (1) damage level by at least one respondent are marked red and blue, respectively.

Question Id																		
	q1 (p)	q2	q3	q4	q5	q6	q7 (p)	q8	q9	q10	q11	q12 (p)	q13	q14	q15	q16	q17	q18 (p)
Graduate students (all cases; n = 44)																		
Min	1	3	6	3.5	3	4	4.5	2	5.5	3	4	2	2	2	6	3	4	4.5
Q1	6	5	8	7	5.4	6	7	3.9	7	6	6	9.4	4.4	5.5	7	6	7	8
Median	6.5	6	9	8	6	8	8	5	7.3	7	7	10	6	6	9	7	8	9
Q3	7.1	7	10	9.6	7	9	8.6	6.1	8	7.6	8	11	7	7	10.1	7.1	9.6	9.5
Max	8.5	10	12	12	9	11	12	10	11	10	10	12	9	8	12	12	12	11.5
Mean	6.3	5.8	9.0	8.4	6.3	7.9	7.7	4.9	7.4	6.6	7	10	5.5	5.9	8.9	6.9	8.3	8.8
Online crowd survey (all cases; n = 610)																		
Min	1	1	1	1				3	1			4	1	1	1	1	1	1
Q1	6	6	8	9				7	6			10	6	6	6	6	6	8
Median	7	7	9	10				8	7			11	6	6	7	7	9	
Q3	8	8	11	11				10	9			12	7	7	8	8	10	
Max	12	12	12	12				12	12			12	11	11	12	12	12	
Mean	6.8	6.8	9.3	9.7				8.4	7.4			10.6	6.2	6.2	7.2	7.2	9.0	

The results of the online survey demonstrate that all ten questions (apart from q14) were answered with a max response of 12, while eight questions (all but q7 and q14) were given a minimum (min) response of 1.

Apart from a single question (q9), the mean response returned for each question of the online survey is higher than that for the corresponding graduate student survey question.

The distribution of both sets of survey responses is shown in the form of boxplots in Figure 2. For the graduate student dataset, an interquartile range (IQR) that is equal to or greater than two damage levels is shown for eight questions (q2–q4, q6, q8, q11, q13 and q15), whereas q6 and q15 present an IQR of three damage levels and above. Questions q1, q12, q14, and q16 tended to have a few extreme outliers, but this effect seems to be negligible considering their overall spread. For the online survey, the largest IQR of three damage levels is presented for questions q3, q7, and q9, while responses to the other questions have a value of two, excluding q14, which has an IQR value of one. For both datasets, an observed large IQR for a given question may indicate that the respondents were less certain of their damage-level selection; thus, their evaluations are more scattered in the boxplots than those for questions with a smaller IQR.

Does Crowd Opinion Correspond with the Experts' Opinion?

In this paper, we examine whether the severity level of damage assigned by semiskilled graduate students and by the online crowd are statistically different from those made by the experts. We interrogate the mean and mode of the values for severity level assigned by the respondents to each of the questions in the two surveyed samples and examine their correlation with the adjacent statistics from the experts' opinion dataset. To test these correlations, we used a paired *t*-test to verify the mean differences between two examined datasets, with a two-tailed F-test to check the equality of the adjacent variances. We have tested the differences between: (1) expert and graduate mean values; (2) expert and online crowd mean values; (3) expert and graduate mode values; (4) expert and online mode

values. The complete statistical *t*- and F-tests, along with detailed explanations, are elaborated upon in Supplementary Materials File S4. Accordingly, it was found that the *t*-tested *p*-value of the four comparisons, excluding the expert/graduate student mean difference, is above 0.05 and, likewise, the F-test *p*-value is above 0.05 (both with a confidence interval of 0.95). The highest Pearson correlation value is achieved for the expert/graduate student mean (0.91), while the correlations of the expert/graduate student mode, expert/online crowd mean, and expert/online crowd mode are 0.6, 0.85, and 0.86, respectively.

