CITIZEN SCIENCE:
THEORY AND PRACTICE

Odenwald, S. 2018. A Citation Study of Citizen Science Projects in Space Science and Astronomy. *Citizen Science: Theory and Practice*, 3(2): 5, pp. 1–11. DOI: https://doi.org/10.5334/cstp.152

RESEARCH PAPER

A Citation Study of Citizen Science Projects in Space Science and Astronomy

Sten Odenwald

This paper presents a citation study of 143 publications in refereed journals resulting from 23 citizen science (CS) projects in space science and astronomy. The projects generated a median of two papers during their average of six years of operation. The 143 papers produced 4,515 citations for a median of 10 citations/paper. These papers were compared to a uniform group of papers published in the year 2000 in refereed space science journals. The CS papers have an average annual peak citation rate that is about four times the average of the year-2000 sample. The CS citation history profiles peak within 3 years after paper publication but decline thereafter at a faster pace than the average paper published in 2000. This suggests that CS papers "burn brighter" but remain of interest for only half as long as other papers in space science and astronomy. Nevertheless, CS papers compare well with some of the most highly ranked "Top-1000" research papers of the modern era. The proportion of CS papers surpassing 200 citations is one-in-26, which is 40-fold higher than the proportion for the typical paper published in 2000. The study concludes that CS projects are not only as good as conventional non-CS research projects in generating publishable results, but can actually outperform the citation rates of the typical non-CS papers in space science and astronomy.

Keywords: citations; space science; citizen science

Introduction

The idea of citizen science (CS) has been around for a very long time, and arguably goes back to the "gentleman science" time of Galileo and his contemporaries. The advent of the Internet has tremendously expanded the opportunity for the public to engage in scientific research using the CS model. A simple search on GOOGLE, for example, turns up projects as simple as merely noting the faintest stars that you can see in a constellation (GLOBE at Night) or as complex as The Milky Way Project, which has participants view thousands of infrared images of the Milky Way from the Spitzer Space Observatory and identify interstellar bubbles.

Among the many concerns that scientists raise to the CS approach for conducting scientific research is whether the results will actually be meaningful to the scientific community at large. This consideration relates to questions about whether novice, non-professional participants are capable of delivering high-quality data to the Principle Investigator or project team (Lewandowski 2015; Kosmala et al. 2016). Labor and software development costs also have to be controlled that are associated with setting up and operating a CS project (Sauermann and Franzoni 2014). The bottom-line concern for most scientists, however, is whether the CS effort will lead to publishable results that will advance scientific knowledge in some measurable way.

This paper describes the application of citation metrics to research papers published specifically within the CS project areas of space science and astronomy. It also attempts to place the products of CS-leveraged research within the context of general space science research. The objective is to use traditional citation analysis to determine whether CS research is quantitatively different from traditional space science research in its depth of penetration into the general research dialog.

Methods

One of the earliest citation studies—of all scientific papers, not just those focused on citizen science—was by de Solla Price (1965), who investigated "networks of scientific papers" by linking each published paper to other papers that later reference them. In 1979, Abt (1981) embarked on a study of 326 astronomy-related research papers published in 1961 and cited from 1961 to 1979. Using data from the Science Citation Index (1962), the study found 6,070 citations to these papers, and established that a typical research paper published at that time enjoyed approximately one citation per year, e.g., 6070/(326 * 18). Only about one-in-eighteen papers generated more than three citations per year irrespective of whether they were primarily observational or theoretical in nature. The peak

ADNET Systems, US
sten.f.odenwald@nasa.gov

citation period occurred approximately five years after publication and declined at a constant rate thereafter.

Subsequent studies by van der Kruit (2004) and Pearce (2004) of the NASA Astrophysics Data System (ADS) refined this initial study by Abt (1981) and provided various ancillary measures of citation lifespans, impact, and sustainability of research papers in astronomy. Among their findings were that one paper in one hundred accumulates more than 91 citations over five years, and that 10% of all astronomers who publish two or more papers in any five-year period receive over 200 citations. Moreover, in astronomy the self-citation rate is about 25%, and citations to research articles are strongly affected by the rankings of the journals in which they publish. In the specific genre of CS, Watson and Floridi (2016) investigated a broad range of projects within the *Zooniverse* platform and found that such projects produced papers that were consistently more cited than traditional research projects.

To assess the impact of CS projects in space science and astronomy, a total of 48 projects in these areas were identified in the CS project catalogs by SciStarter (Cavalier 2014), The National Geographic (1996), Scientific American (2011), Citizen Science Alliance (2007), Science@NASA (2017), and the US Federal Government (Wilson Center 2017). The 23 CS projects in this sample which provide bibliographic information are listed in **Table 1**. Column 2 gives the catalog used to identify the project papers, where SciSt refers to SciStarter, SciAm refers to Scientific American, and CSA refers to Citizen Science Alliance. Column 3 is the total number of a project's formal and informal publications; Column 4 is the start year for the project; and Column 5 is the URL of the project's bibliographic listing accessed December 10, 2017.

Detailed tallies for each project are shown in **Table 2**. The first three columns are self-explanatory. Column 4 gives the number of project operating years, including the inception year, through the end of 2017. Column 5 is the estimated number, P, of known participants in each project obtained from the project's online compilations and other online mentions of the project. Column 6 provides the total number of publications of all types; column 7 gives the total number of refereed publications. Column 8 gives the median number of authors for the project's refereed papers, and column 9 gives the total number of citations to these refereed papers through the end of 2017.

Among the 23 projects in **Table 2**, the publications shown in column 6 total 238 papers. Of these papers, 143 appeared in refereed science journals resulting from

Table 1: Citizen science projects in space science and astronomy with published bibliographies.

