How to Open Science: Analyzing the Open Science Statement Compliance of the Learning @ Scale Conference

Aaron Haim
ahaim@wpi.edu
Worcester Polytechnic Institute
Worcester, Massachusetts, USA

Chris Baxter*
Robert Gyurcsan*
wcbaxter@wpi.edu
rfgyurcsan@wpi.edu
Worcester Polytechnic Institute
Worcester, Massachusetts, USA

Stacy T. Shaw
Neil T. Heffernan
sshaw@wpi.edu
nth@wpi.edu
Worcester Polytechnic Institute
Worcester, Massachusetts, USA

ABSTRACT

There have been numerous efforts documenting the effects of open science in existing papers; however, these efforts typically only consider the author's analyses and supplemental materials from the papers. While understanding the current rate of open science adoption is important, it is also vital that we explore the factors that may encourage such adoption. One such factor may be publishing organizations setting open science requirements for submitted articles: encouraging researchers to adopt more rigorous reporting and research practices. For example, within the education technology discipline, the *ACM Conference on Learning @ Scale* (L@S) has been promoting open science practices since 2018 through a Call For Papers statement. The purpose of this study was to replicate previous papers within the proceedings of L@S and compare the degree of open science adoption and robust reproducibility practices to other conferences in education technology without a statement on open science. Specifically, we examined 93 papers and documented the open science practices used. We then attempted to reproduce the results with invitation from authors to bolster the chance of success. Finally, we compared the overall adoption rates to those from other conferences in education technology. Although the overall responses to the survey were low, our cursory review suggests that researchers at L@S might be more familiar with open science practices compared to the researchers who published in the *International Conference on Artificial Intelligence in Education* (AIED) and the *International Conference on Educational Data Mining* (EDM): 13 of 28 AIED and EDM responses were unfamiliar with preregistrations and 7 unfamiliar with preprints, while only 2 of 7 L@S responses were unfamiliar with preregistrations and 0 with preprints. The overall adoption of open science practices at L@S was much lower with only 1% of papers providing open data, 5% providing open materials, and no papers had a preregistration.

All openly accessible work can be found in an Open Science Framework project[1].

*The second and third authors contributed equally to this work.
[1]https://doi.org/10.17605/osf.io/pj3te

This work is licensed under a Creative Commons Attribution International 4.0 License.

L@S '23, July 20–22, 2023, Copenhagen, Denmark

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0025-5/23/07.
https://doi.org/10.1145/3573051.3596166

CCS CONCEPTS

• **General and reference** → **Reference works**; **General conference proceedings**; *Validation*; *Reliability*; *Verification.*

KEYWORDS

Open Science, Peer Review, Reproducibility, Statement Compliance

ACM Reference Format:
Aaron Haim, Chris Baxter, Robert Gyurcsan, Stacy T. Shaw, and Neil T. Heffernan. 2023. How to Open Science: Analyzing the Open Science Statement Compliance of the Learning @ Scale Conference. In *Proceedings of the Tenth ACM Conference on Learning @ Scale (L@S '23), July 20–22, 2023, Copenhagen, Denmark.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3573051.3596166

1 INTRODUCTION

Efforts towards increasing the adoption of open science and providing robust reproducibility had steadily increased since 2010 and into the current decade [33, 35]. Numerous research studies have been carried out on the effects of open science practices, such as whether open access papers affect the number of citations received [10], or how open science affects intellectual property rights [30]. However, there are numerous factors that have not been investigated. One such factor involved publishing organizations encouraging or settings requirements on submitted articles. The *ACM Conference on Learning @ Scale* (L@S) was one such conference where the website contained a 'Statement on Open Science', educating researchers on some of the available practices.

Previous research has focused on the adoption of open science and the robustness of reproducibility as conducted on the *International Conference on Learning Analytics and Knowledge* (LAK) [7, 21], the *International Conference on Artificial Intelligence in Education* (AIED) [8], and the *International Conference on Educational Data Mining* (EDM) [9]. Findings showed around a 5% adoption rate of open science practices needed for reproducibility, with only 1 paper producing the exact results reported and 2 papers producing results within the reported confidence interval. The previous works also collected responses from authors about their papers, typically providing 55%-65% increase in reports that the dataset or source could be made available for reproducibility. However, the conferences themselves did not encourage any open science or reproducibility practices, while L@S did.

The goal of this work was to compare the adoption of open science and reproducibility within the L@S conference to previous works published in other conferences. Specifically, this work would replicate the results of the previous works across papers published

within the 9th and 8th L@S conferences. Trained reviewers examined each paper for any open science and reproducibility practices that were adopted. In addition, we reached out to authors to ask them about why they did or did not adopt open science practices. We combined the collected information to run extensive, documented reproducibility tests. Finally, we compared the results obtained from this conference to those from prior conferences at LAK, AIED, and EDM to view the effect of L@S encouraging researchers to adopt open science practices.