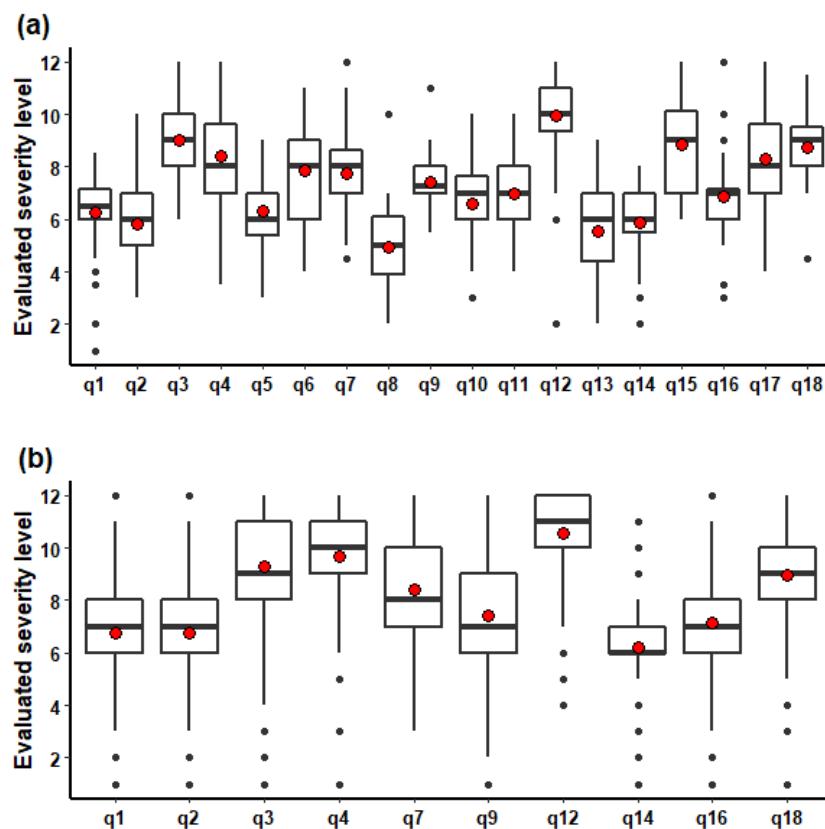


Figure 2. Boxplots presenting the distribution of replies to each question in the two datasets: **(a)** results of the questionnaire to 44 graduate students (Supplementary Materials File S2); **(b)** results of the online survey (Supplementary Materials File S2). The horizontal axis represents the questions (i.e., q1, q2, ..., qn), while the vertical axis shows the distribution of the answers according to the evaluated damage severity levels. Note that the lower image **(b)** shows fewer cases since the online survey included only 10 questions. The boxes represent the interquartile range (IQR, or the Q1–Q3 range) with a horizontal line denoting the median and a red circle representing the mean value (see also Table 1).

7. Discussion

7.1. The Similarity between the Two Surveys and the Experts' Opinion

The major question raised by this study is whether semiskilled graduate students' evaluations and an online crowd's evaluations of earthquake damage severity levels resemble those made by experts. The *t*- and F-tests of the comparison of IDs 2–4 in Supplementary Materials File S4 and Table S2 show that one cannot reject the null hypothesis of there being no difference between the mean range and variance of the two examined datasets. However, the conclusion drawn for the semiskilled-graduate students' mean-range difference (Supplementary Materials File S4, Table S2: ID 1) was that one should reject the null hypothesis, based on the *t*-test of a *p*-value that is lower than 0.05 (assuming that alpha = 0.05). Nevertheless, this *p*-value (0.045) is close enough to the alpha value that, had the mean ranges been slightly different, the results might have been different. Following the later notation, one should pay attention to the two questions that were based

on the same report but that were presented differently to the respondents (q6 and q15). The mean value of replies of all respondents to questions q6 and q15 indicated severity levels of 7.9 and 8.9, respectively, while the experts evaluated the same damage report at a level of 9. The difference indicates that the low evaluation value assigned as a response to q6 could have been related to the blurred scan, leading perhaps to a lack of attention in comparison with the clear and enlarged transcript shown for q15. Therefore, one can omit the mean value of q6 and rerun the *t*-test. Accordingly, the resulting *t*-test value and *p*-value were 1.8672 and 0.07922, respectively, with a Pearson correlation value of 0.92 (Supplementary Materials File S4, Table S2: ID 5)—that is, under this *t*-test iteration, one cannot reject the null hypothesis, meaning that the graduate students' evaluations are not statistically different from those of the experts.

7.2. What Is the Best Proxy for the Experts' Opinion Evaluation?

The absolute deviation of the graduate students' or crowd evaluations from those of the experts is demonstrated by the absolute mean difference between the corresponding evaluations of the two datasets (Supplementary Materials File S4, Table S2: IDs 1 and 2). The absolute means of comparisons 1 and 2 (mean severity level) that are presented are 0.46 and 0.54, respectively, while for comparisons 3 and 4 (mode severity level), they are 0.94 and 0.75, respectively. Additionally, the correlations of comparisons 1 and 2 are equal to or higher than 0.85 and seem to be better clustered within a confidence level of 0.95 (Supplementary Materials File S4, Figure S1: a1, b1). This notion is useful to justify the use of the mean value over the mode value in selecting the better proxy for the experts' opinion. The advantage of using the mean over the mode values is also demonstrated in Supplementary Materials File S4, Table S1 when examining individual questions with differences from the experts that are greater than 1 level. It appears that, for the mean values, the numbers of such cases are smaller than for the mode values (7 vs. 3 in the graduate student dataset and 4 vs. 2 in the online dataset). Altogether, these findings indicate that a resemblance to the experts' opinion evaluation is achieved more successfully when using the mean value of the questions rather than the mode value (Figure 3). In other words, although the mean is expressed in this case as a real number (and is not discrete), it is a more suitable indicator for use in severity-level evaluations than the mode value.