(1)	(2)	(3)	(4)	(5)
Name	**Catalog**	**Pub**	**Year**	**Bibliography URL**
Galaxy Zoo	SciSt	58	2008	https://www.zooniverse.org/about/publications
HamSCI	SciSt	38	2013	http://www.hamsci.org/publications?page=1
Sungrazer Project	SciSt	27	1996	https://tinyurl.com/ybzxsq3h
Einstein@Home	SciSt	22	2008	https://einsteinathome.org/science/publications
Radio JOVE	SciSt	21	1998	https://radiojove.gsfc.nasa.gov/library/pubs.htm
Planet Hunters	SciSt	12	2010	https://www.zooniverse.org/about/publications
Solar Storm Watch	SciAm	7	2012	https://www.zooniverse.org/about/publications
Supernova Hunters	CSA	7	2010	https://www.zooniverse.org/about/publications
Inspire-Ionosphere	SciSt	6	2007	http://solar-center.stanford.edu/SID/science/
Galaxy Zoo-S. Nova	CSA	5	2009	https://www.zooniverse.org/about/publications
Milky Way Project	SciSt	5	2012	https://www.zooniverse.org/about/publications
Galaxy Zoo-Mergers	CSA	5	2009	https://en.wikipedia.org/wiki/Galaxy_Zoo
Aurorasaurus	SciSt	6	2013	https://tinyurl.com/y8ws2u4v
Space Warps	SciAm	4	2015	https://www.zooniverse.org/about/publications
PANOPTES	SciSt	4	2015	http://www.projectpanoptes.org/status.html
Disk Detective	SciSt	2	2014	https://www.zooniverse.org/about/publications
Snapshot Supernova	CSA	2	2015	https://www.zooniverse.org/about/publications
Radio Galaxy Zoo	CSA	2	2015	https://www.zooniverse.org/about/publications
Ice Hunters	CSA	1	2012	https://www.zooniverse.org/about/publications
Andromeda Project	CSA	1	2012	https://www.zooniverse.org/about/publications
Planet Four	SciSt	1	2016	https://www.zooniverse.org/about/publications
Backyard Worlds	SciSt	1	2017	https://www.zooniverse.org/about/publications
Moon Zoo	SciSt	1	2016	https://www.zooniverse.org/about/publications

Table 2: Statistics for publications ordered by project start year.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Project	**Start Year**	**Operating Years**	**P**	**Total Pubs**	**Ref. Pubs**	**Media Authors**	**Total Cites**	**Pub. Rate**
A	SunGrazer	1996	22	100	27	11	2	274	0.25
B	RadioJove	1998	20	1000	21	0			
C	Inspire-Ionosphere	2007	11	1000	6	1	5	29	0.02
D	Galaxy Zoo	2008	10	250,000	58	57	12	2984	0.48
E	Einstein@Home	2008	10	500,000	22	22	20	384	0.11
F	Galaxy Zoo-Supernova	2009	9	3,000	5	5	24	150	0.02
G	Galaxy Zoo – Mergers	2009	9	140,000	5	4	6	135	0.07
H	Planet Hunters	2010	8	300,000	12	12	21	291	0.07
I	Supernova Hunters	2010	8	3,000	7	1	25	0	0.01
J	Solar Storm Watch	2012	6	16,000	7	7	11	43	0.11
K	Milky Way Project	2012	6	20,000	5	4	5	134	0.13
L	Ice Hunters	2012	6	100	1	1	22	9	0.01
M	Andromeda Project	2012	6	10,000	1	1	15	10	0.01
N	HamSci	2013	5	50	38	1	9	3	0.02
O	Aurorasaurus	2013	5	5,000	6	5	3	7	0.33
P	Disk Detective	2014	4	30,000	2	2	20	5	0.03
Q	Space Warps	2015	3	4000	4	4	21	40	0.06
R	PANOPTES	2015	3	20	4	0			
S	Snapshot Supernova	2015	3	60,000	2	0			
T	Radio Galaxy Zoo	2015	3	10,000	2	3	25	14	0.06
U	Planet Four	2016	2	100,000	1	0			
V	Moon Zoo	2016	2	9,000	1	1	8	3	0.06
W	Backyard Worlds	2017	1	50,000	1	1	12	0	0.08
	Median Values:	**2012**	**6**	**10,000**	**5**	**2**	**12**	**10**	**0.06**

19 of the projects (Column 7). Abstracts to professional conferences or web-based technical essays were not included, nor were articles appearing only on preprint servers such as *Astro-ph* (e.g., arXiv). The refereed papers are distributed among the various journals such that the *Monthly Notices of the Royal Astronomical Society* (49%) and *The Astrophysical Journal* (25%) account for 74% of the total, with *Physics Review D*, *Astronomical Journal*, *Space Weather,* and *Icarus* accounting for 19%, and a variety of other journals making up the remaining 8%. The preponderance of publications in MNRAS is largely due to the *Galaxy Zoo* project.

The 4,515 citations generated by the 143 refereed papers were found using the Web of Science "Science Citation Index" (hereafter WoS), which is a widely used citation research tool in the bibliometric community. Eight papers were published in 2017 and have not had enough time to generate citations in refereed journals. In addition, 16 papers could not be found in the WoS archive, so the SAO/NASA Astrophysical Data System (ADS) Database (ADS 2018) was used in these cases. The raw citation tallies for each project and paper are found in the accompanying supplemental materials.

The original sample of 48 space science-related CS projects included 25 that did not have identifiable bibliographies. Many of these projects have been in operation for five or more years. One might expect CS projects to be heavily biased in favor of announcing their results, because of their novelty as a research approach with massive public involvement. However, taking the lack of announced publications at face value suggests that nearly half of all CS space-science projects never lead to publishable results in either the informal or formal literature. This may result from the project not being in progress long enough to return publishable results, or the scientific value of the findings not being deemed significant, or both.

Nevertheless, this result may not be exceptional in the larger context of astronomical research. A CS project can be compared to a specific research program undertaken by an astronomer or a team of astronomers at a research facility. In 2018, a study of 1,278 teams of astronomers who had observation programs on the European Southern

Observatory Very Large Telescope between 2006 and 2013 found that even after ten years from the time of the observation session, as many as 50% of the teams had no publications (Clery 2018). Surveys of other telescope facilities by Smith (2018) yield similar 30–50% non-publication rates. This result is potentially problematic for CS projects, because a key argument often made to funding agencies is that a CS project will lead to publishable results. A further study of this "silent" CS cohort is warranted, but it may simply show that CS projects suffer from the same issues faced by ground- and facility-based research.

Additional insights gleaned from **Table 2** based on the tally of papers in refereed journals are that in terms of median values, the number of refereed papers per project is about two (**Table 2**: col. 7). These papers had about 12 authors and co-authors (**Table 2**: col. 8), and by 2017 the projects had been in operation for about six years (**Table 2**: col. 4). In terms of the scale of these projects, because the participation levels span many orders of magnitude, the median of 10,000 participants is likely to be a fair indicator of the typical project. In the following sections, this study "drills down" through these data to try to uncover other measures of how well these projects and their results have become integrated into the general scientific literature.