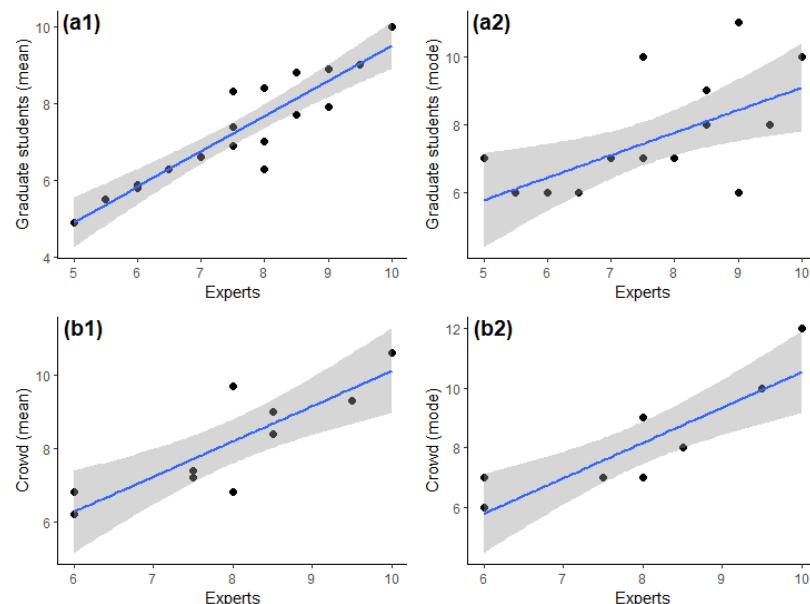


Figure 3. Correlations between the evaluations of the experts and the corresponding estimates made by the graduate students or the crowd: (a1) experts vs. graduate students' mean severity level; (a2) experts vs. graduate students' mode severity level; (b1) experts vs. the online crowd's mean severity level; (b2) experts vs. the online crowd's mode severity level.

7.3. The Characteristics of Successful Evaluation

Next, we examined the characteristics that lead to a successful evaluation and what might be the cause of a failed evaluation. Among the questions that were asked in the surveys given to both the graduate students and online respondents, three responses (q1, q6, and q11) exhibited a mean value that was greater than or equal to a level of 1 from the experts' opinion, while question q4 only showed such a difference in the online survey (Supplementary Materials File S4, Table S1). Of these, a single question (q1) was based on a photograph, while the other two (q6 and q11) were based on textual evidence. Seemingly, photographs are much more descriptive and intuitive than textual accounts and, thus, are expected to be simpler and more straightforward for the purposes of evaluation. However, they may contain information that only skilled experts are knowledgeable enough to resolve, while others are likely to ignore it. For example, in q1, the minaret of the mosque (on the left side of the image) appears at first to be complete, but a close inspection reveals that its top is ruined. However, the other three photographs (q7, q12, and q18) document the (clearly apparent) collapsed walls and buildings; two are associated with the 1927 Jericho earthquake (q7 and q18) and one (q12) is of a modern structure after the Spitak, Armenia earthquake in 1988 ([5], p. 83). The apparent collapse shown in the three photographs was detected effectively by the graduate students and the online respondents but was under-rated in q1. Therefore, it seems that in questions such as q1, experienced and skilled eyes are of great importance in fully identifying the damage and correctly evaluating its level of severity [52,61,62].

The level of damage presented for the two questions containing textual evidence (q6 and q11) was underestimated by the graduate students, whereas that of the damage presented for the third question (q4) was overestimated in the online survey (in relation to the experts' opinions). Question q6 is based on a newspaper report (*The Times*, 14 July 1927), while the second (q11) is based on a fragmentary scan of a letter sent by J.L. Magness, who was the president of the Hebrew University of Jerusalem at the time, to the engineers Chaikin and Kornberg (17 July 1927) concerning the damage caused to the university buildings by the 1927 earthquake. The damage presented in q6 was underestimated by the graduate students, perhaps due to the poor quality and blurriness of the scan. This may also be the case with q11, as the scan of the original report was also vague and may have hindered an accurate interpretation by the observer. The damage seen in q4 was overestimated significantly in the online survey, perhaps due to the report revealing that "a Russian maid-servant was killed [emphasis by the present authors] in the building by falling stones" (*The New York Times*, 13 July 1927). This is the only question in which a fatality was mentioned; it may have evoked an intuitive association with the greater severity of the damage rather than relating directly to the severity level of the damage as such.

On the other hand, q13, which is also based on a low-quality scan of a newspaper report (*The Times*, 14 July 1927) and which includes the word "unhurt", may have led to an accurate evaluation by the graduate students. Also notable are questions q2, q5, q8, and q10, which are based on a report made by the engineer, Michaeli [42]. Apart from q2, these instances were evaluated with only minor differences from the experts' opinion (less than a level of 0.5), implying that the reports were clear and received the proper level of attention from the readers, perhaps due to the mention of an "engineer" in all four questions, along with his detailed description.

7.4. Influence/Effects of Social Characteristics

The respondents of the online survey were also asked about their socioeconomic characteristics. First, we tested whether there was a correlation between the age of the respondents (18–89) and the differences in the evaluation of the experts' opinions and found no significant correlation (Figure 4a: $R^2 = 0.08$; p -value = 0.038). This is also the case when inspecting the difference with the respondent's age (Figure 4b: $R^2 = 0.09$; p -value = 0.017). This means that there was no age effect on the evaluation. Further on, we conducted *t*-and F-tests, along with the Pearson correlation value of the mean severity level difference

regarding the respondents' gender (male/female), their previous experience of feeling earthquakes (yes/no), and their levels of education (basic: elementary school or less; high school, with or without a matriculation certificate and other pre-academic studies; academic: bachelor's and graduate degrees and above). The results are presented in Table 2. Accordingly, while there are some differences between the various classifications, they seem to be negligible overall. For instance, the Pearson correlation value ranges slightly between 0.83 and 0.86. In other words, the classification factors of gender, the respondent's previous earthquake experience, and education level are not dominant when evaluating the severity level of damage, and their effect is minor.

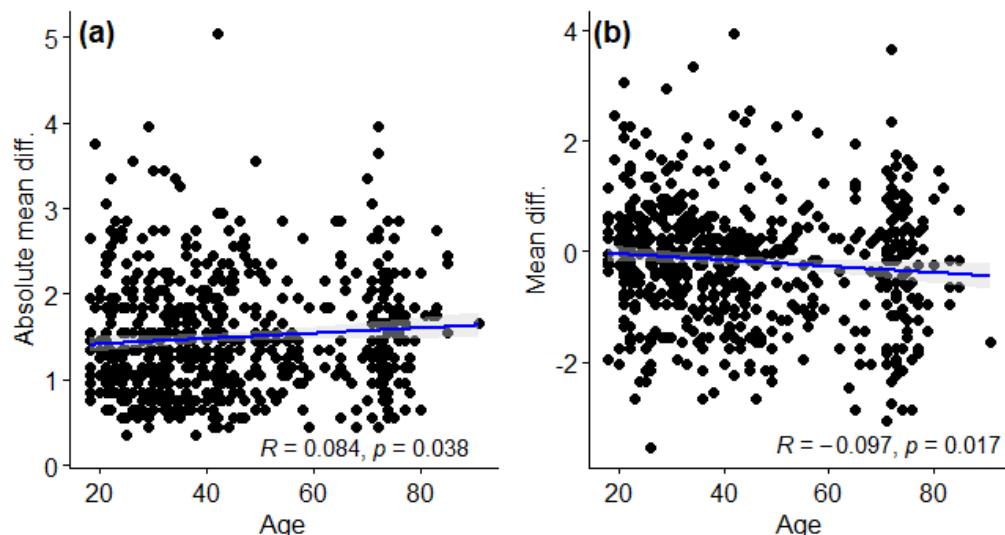


Figure 4. Scatter plot and correlation between the age of the online respondents and: (a) the absolute mean difference between their evaluation and the experts' evaluation; (b) the mean difference between their evaluation and the experts' evaluation.

7.5. The Importance and Potential Contribution of Crowdsourced Evaluation

Our findings show that crowdsourced damage assessments may complement existing methodologies for resolving and grading the levels of earthquake damage. That is, the crowdsourced contribution may assist experts in the exhaustive work of estimating seismic intensities, especially during and immediately after a damaging earthquake, when a wealth of information is accumulated in a very short period of time. We are aware that the whole process of deriving intensity degrees on the basis of crowdsourced evaluation is not a straightforward process and one that should be guided and supervised by experts. Nevertheless, although we stress that our insights are not intended to replace the necessity of professionals, they do suggest efficient and supportive methods that can assist experts in making those assessments.

A potential procedure embedding such approaches may include a process for transferring and accumulating damage reports in a hub, where they will be classified into damage reports that can be processed by the crowd and those that will be reserved only for the experts. Then, respondents will be given access to these reports via online platforms, and they can then assess the severity levels using a friendly, intuitive interface. The received responses will be centrally supervised, analyzed for their reliability and accuracy, and then disseminated publicly. Such a procedure is likely to speed up processing time and increase efficiency. However, in order to verify our preliminary findings and validate such a process, there is a need to consider the impact of local site-effects [63] as well as examining additional earthquake scenarios (e.g., [64–66]) and destructive events, such as the recent 2023 Turkey–Syria earthquakes [67].

Table 2. Pearson correlation, *t*-test, and F-test comparisons between the several classifications of online survey evaluations and experts' evaluations. Columns: Experts—the experts' evaluation; F—females' evaluations; F. Diff—the difference between Experts and F; M—males' evaluations; M. Diff—the difference between Experts and M; Ex. N—the respondents experienced an earthquake in the past; Ex. N. Diff—the difference between Experts and Ex. N; Ex. Y—respondents who did not experience an earthquake in the past; Ex. Y. Diff—the difference between Experts and Ex. Y; Ac—respondents with academic-level education; Ac. Diff—the difference between Experts and Ac; Ba—respondents with a basic education; Ba. Diff—the difference between Experts and Ba.