Results and discussion

Paper publication rates

The publication rate of an astronomer, a project, or an institution is the simplest statistic commonly used to gauge impact or productivity. It is based upon the number of research articles (N) divided by the number of participants, P, and the span of time over which the research effort was performed, T, and represented by Rp = N/(PT) in units of papers/year/astronomer. To compare Rp between different time periods, normalizing the rate by the number of authors partially compensates for historical trends in the growth of authorship lists, or in the number of professional astronomers employed to conduct research.

According to a study by Abt (1998), the growth in the number of citable papers in the popular journals *Astronomy and Astrophysics* (A&A), *The Astronomical Journal* (AJ), ApJ, PASP, and MNRAS from 1960–1996 was directly proportional to the number of professional astronomers who are members of the American Astronomical Society (AAS), with an average of 0.4 papers/year/astronomer.

The ADS publication database was used to identify a total of 9,000 papers published in 2017 in the journals surveyed by Abt (1981, 1998). The total AAS membership included about 7,000 astronomers for an average rate of 1.3 papers/year/astronomer. This estimate is actually a lower limit, because during any given year, only a fraction of AAS members publish research. Nevertheless, this three-fold increase in publication rate between 1961 and 2017, despite only a 50% increase in AAS membership, speaks to a dramatic and non-linear change in the "productivity" of modern-day astronomers no longer proportional to AAS membership levels. This suggests that comparing publication rates across many decades will be difficult, because factors other than population size seem to be at play in deciding how a specific paper will fare using publication-based metrics alone.

For the CS papers in **Table 2**, the publication rate, which was previously defined as Rp, can be calculated for each project (col. 10) by dividing the number (N) of refereed papers (col. 7) by the median number, P, of authors (col. 8) and then dividing by the duration, T, of the project (col. 4). The resulting average publication rate across all 19 projects with refereed papers is 0.10 papers/year/astronomer. This aggregate average for CS projects is more than ten-fold lower than the nominal rate calculated for 2017 for non-CS research in astronomy. One reason may be that CS papers have a larger-than-average number of co-authors involved in the paper publication. In a study of general trends in astronomical publications, Hennekin (2012) noted that in 1960, 80% of papers involved 1–2 authors, while only 15% did so by 2009, with a current median level near 4 authors/paper. The median number of authors involved in all of the CS projects in **Table 2** is 12, which is significantly higher than typical non-CS publications in astronomy, perhaps providing some of the reason for a lower CS publication rate per author. CS publications appear to attract more co-author participation than typical non-CS papers. The aggregate value of 0.10 papers/year/CS author may actually be a factor of three or four times higher when corrected for general co-authorship increases, but is still significantly lower than the previously discussed estimate for the 2017 rate of 1.3 papers/year/astronomer. Taken by itself, this number suggests that CS papers are being produced by projects at a significantly lower rate than the average modern astronomical papers. This result may also be related to the non-publication effect found by Clery (2018), in which scientific data from a research program, once available, are more difficult to analyze than expected and lead to few or no publications. Beyond simple paper-counting methods, there are other ways to explore how well CS papers and research fares in contributing to the general professional research conversation. The next section discusses how frequently CS research is subsequently cited by the scientific community.

Citation statistics

The 4,515 citations produced by the 143 refereed papers in this study can be summarized in a variety of second-order metrics. A simple estimate of citation output from CS projects can be obtained by dividing the total citations by the total papers, yielding 31 citations/paper. This result does not offer a fair assessment of what to expect from the average paper, however, because averages can be dominated by a small number of highly successful papers. As shown in **Table 3**, dividing the number of citations (col. 6) by the number of papers published (col. 4) yields the average number of citations per paper for each project (col. 7). Several projects generated the dominant share of the citations, while six projects generated no citations at all by the end of 2017. A better aggregate measure than the average citation rate is the median citation rate, which for this ensemble is 6 citations/paper.

A second measure also normalizes the number of publications to the number of authors and project duration.

Table 3: CS Paper Publication and Citation Rates.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Project	**Project Duration**	**Ref. Pubs**	**Median Authors**	**Total Cites**	**Citations/ Paper**	**Paper Rates**	**Citation Rates**
A	SunGrazer	22	11	2	274	25	0.25	0.57
B	RadioJove	19	0	0	0			
C	Inspire-Ionosphere	11	1	5	29	29	0.02	0.53
D	Galaxy Zoo	10	57	12	2984	52	0.48	0.44
E	Einstein@Home	10	22	20	384	18	0.11	0.09
F	GalaxyZoo-Supernova	7	5	24	150	30	0.03	0.18
G	GalaxyZoo – Mergers	9	4	6	135	34	0.07	0.63
H	Planet Hunters	6	12	21	291	24	0.10	0.19
I	Supernova Hunters	3	1	25	0			
J	Solar Storm Watch	6	7	11	43	6	0.11	0.09
K	Milky Way Project	6	4	7	134	34	0.10	0.80
L	Ice Hunters	7	1	22	9	9	0.01	0.06
M	Andromeda Project	2	1	15	10	10	0.03	0.33
N	HamSci	5	1	9	3	3	0.02	0.07
O	Aurorasaurus	5	5	3	7	2	0.33	0.09
P	Disk Detective	2	2	20	5	3	0.05	0.06
Q	Space Warps	3	4	21	40	10	0.06	0.16
R	PANOPTES	2	0	0	0			
S	Snapshot Supernova	2	0	0	0			
T	Radio Galaxy Zoo	3	3	25	14	5	0.04	0.06
U	Planet 4	1	0	0	0			
V	Moon Zoo	1	1	8	3	3	0.13	0.38
W	Backyard Worlds	1	1	12	0		0.08	
	Median values:	**5**	**4**	**12**	**29**	**10**	**0.05**	**0.09**

The idea is to compare project output in a manner that compensates for the fact that some projects have more citations and papers simply because they have been in operation for a long time. It is also advantageous in performing citation studies to "level the playing field" so that a paper with a large author list does not yield a higher impact for that reason alone. This can be accomplished by dividing column 4 by the product of column 3 and column 5, resulting in the "Paper Rates" shown in column 8. For example, the *SunGrazer* project resulted in 11 refereed publications written by a median of two authors during its 22-year project duration. The normalized paper rate, Rp, for this project is then Rp = 11/(22 * 2) = 0.25 paper/year/author, which appears in column 8. This results in a median annual paper rate for the CS projects of Rp = 0.05 papers/year/author. Similarly, the normalized citation rate is defined by dividing the total citations (col. 6) by the product of the number of published papers (col. 4), the median number of authors (col. 5), and the project duration (col. 3). For example, the *SunGrazer* project generated 274 citations from 11 papers published by an average of 2 authors over the course of 22 years for Rc = 274/(22 × 11 × 2) = 0.57 citations/paper/year/author, shown in column 9. The table also shows the median values for the 19 projects with refereed papers.