	Experts	F.	F. Diff	M	M. Diff	Ex. N	Ex. N. Diff	Ex. Y	Ex. Y. Diff	Ac.	Ac. Diff	Ba.	Ba. Diff
q1	8	6.6	1.4	6.9	1.1	6.6	1.4	6.9	1.1	6.9	1.1	6.7	1.3
q2	6	6.8	-0.8	6.7	-0.7	6.7	-0.7	6.8	-0.8	6.7	-0.7	6.8	-0.8
q3	9.5	9.6	-0.1	9	0.5	9.3	0.2	9.3	0.2	9.3	0.2	9.4	0.1
q4	8	10	-2	9.4	-1.4	9.7	-1.7	9.7	-1.7	9.7	-1.7	9.7	-1.7
q7	8.5	8.4	0.1	8.4	0.1	8.4	0.1	8.4	0.1	8.4	0.1	8.5	0
q9	7.5	7.5	0	7.4	0.1	7.5	0	7.4	0.1	7.3	0.2	7.5	0
q12	10	10.7	-0.7	10.4	-0.4	10.7	-0.7	10.4	-0.4	10.5	-0.5	10.6	-0.6
q14	6	6.1	-0.1	6.3	-0.3	6.2	-0.2	6.2	-0.2	6.2	-0.2	6.2	-0.2
q16	7.5	7.1	0.4	7.2	0.3	7	0.5	7.2	0.3	7.1	0.4	7.3	0.3
q18	8.5	9.1	-0.6	8.9	-0.4	9	-0.5	9	-0.5	8.9	-0.4	9	-0.5
Shapiro-Wilk		0.95		0.98		0.97		0.95		0.95		0.96	
p-value		0.697		0.983		0.898		0.708		0.838		0.81	
Pearson		0.83		0.86		0.84		0.85		0.85		0.84	
t-test		-0.911		-0.502		-0.656		-0.797		-0.588		-0.885	
p-value		0.385		0.627		0.528		0.445		0.57		0.399	
lower		-0.866		-0.606		-0.76		-0.708		-0.685		-0.781	
upper		0.368		0.386		0.418		0.339		0.402		0.341	
mean		-0.25		-0.11		-0.17		-0.18		-0.14		-0.22	
Mean. Abs		0.62		0.53		0.6		0.54		0.55		0.54	
F-test		0.681		0.909		0.728		0.838		0.802		0.769	
p-value		0.576		0.890		0.644		0.796		0.747		0.702	

8. Conclusions

For this paper, we asked whether evaluations of the level of severity of earthquake damage conducted by nonprofessional and semiskilled crowds can, under certain conditions, serve as proxies for experts' opinions. Using the 1927 Jericho earthquake as a test case, the mean and mode of the assessment datasets made by two crowd groups were found to be close enough to the experts' estimates and, thus, be utilized as proxies. Among these results, the mean value seems to be the most indicative. No apparent dominance was found when classifying the respondents by age, gender, religiosity, previous earthquake experience, and level of education, at least within the presented culture.

The crowd response to some questions showed an apparent dissimilarity between the surveys and it required the intervention of skilled experts. We have noticed that subtle nuances in textual description or the quality of a given photograph hinder non-experts from making a decisive evaluation. That is, the work of skilled experts cannot be replaced in the process of evaluating the level of severity of earthquake damage. However, we tentatively suggest that the contribution of crowdsourced evaluation may support experts and decision-makers in the time-consuming work of evaluating the damage, in particular, immediately after a destructive earthquake, when vast quantities of information flow, and time is of the essence. However, there is a need to expand the investigation into other cases, in order to draw firm theoretical and practical conclusions.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/data8060108/s1>. This manuscript has 4 Supplementary Materials files enclosed. Supplementary Materials File S4 includes the following references: [68,69].

Author Contributions: Conceptualization, M.Z.; methodology, M.Z., A.S. and C.R.; validation, A.S and C.R.; formal analysis, M.Z., A.S. and C.R.; investigation, A.S. and C.R.; resources, M.Z.; writing—original draft, M.Z. and A.S.; writing—review and editing, M.Z., A.S. and C.R.; visualization, M.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministry of Energy, State of Israel, Grant Number 3-14604. The APC was funded by the University of Haifa.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data is contained within the article or supplementary material. The online survey dataset is available at <https://figshare.com/s/282f9e0c035102985e1e>.

Acknowledgments: We would like to acknowledge the contributions of Shai Mizrahi from iPanel, Einat Magal from the Ministry of Energy, and Almog Arad from the University of Haifa.

Conflicts of Interest: The authors acknowledge there are no conflicts of interest that have been recorded.

Notes

- 1 Questions 3 (text) and 12 (image) in the questionnaire that was intended for skilled personnel.
- 2 The 1927 photographs were taken shortly after the earthquake by the American Colony Photography Department and were downloaded from the (G. Eric and Edith) Matson Photograph Collection at the Library of Congress (<https://www.loc.gov/pictures/collection/matpc/>, accessed on 10 April 2023). See Refs. [51,60].
- 3 Damage definitions were obtained from the European macroseismic scale EMS project at https://media.gfz-potsdam.de/gfz/sec26/resources/documents/PDF/EMS-98_short_form_English_PDF.pdf, last accessed on 10 April 2023.
- 4 A confidence level of 95% and a marginal error of 4%.