The Abt (1981) study identified 326 papers published in 1961 that generated 6,070 citations after 17 years. To compare this with a modern re-calculation, the ADS was used to identify 2,479 papers published in the same journals as studied by Abt (1981), which were followed until the end of 2017 to determine the citation statics using WoS. These modern papers yielded 26,020 citations, for an average rate of 10 citations/paper and 0.6 citations/paper/year. By comparison, the 143 CS papers published over a variety of years since the start of the earliest CS project (*SunGrazer*) in 1996 (summarized in **Table 3**) generated 4,515 citations for a median aggregate rate of 10 citations/paper, and 0.09 citations/paper/year. The CS papers thus appear to be cited at a comparable median rate to the Year-2000 papers, but the annual rate at which they are cited is dramatically lower. If the aggregate citations/paper is

an indicator of how well CS papers and their results are received (e.g., impact) by the scientific community, does the low annual rate indicate that CS papers do not stimulate intense interest to cite them? Not necessarily.

Making a direct comparison between the Abt (1981), Year-2000, and CS papers is difficult, because the papers in the CS study were not all published in the same calendar year but over a staggered publication timeline beginning in 1996 and ending in 2017. Hypothetically, for a 7-year project that started in 2010, ended in 2017, and published two papers in 2011 and 2016, the citations for the second paper will be reduced by the same project duration factor of 7 years as the citations for the earlier paper. This unfairly penalizes the more recent project papers and reduces the overall project annualized citation rate. To explore this aspect of the citation comparison in more detail, instead of looking at the aggregated citation rate, this study examines the actual history of paper citations over the project timeline, which henceforth will be called the Project Citation History.

Project citation history

In typical citation studies such as the one undertaken by Abt (1981), all papers published during a single year (e.g., 1961) are followed forward in time. A similar study can be performed for the 143 CS papers by shifting the citation histories for each paper to a uniform publication Year Zero for each project. **Figure 1** shows these citation histories for the first ten years of operation, with the citations summed across all papers in each project. The resulting cumulative project citation histories are then plotted on a common annual citation scale.

Although most of the project citation profiles reach their peak within two to three years after the publication of the primary paper, one project, *SunGrazer (Plot symbol: A)*, has a significantly different citation history with a much longer tail to its distribution extending beyond the plot to a maximum of 18 years. None of these distributions are similar to the aggregate of the papers studied by Abt (1981), which reached a peak after five years followed by a slow decline extending to twenty years. This suggests that CS papers generate considerable interest much sooner than conventional papers, however, they also seem to decline in citation popularity at a faster rate.

This effect can be seen more clearly in **Figure 2**, where the CS citation history for all projects is compared to the Abt (1981) and the Year-2000 publication cohorts. The number of citations each year has been normalized by the total number of papers in each group to determine the annual citations per paper. What is immediately obvious is that CS papers in the aggregate have a higher annual citation rate than conventional papers published in 1961 or 2000, implying that the historical context of these papers is less of an issue than, apparently, their origin in traditional versus CS research activities. Citizen science research results, when appearing in refereed publications, do indeed seem to burn more brightly and fade more quickly in interest compared to traditional research results. Also interesting is that after citation Year 10, the modern paper rates are half the 1961 rates, and are also at least double the CS paper citation rates. Evidently, modern papers represented by the Year-2000 cohort experience systematically fewer long-term citations than older papers, but consistently exceed the citation rates of the roughly contemporaneous CS papers. The time at which the CS papers fall below both the older and the modern papers is roughly 8–10 years after publication. This is probably a useful marker for CS project developers, which defines a maximum interest lifespan for CS papers relative to non-CS papers.

The previous discussion described how CS citations follow the normal publication year profiles of research papers. An interesting subsidiary question is how citations follow the project year. This would be similar in spirit to

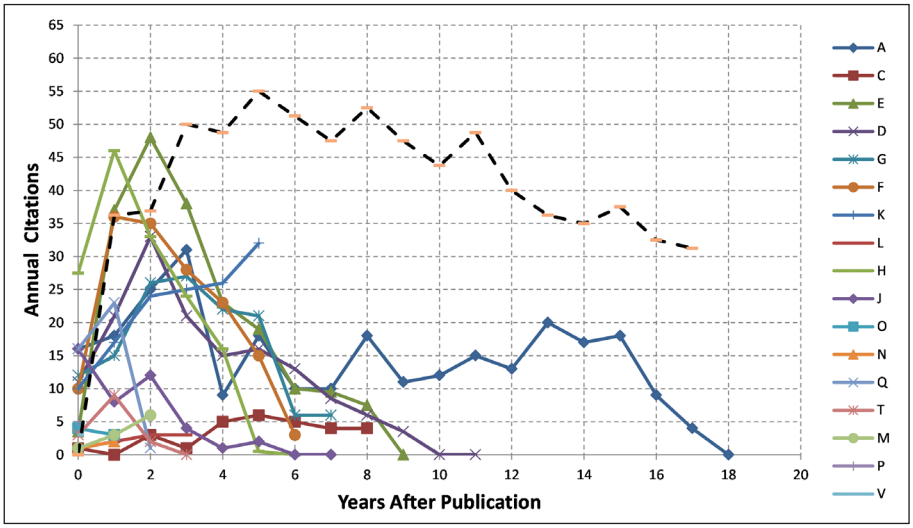

Figure 1: Citation histories for each project shifted to the start of the first citation year. The legend gives the symbol identifications in terms of the A-W project entries in Table 3. For plotting purposes, the history values for the three largest projects (D: *Galaxy Zoo*, E: *Einstein@Home*, and H: *Planet Hunters*) have been reduced by a factor of 2, 2, and 20 respectively, to focus attention on their profile shapes. The dashed line is the aggregate citation history by Abt (1981) for the 326 papers published in 1961 but reduced by a factor of eight.