References

1. Wood, H.O.; Neumann, F. Modified Mercalli Intensity Scale of 1931. *Bull. Seismol. Soc. Am.* **1931**, *21*, 277–283. [[CrossRef](#)]
2. Medvedev, S.W.; Sponheuer, W.; Karnik, V. *Seismic Intensity Scale Version MSK 1964*; United Nation Educational, Scientific and Cultural Organization: Paris, France, 1965; p. 7.
3. Ferrari, G.; Guidoboni, E. Seismic scenarios and assessment of intensity: Some criteria for the use of the MCS scale. *Ann. Di Geofis.* **2000**, *43*, 707–720. [[CrossRef](#)]
4. Musson, R.M.W.; Cecić, I. Intensity and intensity scales. In *New Manual of Seismological Observatory Practice (NMSOP)*; Deutsches GeoForschungsZentrum GFZ: Potsdam, Germany, 2012; pp. 1–41. [[CrossRef](#)]
5. Grünthal, G. European Macroseismic Scale 1998 EMS-98. In Proceedings of the European Seismological Commission, Luxembourg, 7–12 September 1998.
6. Karcz, I. Implications of some early Jewish sources for estimates of earthquake hazard in the Holy Land. *Ann. Geophys.* **2004**, *47*, 759–792.
7. Ambraseys, N.N. Historical earthquakes in Jerusalem—A methodological discussion. *J. Seismol.* **2005**, *9*, 329–340. [[CrossRef](#)]
8. Cecić, I.; Musson, R.M.W. Macroseismic surveys in theory and practice. *Nat. Hazards* **2004**, *31*, 39–61. [[CrossRef](#)]
9. Musson, R.M.W. Intensity assignments from historical earthquake data: Issues of certainty and quality. *Ann. Geophys.* **1998**, *41*, 79–91.
10. Musson, R.M.W.; Grunthal, G.; Stucchi, M. The comparison of macroseismic intensity scales. *J. Seismol.* **2009**, *14*, 413–428. [[CrossRef](#)]
11. Kankanamge, N.; Yigitcanlar, T.; Goonetilleke, A.; Kamruzzaman, M. Can volunteer crowdsourcing reduce disaster risk? A systematic review of the literature. *Int. J. Disaster Risk Reduct.* **2019**, *35*, 101097. [[CrossRef](#)]
12. Miller, H.J.; Goodchild, M.F. Data-driven geography. *GeoJournal* **2015**, *80*, 449–461. [[CrossRef](#)]
13. Goodchild, M.F. Citizens as sensors: The world of volunteered geography. *GeoJournal* **2007**, *69*, 211–221. [[CrossRef](#)]
14. Fritz, S.; McCallum, I.; Schill, C.; Perger, C.; See, L.; Schepaschenko, D.; Obersteiner, M. Geo-Wiki: An online platform for improving global land cover. *Environ. Model. Softw.* **2012**, *31*, 110–123. [[CrossRef](#)]
15. Goodchild, M.F.; Guo, H.; Annoni, A.; Bian, L.; De Bie, K.; Campbell, F.; Woodgate, P. Next-generation digital earth. *Proc. Natl. Acad. Sci.* **2012**, *109*, 11088–11094. [[CrossRef](#)] [[PubMed](#)]
16. Dror, T.; Dalyot, S.; Doytsher, Y. Quantitative evaluation of volunteered geographic information paradigms: Social location-based services case study. *Surv. Rev.* **2015**, *47*, 349–362. [[CrossRef](#)]
17. Yu, C.; Chai, Y.; Liu, Y. Literature review on collective intelligence: A crowd science perspective. *Int. J. Crowd Sci.* **2018**, *2*, 64–73. [[CrossRef](#)]
18. Kaplan, C.A. Collective intelligence: A new approach to stock price forecasting. In Proceedings of the 2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace, Tucson, AZ, USA, 7–10 October 2001; pp. 2893–2898.
19. Liao, P.; Wan, Y.; Tang, P.; Wu, C.; Hu, Y.; Zhang, S. Applying crowdsourcing techniques in urban planning: A bibliometric analysis of research and practice prospects. *Cities* **2019**, *94*, 33–43. [[CrossRef](#)]
20. Salk, C.F.; Sturn, T.; See, L.; Fritz, S.; Perger, C. Assessing quality of volunteer crowdsourcing contributions: Lessons from the Cropland Capture game. *Int. J. Digit. Earth* **2016**, *9*, 410–426. [[CrossRef](#)]