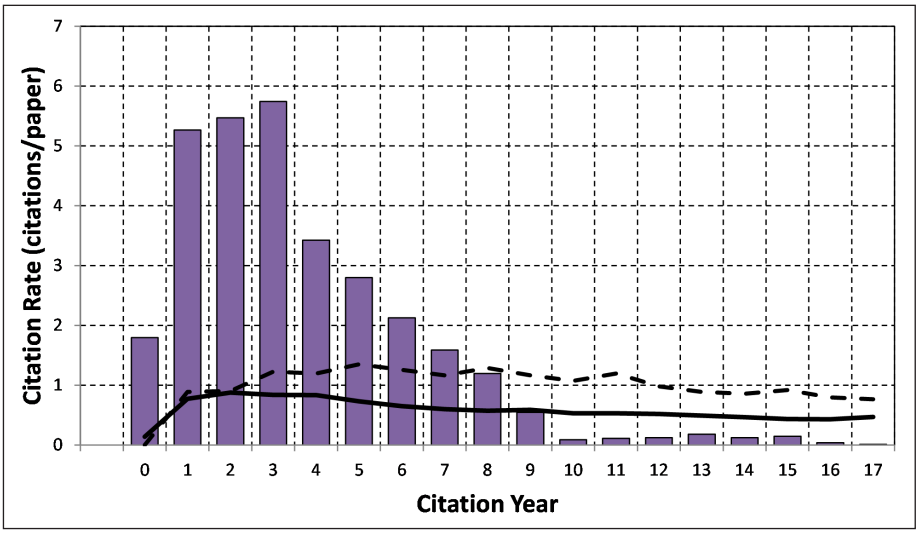

Figure 2: A normalized comparison of the modern Year-2000 citation histories (N = 26,020, solid line) with an earlier Year-1961 study by Abt (1981: N = 326, dashed line) and the normalized citation rate for the current CS sample (N = 143, bars) obtained by summing over the individual project profiles in Figure 1.

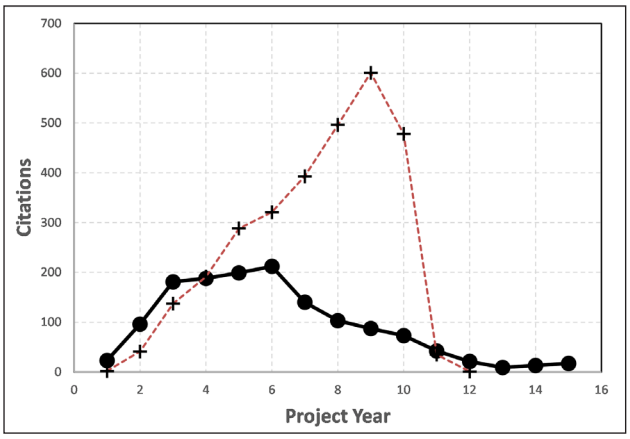

Figure 3: Citation history of the 86 papers for the 22 published CS programs that generated 1,531 citations (solid line) and the 57 papers from *Galaxy Zoo* (dotted line), which contributed 2,984 citations.

typical paper citation studies that begin in the year of paper publication such as the Abt (1981) study beginning in 1961, or the previously mentioned modern study for all papers published in the year 2000. For example, a project starting in 2005 has a paper published in 2008 that produces 10 citations in 2010. The 10 citations are counted in the tally for Project Year 4, and this calculation is re-performed for all papers in the project to create the Project Citation History. Taken as an ensemble, the annual citation counts can be aggregated for all 143 papers by representing the citations in terms of the Project Year. This effectively treats a CS project as a single meta-paper published in the project's inception year and followed to 2017. The individual Project Citation Histories are shown in **Figure 3**. The citations for *Galaxy Zoo* have been plotted separately. The sharp drop off in citations for *Galaxy Zoo* in Years 10 and 11 corresponds to citations recorded in the calendar years 2017 and 2018, for which the later year is incomplete. The smaller drop between Year 9 and 10 is, however, significant because the citation tallies for the corresponding years 2016 and 2017 are believed to be complete, suggesting that this project may have reached its peak citation year around 2016, Project Year 9.

Apparently for the non-*Galaxy Zoo* papers, citations can be expected to increase to a maximum near Project Year 5, and then decline almost monotonically thereafter. This profile is similar to the aggregate of the papers studied by Abt (1981), which also reached a peak after five years followed by a slow decline. This indicates that the aggregate citation output for individual CS projects as a function of their project year strongly resembles the citation profile for average non-CS research papers.

Alternative metrics

In 1994, the Institute of Physics (IOP) became the first publisher to publish a journal on the World Wide Web. The collection of IOP-accessible journals now spans 75 titles. Because the number of "views" and "downloads" to articles in the online IOP journals could be tallied easily each month, indices for "impact" were soon developed as an adjunct to the more traditional paper citation studies. The expansion of these Internet and social media-based indicators of article popularity and impact, collectively called "altmetrics," has led to the creation of a number of real-time services such as Altmetric.com (2011), which offers an on-the-spot index of an article's popularity via an embedded app. *Altmetric* (Altmetric.com 2018) gives the WoS index, but also aggregates all online and non-technical references including "tweets" and downloads of articles, mentions in blogs, and mentions in the news media in its Altmetric Attention Score. In this score, mentions in the news media and blogs received most of the weighting compared to mentions in Twitter, Facebook, or YouTube (Huang, Wang, and Wu 2018).

The previous discussions examined, in detail, how the citation data for WoS used in the current CS study compares with the other citation index services: iopscience.com, the ADS Database, ScienceDirect.com, Institute of Physics (IOPscience), and Altmetric.com. The results for an

Table 4: Comparison of citation statistics for *Planet Hunters*.