21. Tavra, M.; Šerić, L.; Liseč, A.; Ivanda, A.; Divić, M.G. Transforming Social Media Posts into Volunteered Geographic Information using Data Mining Methods. In Proceedings of the 2021 6th International Conference on Smart and Sustainable Technologies (SpliTec), Bol and Split, Croatia, 8–11 September 2021; pp. 1–6.
22. Earle, P.S.; Bowden, D.C.; Guy, M. Twitter earthquake detection: Earthquake monitoring in a social world. *Ann. Geophys.* **2012**, *54*, 708–715. [[CrossRef](#)]
23. Crooks, A.; Croitoru, A.; Stefanidis, A.; Radzikowski, J. #Earthquake: Twitter as a distributed sensor system. *Trans. GIS* **2013**, *17*, 124–147.
24. Bhuvana, N.; Aram, I.A. Facebook and WhatsApp as disaster management tools during the Chennai (India) floods of 2015. *Int. J. Disaster Risk Reduct.* **2019**, *39*, 101135. [[CrossRef](#)]
25. Zohar, M.; Genossar, B.; Avny, R.; Tessler, N.; Gal, A. Spatiotemporal analysis in high resolution of tweets associated with the November 2016 wildfire in Haifa (Israel). *Int. J. Disaster Risk Reduct.* **2023**, *92*, 103720. [[CrossRef](#)]
26. Warby, S.C.; Wendt, S.L.; Welinder, P.; Munk, E.G.S.; Carrillo, O.; Sorensen, H.B.D.; Jenum, P.; Peppard, P.E.; Perona, P.; Mignot, E. Sleep-spindle detection: Crowdsourcing and evaluating performance of experts, non-experts and automated methods. *Nat. Methods* **2014**, *11*, 385–392. [[CrossRef](#)] [[PubMed](#)]
27. See, L.; Comber, A.; Salk, C.; Fritz, S.; Van Der Velde, M.; Perger, C.; Schill, C.; McCallum, I.; Kraxner, F.; Obersteiner, M. Comparing the quality of crowdsourced data contributed by experts and non-experts. *PLoS ONE* **2013**, *8*, e69958. [[CrossRef](#)] [[PubMed](#)]
28. Liang, W.-T.; Lee, J.-C.; Hsiao, N.-C. Crowdsourcing platform toward seismic disaster reduction: The Taiwan scientific earthquake reporting (TSER) system. *Front. Earth Sci.* **2019**, *7*, 79. [[CrossRef](#)]
29. Wald, D.J.; Quitoriano, V.; Worden, C.B.; Hopper, M.; Dewey, J.W. USGS “Did You Feel It?” internet-based macroseismic intensity maps. *Ann. Geophys.* **2012**, *54*, 688.
30. Bondár, I.; Steed, R.; Roch, J.; Bossu, R.; Heinloo, A.; Saul, J.; Strollo, A. Accurate locations of felt earthquakes using crowdsource detections. *Front. Earth Sci.* **2020**, *8*, 272. [[CrossRef](#)]
31. Steed, R.J.; Fuenzalida, A.; Bossu, R.; Bondár, I.; Heinloo, A.; Dupont, A.; Strollo, A. Crowdsourcing triggers rapid, reliable earthquake locations. *Sci. Adv.* **2019**, *5*, 9824. [[CrossRef](#)]
32. Tosi, P.; Sbarra, P.; De Rubeis, V.; Ferrari, C. Macroseismic intensity assessment method for web questionnaires. *Seismol. Res. Lett.* **2015**, *86*, 985–990. [[CrossRef](#)]
33. Deatrick, E. Crowdsourced seismology. *Eos* **2016**, *97*. [[CrossRef](#)]
34. Quitoriano, V.; Wald, D.J. USGS “Did You Feel It?”—Science and Lessons From 20 Years of Citizen Science-Based Macroseismology. *Front. Earth Sci.* **2020**, *8*, 120. [[CrossRef](#)]
35. Sbarra, P.; Tosi, P.; De Rubeis, V. How observer conditions impact earthquake perception. *Seismol. Res. Lett.* **2014**, *85*, 306–313. [[CrossRef](#)]
36. Avni, R. *The 1927 Jericho Earthquake, Comprehensive Macroseismic Analysis Based on Contemporary Sources*; Ben Gurion University: Beer-Sheva, Israel, 1999.
37. Shapira, A.; Avni, R.; Nur, A. A new estimate for the epicenter of the Jericho earthquake of 11 July 1927. *Isr. J. Earth Sci.* **1993**, *42*, 93–96.
38. Zohar, M.; Marco, S. Re-estimating the epicenter of the 1927 Jericho earthquake using spatial distribution of intensity data. *J. Appl. Geophys.* **2012**, *82*, 19–29. [[CrossRef](#)]
39. Willis, B. *To the Acting High Commissioner Lt-Col. G. S. Symes; Jerusalem, Israel*; British National Archives, CO 733/142/13: Richmond, UK, 1927; pp. 2–7.
40. Symes, G.S. *Earthquake Damage to Historic Buildings*; Department of Antiquities: Jerusalem, Israel, 1927.
41. Braver, A.I. Earthquakes in Eretz Israel from July 1927 till August 1928. In *Jerusalem*; Suckenic, L., Peres, I., Eds.; Darom Publishing: Jerusalem, Israel, 1928; pp. 316–325.
42. Michaeli, C.E. Notes on the Earthquake. *Constr. Ind.* **1928**, *11–12*, 9–12.
43. Anonymous. After the Earthquake. *Jewish Missionary Intelligence: Jerusalem, Israel*, 9 September 1927; 121.
44. Anonymous. The Earthquake in Eretz Israel. *Davar Tel Aviv, Israel*, 13 July 1927; 1.
45. Anonymous. The Earthquake in Eretz Israel. *Haaretz Tel Aviv, Israel*, 12 July 1927; 1–2.
46. Anonymous. The Earthquake. *Doar Hayom, Tel Aviv, Israel*, 12 July 1927; 1.
47. Anonymous. The Earthquake in Eretz Israel. *Haaretz, Tel Aviv, Israel*, 13 July 1927; 1.
48. Anonymous. The Earthquake in Palestine. *The Times, Tel Aviv, Israel*, 15 July 1927; 2.
49. Earthquake Turns Out a Real Catastrophe: Four Hundred Dead Bodies Recovered. Material Damage Exceeds Quarter Million Pounds. *The Palestine Bulletin, Jerusalem, Israel*, 13 July 1927.
50. Anglo Palestine Bank advertisement. *The Palestine Bulletin; Jerusalem, Israel*, 18 July 1927.
51. Biger, G.; Schiller, E. Rare collection of photographs following the 1927 earthquake. *ArielK* **1991**, *55–56*, 127–137.
52. Zohar, M.; Rubin, R.; Salomon, A. Earthquake damage and repair: New evidence from Jerusalem on the 1927 Jericho earthquake. *Seismol. Res. Lett.* **2014**, *85*, 912–922. [[CrossRef](#)]
53. Vered, M.; Striem, H.L. A macroseismic study and the implications of structural damage of two recent major earthquakes in the Jordan rift. *Bull. Seismol. Soc. Am.* **1977**, *67*, 1607–1613. [[CrossRef](#)]

54. Darvasi, Y.; Agnon, A. Calibrating a new attenuation curve for the Dead Sea region using surface wave dispersion surveys in sites damaged by the 1927 Jericho earthquake. *Solid Earth* **2019**, *10*, 379–390. [[CrossRef](#)]
55. Selzer, A.; Zohar, M. The ruins in the chemistry building are revised. The Hebrew university competes with the damages of the earthquake of July 1927. *Cathedra* **2019**, *171*, 75–96.
56. Salamon, A.; Katz, O.; Crouvi, O. Zones of required investigation for earthquake-related hazards in Jerusalem. *Nat. Hazards* **2010**, *53*, 375–406. [[CrossRef](#)]
57. Avni, R.; Bowman, D.; Shapira, A.; Nur, A. Erroneous interpretation of historical documents related to the epicenter of the 1927 Jericho earthquake in the Holy Land. *J. Seismol.* **2002**, *6*, 469–476. [[CrossRef](#)]
58. Rotstein, Y. Gaussian Probability Estimates for Large Earthquake Occurrence In The Jordan Valley, Dead-Sea Rift. *Tectonophysics* **1987**, *141*, 95–105. [[CrossRef](#)]
59. Ben-Menahem, A.; Nur, A.; Vered, M. Tectonics, seismicity and structure of the Afro-Eurasian junction—the breaking of an Incoherent plate. *Phys. Earth Planet. Inter.* **1976**, *12*, 1–50. [[CrossRef](#)]
60. Gavish, D. The American Colony and Its Photographers. In *Zev Vilnay's Jubilee volume*; Schiller, E., Ed.; Ariel Publishing House: Jerusalem, Israel, 1984; Volume 1, pp. 127–144.
61. Ambraseys, N.N.; Karcz, I. The earthquake of 1546 in the Holy Land. *Terra Nova* **1992**, *4*, 253–262. [[CrossRef](#)]
62. Zohar, M.; Rubin, R.; Salomon, A. Why is the minaret so short? Evidence on earthquake damage in Mt. Zion. *Palest. Explor. Q.* **2015**, *147*, 230–246. [[CrossRef](#)]
63. Aki, K. Local site effects on weak and strong ground motion. *Tectonophysics* **1993**, *218*, 93–111. [[CrossRef](#)]
64. Zohar, M. Temporal and spatial patterns of seismic activity associated with the Dead Sea Transform (DST) during the past 3000 Yr. *Seismol. Res. Lett.* **2020**, *91*, 207–221. [[CrossRef](#)]
65. Kagan, Y.Y.; Jackson, D.D.; Rong, Y. A new catalog of southern California earthquakes, 1800–2005. *Seismol. Res. Lett.* **2006**, *77*, 30–38. [[CrossRef](#)]
66. Tan, O.; Tapirdamaz, M.C.; Yörük, A. The earthquake catalogues for Turkey. *Turk. J. Earth Sci.* **2008**, *17*, 405–418.
67. Dal Zilio, L.; Ampuero, J.-P. Earthquake doublet in Turkey and Syria. *Commun. Earth Environ.* **2023**, *4*, 71. [[CrossRef](#)]
68. Peck, R.; Olsen, C.; Devore, J.L. *Introduction to Statistics and Data Analysis*, 3rd ed.; International Student Edition; Thomson Brooks/Cole: Geneva, Switzerland, 2008.
69. Shapiro, S.; Wilk, M. An analysis of variance test for normality. *Biometrika* **1965**, *52*, 591–611. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.