(1)	(2)	(3)	(4)	(5)
Paper	**WoS**	**ADS**	**iopScience**	**Altmetric**
2013a	86	96	73	257
2012a	47	49	–	64
2016a	33	37	–	2139
2012b	23	21	18	13
2013b	23	26	22	102
2014a	17	21	17	183
2014b	17	17	17	73
2015	14	16	10	1
2014c	12	9	–	–
2013c	11	18	11	1
2016b	5	6	6	0
2013d	0	0	0	0

example CS project, *Planet Hunters*, are shown in **Table 4**. For uniformity, only citations appearing through the end of 2017 are included. Not included were citations to arXiv, BAAS, and RNAAS. A dashed line indicates that the journal was not covered by the citation service. Column 1 is the paper publication year. Columns 2–5 are the total number of citations for each paper from the year of publication to 2017.

The citation statistics for the WoS, ADS, and iopScience follow a nearly linear correlation as expected, because these services generally refer to the same body of research articles and employ the same bibliographic citation measures. However, the *Altmetric* scores are not obviously related to the bibliographic citation statistics. In one case a paper, 2013a, generated between 73 to 96 citations and an *AAS* of 257, while 2016a generated between 33 to 37 citations and a dramatically higher *AAS* of 2,139. Similar disparities can be found among the other papers considered, suggesting that it is not a simple matter (linear transformation) to relate the Altmetric.com *AAS* score to a paper's placement in the research journals, which in the past (e.g., Abt 1981) has been considered a good gauge of scientific impact.

Along with the growing popularity of altmetrics for assessing article popularity has come intense scrutiny of exactly how to interpret these new metrics as compared to older forms of measure such as paper citations. According to Bastow, Dunleavy, and Tinkler (2014), canonical bibliographic measures such as citations are best interpreted as a measure of the value of a published research article to scientists, while altmetrics are primarily a measure of a research article's impact and benefit to society and social discourse. Bibliometricians such as Bornmann (2014) are increasingly seeing altmetrics as a supplement to citation counting. The relationship between the newer and older citation indices remains murky, largely because of the very different databases being sampled to form the index. Citations are generally targeted at professional research activity, while altmetrics include non-scientists, students, teachers, and many other groups. Also, while "gaming" citation counts is difficult except through self-citation, counts for altmetrics can be bought or even generated by robots. Also, the number of journals that can provide citations remains relatively fixed over time, while the number of websites, social media platforms, and data providers changes almost annually for altmetrics tabulations. A sudden change in index value may be interpreted as a change in social interest; alternatively, it may show that a provider entered or left the data pipeline. Correlations between altmetrics scores and citations have been found by Evans and Krauthammer (2011) for Wikipedia and by Thelwall et al. (2013) for Twitter. However, the degree of correlation between WoS citation rates and AAS varies significantly from journal to journal and, as discussed by Huang, Wang, and Wu (2018), in most cases no clear relationship exists between them, which was also found for the CS papers.

Citation biases

Although citation studies are becoming widely used to evaluate author productivity in both research and academic settings, using them appears to present a variety of biases. For instance, Letchford et al. (2015) found that brief titles for research papers garner more citations than longer titles. King et al. (2017) found that men cite their own research more often than women researchers. According to Fowler and Aksnes (2007), self-citation accounts for about 10% of a paper's citation history. Even the number of references in a paper's bibliography has some impact on citations, according to Vieira and Gomes (2017).

Among the other biases examined in this study was whether the number of papers produced by a project, or the number of ensuing citations, was in any way related to the number of co-authors in the typical paper. Such correlations have been found by, for example, Vieira and Gomes (2009). The CS papers offer an enormous range, by a factor of nearly 200, in the number of co-authors listed on each paper. For example, one project (*Einstein@Home*) produced several papers with more than 500 co-authors (e.g., the LIGO Consortium), while many of the smaller CS projects involve fewer than four co-authors. The median number of authors for each project is presented in **Table 2** column 7. Comparing the number of authors in column 7 with the number of refereed papers in column 6 shows no clear correlation, suggesting that only a small number of authors may be responsible for the bulk of the publications from the groups affiliated with each CS project. In terms of citations, one might presuppose that projects with large teams of authors have greater opportunities for the CS research to be cited in subsequent team publications. However, this expected trend is not borne out by the citation data. For example, one project had a median of twenty authors per paper and generated 324 citations, while a second project had a median of only one author per paper yet generated a comparable 239 citations. This large variation in authors-per-paper suggests that the analysis in the section on paper publication rates, above, normalized by the median number of authors leading to a dramatically lower median estimate for CS citations/year/paper/astronomer (e.g.,

0.09 c/y/p/a), is not a reliable indicator of a project's productivity compared to other non-CS surveys.

Self-Citations

An important issue to consider when dealing with papers in which a small number of citations are involved is "self-citation." Researchers commonly cite their own relevant research when writing new papers, and in the novel field of citizen science, self-citation of previous CS publications would be expected. Moreover, as shown by Aksnes (2003), the larger the author list on a given paper, the more often self-citation can be expected to occur. According to Shema (2012), some journals even encourage authors to cite other articles in the same journal, presumably to increase that journal's impact score. According to Fowler and Aksnes (2007), within the general population of non-CS research papers, the self-citation rate is about 11%, and the incentive for self-citing at rates as high as 20% is free from penalty. For example, if the author is being evaluated for tenure, self-citations are not counted against the author as superfluous. Also, not all self-citations are made merely to "pad" a bibliography. If an author is opening up a new field or scientific line of inquiry, self-citations are necessary in a positive way. Among the 4,515 citations in this study of 143 CS papers, about 840 were self-citations, just over an aggregate average of 18%. These included citations to papers written by other members of the project team, or to previously published findings from the project. The distribution of self-cited papers with respect to the total citations for all 143 papers is shown in **Figure 4**.

The self-citation rates can be binned into three groups based on the total paper citations shown in **Figure 4**. Group A, with between 1 and 10 citations, displayed a wide range of self-citation rates between 0% to 100%. Group B, with 11 to 100 citations, had an average of 20% ± 10%. Group C with >100 citations had an average self-citation rate of 10% ± 10%. This suggests that self-citation among the refereed journal articles for CS is a large issue, especially among papers that yield only small numbers of citations during their publication lifetimes. In these cases, the papers are being cited only by other CS papers, often within their own project, rather than by a larger base of non-CS investigators. However, because the median number of self-citations per paper is 4, papers in Groups B and C that garner more than 40 citations achieve a balance that is similar to the 11% self-citation rates seen in the general population of papers as identified by Fowler and Aksnes (2007). This suggests a citation threshold at approximately Group B, in which CS papers are being more widely cited by other research communities than by the CS project members.

Heavily-cited papers

What can be said about the proportion of projects that generate the largest numbers of citations in the literature? Based on the tallies in **Table 2** of 19 projects with refereed publications, the median citations per published CS research paper is 10 citations/year. The top four projects with papers that cumulatively exceeded 200 citations (upper 21% of projects) account for 3,933 citations, while projects in the lower 37% account for 37 citations. All other things being equal, the odds that any given CS project will exceed 200 citations of its research in its roughly 6-year average lifetime are about one in five.

Among the 58 million papers indexed by the *WoS*, only 14,500 papers (one-in-3,800) achieved more than 1,000 citations by 2014 (van Noorden, Maher, and Nuzzo 2014). Space science and astronomy topics are a small subset that is drowned out by the far more numerous papers in physics, chemistry, and biology. For space science papers in particular, Pearce (2004) examined citation impact statistics for the 1,000 most-cited papers out of the 439,000 papers published previous to November 2003 that had at least 257 citations by that time. After five years, one paper in 1,000 had reached 253 cumulative citations. For the current CS study, 77 CS papers were found that had been published for five years or longer to match the Pearce (2004) study baseline. These CS papers generated

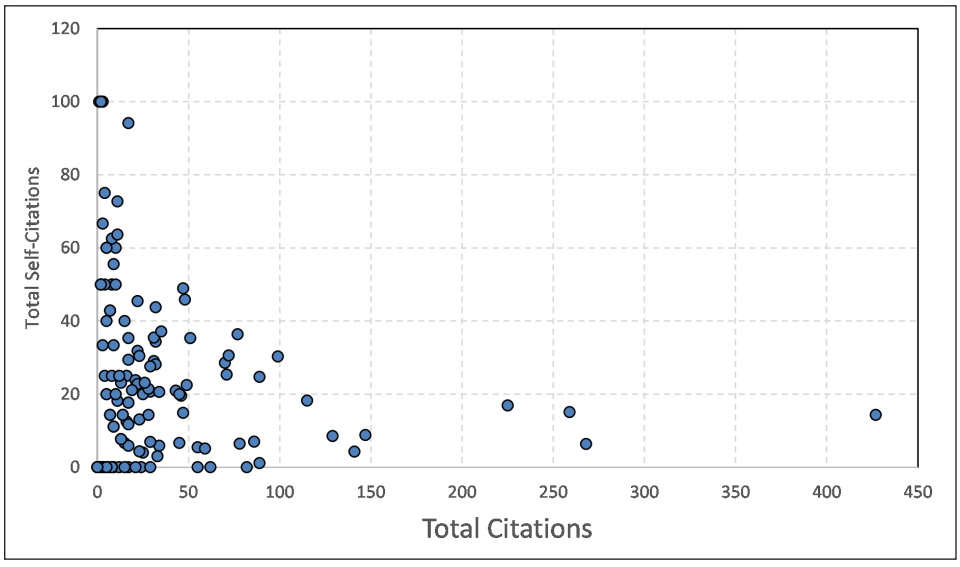

Figure 4: Self-citations for the CS papers in this study. The total number of self-citations to the 143 CS papers is compared to the total number of all citations to each paper.

3,884 citations (87% of the total), with one paper out of 26 having more than 253 citations after 5 years. Papers presenting CS research are significantly more likely than traditional research papers to reach the ranks of the Top-1000 papers in space science and astronomy.

Conclusion

A total of 143 publications in refereed journals resulting from 23 CS projects in space science or astronomy have been investigated for their citation histories. The projects generated a median of two papers during their average of 6 years of activity. The citation history profiles show a marked trend to peak within 2–3 years after paper publication but decline thereafter at a faster pace than the average science paper published after 2000. Moreover, CS research papers do not follow the long citation tail seen in typical science papers published in 2000, but experience a sharp drop-off by year 8 after publication. This pattern suggests that as an aggregate population, CS papers have significantly higher peak citation rates—as much as a factor of four times higher than non-CS papers—and remain of interest for about half the time of non-CS papers. The apparent fact that CS papers "burn brighter" and last half as long as non-CS papers may partially involve the somewhat higher self-citation rates for CS papers that generate fewer than 40 citations, or in the novelty of the CS approach for conducting research, which leads to more reader interest in these types of papers. Meanwhile, the citation rate does not seem to depend on other features of the initial papers such as the number of co-authors, although as a population, CS papers have significantly more co-authors than non-CS papers. In terms of overall ranking, the proportion of CS papers surpassing 200 citations is higher than for the typical paper published in 2000. Some CS projects lead to papers that compare well with some of the most highly ranked "Top-1000" research papers in astronomy and space science that were cited more than 200 times. Although about half of all CS projects in space science and astronomy may not have a publication record after ten years, this pattern resembles that found at a variety of major ground-based observatories.

In concluding the 1981 study, Helmut Abt commented on the impact that citation studies can have on legitimate research projects, and his words are worth reflecting upon: "... *We should remember that the reason for doing astronomical research is to learn important facts about the universe, not to produce citations. The two goals are roughly correlated, but each of us must use our judgment as to what is needed. Sometimes that judgment tells us to do a project even if it is not likely to be cited frequently.*" Some citizen science projects are as much, if not more, about introducing the public to the hard work and thrilling discoveries of hands-on scientific exploration as they are about developing publishable work. Nevertheless, this study shows that CS projects are not only as good as conventional day-to-day research projects in generating publishable results, but appear to significantly out-perform the citation rates of typical non-CS papers published in the year 2000.

Supplementary Files

The supplementary file for this article can be found as follows:

- **Citation Data and Analysis spreadsheets.** These spreadsheets contain the original list of 143 publications and their annual citation counts, in addition to a variety of worksheets containing the calculations that form the basis for the figures and tables in this study. DOI: https://doi.org/10.5334/cstp.152.s1

Acknowledgements

I would like to thank Dr. Marc Kuchner (NASA Goddard Citizen Science Working Group) for his many helpful comments and suggestions, which significantly improved the clarity and accuracy of this work. I also thank the two anonymous referees and the editor of the journal, whose thoughtful comments led to additional work on self-citations, assessing the use and interpretation of altmetrics more fully, and improving the clarity of the discussion. This work was supported by the NASA Space Science Education Consortium through ADNET contract SESDA-IV (80GSFC17C0003).

Competing Interests

The author has no competing interests to declare.

References

Abt, H. 1981. Long-Term Citation Histories of Astronomical Papers. *Publications of the Astronomical Society of the Pacific*, 93: 207–210. DOI: https://doiorg/10.1086/130806

Abt, H. 1998. Is the Astronomical Literature Still Expanding Exponentially? *PASP*, 110: 210–213. http://iopscience.iop.org/article/10.1086/316123/pdf accessed May 2018.

ADS. 2018. SAO/NASA Astrophysical Data System. http://adsabs.harvard.edu accessed December 2018.

Aksnes, DW. 2003. A macro study of self-citation. *Scientometrics*, 56(2): 235–246. DOI: https://doi.org/10.1023/A:1021919228368

Altmetric.com. 2011. About US. https://www.altmetric.com/about-us/ accessed May 2018.

Altmetric.com. 2018. Altmetric: Sources of Attention. https://www.altmetric.com/about-our-data/our-sources/ accessed May 2018.

Bastow, S, Dunleavy, P and **Tinkler, J.** 2014. *The Impact of the Social Sciences*. London, UK: Sage.

Bornmann, L. 2014. Do altmetrics point to the broader impact of research?: An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics*, 8(9): 895–958. https://arxiv.org/ftp/arxiv/papers/1406/1406.7091.pdf accessed May, 2018. DOI: https://doi.org/10.1016/j.joi.2014.09.005

Cavalier, D. 2014. SciStarter. https://scistarter.com/about accessed 12/14/2017.

Citizen Science Alliance. 2007. https://www.citizensciencealliance.org/projects.html accessed 12/14/2017.

Clery, D. 2018. Still working: Astronomers explain why they don't publish. http://www.sciencemag.

org/news/2018/02/still-working-astronomers-explain-why-they-don-t-publish accessed August 23, 2018.

De Solla Price, DJ. 1965. Networks of Scientific Papers. *Science*, 149: 510–515. DOI: https://doi.org/10.1126/science.149.3683.510

Evans, P and **Krauthhammer, M.** 2011. Exploring the use of social media to measure journal article impact. In: *AMIA Annual Symposium Proceedings*, 374. American Medical Informatics Association.

Fowler, JH and **Aksnes, DW.** 2007. Does self-citation pay? *Scientometrics*, 72(3), 427–437. DOI: https://doi.org/10.1007/s11192-007-1777-2

Hennekin, EA. 2012. Publication Trends in Astronomy: The Lone Author. https://arxiv.org/abs/1202.4646v1.

Huang, W, Wang, P and **Wu, Q.** 2018. A correlation comparison between Altmetric AttentionScores and citations for six PLOS journals. *PLoS ONE*, 13(4): e0194962. DOI: https://doi.org/10.1371/journal.pone.0194962

King, MM, Bergstrom, CT, Correll, SJ, Jacquet, J and **West, JD.** 2017. Men Set Their Own Cites High: Gender and Self-citation Across Fields and Over Time. *Socius: Sociological Research for a Dynamic World*, 3: 1–22. DOI: https://doi.org/10.1177/2378023117738903

Kosmala, M, Wiggins, A, Swanson, A and **Simmons, B.** 2016. Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, 14: 551–560. DOI: https://doi.org/10.1002/fee.1436

Letchford, A, Moat, HS and **Pries, T.** 2015. The Advantages of Short Paper Titles. *Royal Society Open Science*, 2. DOI: https://doi.org/10.1098/rsos.150266

Pearce, F. 2004. Citation measures and impact within astronomy. *Astronomy & Geophysics*, 45(2): 2.15–2.17. 1 April 2004. DOI: https://doi.org/10.1046/j.1468-4004.2003.45215.x

Sauermann, H and **Franzoni, C.** 2014. Crowd science user contribution patterns and their implications. *Publications of the National Academy of Sciences*, 112(3): 679–684. http://www.pnas.org/content/112/3/679 accessed 12/14/2017.

Science Citation Index. 1962. Philadelphia, Pennsylvania: Institute for Scientific Information, Inc.

Scientific American. 2011. Education: Citizen Science. https://www.scientificamerican.com/citizen-science/?page=1 accessed 12/14/2017.

Shema, H. 2012. On Self-Citation. *Scientific American*. https://blogs.scientificamerican.com/information-culture/on-self-citation/ accessed May 2018.

Smith, KT. 2018. Why don't astronomers publish observations. *Science*, 359: 1005–1006. http://science.sciencemag.org/content/359/6379/1005.3. DOI: https://doi.org/10.1126/science.359.6379.1005-c

Thelwall, M, Tsou, A, Weingart, S, Holmberg, K and **Haustein, S.** 2013. Tweeting links to academic articles. *Cybermetrics*, 17(1). http://cybermetrics.cindoc.csic.es/articles/v17i1p1.html.

The National Geographic. 1996. Citizen Science Projects. https://www.nationalgeographic.org/idea/citizen-science-projects/ accessed 12/14/2017.

Van der Kruit, PC. 2004. Citation Analysis and the NASA Astrophysics Data System (ADS). www.astro.rug.nl/~vdkruit/jea3/homepage/ads.pdf accessed 12/14/2017.

Van Noorden, R, Maher, B and **Nuzzo, R.** 2014. The Top 100 Papers. *Nature*, 514: 550–553. DOI: https://doi.org/10.1038/514550a

Vieira, ES and **Gomes, JANF.** 2009. Citations to Scientific Articles: Its Distribution and Dependence on the Article Features. *Journal of Informetrics*, 4: 11–13. DOI: https://doi.org/10.1016/j.joi.2009.06.002

Watson, D and **Floridi, L.** 2016. Crowdsourced science: sociotechnical epistemology in the e-research paradigm. *Synthese*, 195: 741–764. DOI: https://doi.org/10.2139/ssrn.2914230

Wilson Center. 2017. Federal Crowdsourcing and Citizen Science Catalog. https://ccsinventory.wilsoncenter.org/ accessed 12/14/2017.

How to cite this article: Odenwald, S. 2018. A Citation Study of Citizen Science Projects in Space Science and Astronomy. *Citizen Science: Theory and Practice*, 3(2): 5, pp. 1–11. DOI: https://doi.org/10.5334/cstp.152

Submitted: 08 February 2018 **Accepted:** 17 July 2018 **Published:** 26 October 2018