Specifically, this work aimed to answer the following questions (henceforth referred to as Research Questions, or RQ):

(1) How many of the papers within the proceedings of the *ACM Conference on Learning @ Scale* (L@S) adopted open science practices and associated subcategories of interest?
(2) For L@S papers, what reasons do authors give for use or non-use of open science practices?
(3) Can the reported results of the papers be reproduced within a 6 hour timeframe, fixing any issues necessary?
(4) Do conference policies around open science show evidence for increased adoption compared to similar conferences without open science policies (e.g. AIED, EDM)?

2 BACKGROUND

2.1 Open Science

Open Science describes a set of principles or practices aimed toward improving transparency, reproducibility, and replicability in science. Practices of open science include all levels of scientific research, including methodologies, data, software, outcomes, and publications, with the goal of creating open and accessible research content to all regardless of societal standing [16, 33]. Numerous subcategories of open science had been created before the unifying term was created. Open science, as a term, only started to gain traction in the early 2010s [33]. During this time, researchers were discovering numerous issues when attempting to replicate, or even reproduce, other researchers' work: ambiguous methodologies on how the experiment was run, materials used incorrectly or overused within the same study, numerous replication failures when attempting to produce the results, etc. Open science provided a simple way to mitigate these issues by developing standards, practices, and common recommendations. Then during the mid-2010s, large-scale studies in psychology [4] and other disciplines [2] began casting doubt on the reproducibility and replicability of previous research. As a result, numerous disciplines began adopting open science practices to make future research more robust and provide greater transparency.

It is important to note that our investment in documenting the adoption of open science and robustness of research began from our own shortcomings in reproducibility. In our lab, we were unable to reproduce research we conducted only a few years ago. Only after communicating with the prior student were we able to obtain the necessary setup instructions to execute the analysis and documentation to understand the reported results. This event motivated us to do a better job increasing transparency and improving the robustness of our own research. Admitting that we too had room to improve, the current work aims to 1) investigate the current adoption rate of open science, 2) survey authors for their own understanding and adoption of such practices, 3) attempt to reproduce their work and diagnose and available issues, and 4) determine the effect of an open science statement on the Call For Papers (CFP) of a conference. The overall goal of this project is to identify ways to promote open science adoption and reproducibility in our field.

2.2 Learning at Scale

Learning at Scale is defined as "the study of the technologies, pedagogies, analyses, and theories of learning and teaching that take place with a large number of learners and a high ratio of learners to facilitators" [31]. At-scale systems had collected large amounts of data to develop numerous features, models, and reports to improve student learning and provide better tools for teachers. Since 2014, the *ACM Conference on Learning @ Scale* (L@S) had been one such conference publishing information provided by these large-scale systems. The systems developed and the data provided generally conformed, at least internally, to open science practices. As such, this work used papers published in the proceedings of the L@S conference as the dataset to analyze.

3 METHODOLOGY

3.1 Open Science Peer Review

To accomplish RQ1, we adopted the methodology from the previous works [7–9]. We evaluated every research article, short paper, and poster published in the proceedings of the previous two years: the *9th ACM Conference on Learning @ Scale* [19] and the *8th ACM Conference on Learning @ Scale* [18]. Only the last two years were considered as reproducibility and resources generally become more difficult to execute and/or locate respectively the older the work is. L@S publishes conference proceedings to the Association for Computing Machinery Digital Library (ACMDL). As such, a Hypertext Transfer Protocol (HTTP) request was used to query the associated papers[2]. As the request queried all papers from L@S in decending order, only those meeting the mentioned requirements were used. A Qualtrics[3] survey was used to mark which open science practices were adopted and whether a paper was completely reproducible. Each paper in the survey was uniquely identified and recorded by its Digital Object Identifier (DOI). The survey also recorded the corresponding author's email[4], the proceedings the paper was a part of, and the ACMDL paper classification (Research Article, Short Paper, or Poster). If the ACMDL paper classification was incorrect, it was marked as a separate entry; however, it was reviewed as though it was the ACMDL classified paper. Each paper review was given a maximum time limit of 15 minutes to complete due to logistical constraints. This could occur when a resource was not specified in the paper and required searching the internet or through citations, for example. A supplemental document, referred to as the 'explanations document' within this work, was used to document the justifications of a certain choice made within the paper. A justification might have been omitted if it was considered

[2]The query can be obtained from the Open Science Framework project wiki: https://osf.io/pj3te/wiki
[3]https://www.qualtrics.com/
[4]The corresponding author's email was omitted from the dataset in compliance with our International Review Board (IRB) protocol.

to be redundant or self-explanatory (e.g., no link to preregistration within the paper, a source had no license). If any of the links within the paper no longer referenced the resource associated with it, the link was added to the explanations document and marked within the survey.

Open Methodology is defined as the methods and evaluation details used to conduct the research are accessible by all [16]. The methods were typically more expansive than those provided within a published paper [14, 17]. For the purposes of this research, we assumed that the methodologies provided everything necessary to reproduce the paper. The field was broken down into the ACMDL access tags: 'Public Access', 'Open Access', or 'Available' if the paper had no access tags. Additionally, if the paper was not available, it would be marked as 'No'. The 'Open Access' tag defined if a paper was freely available to anyone online [1].[5]. The 'Public Access' tag marked that a paper would become freely available one year, at the latest, after its publication[6].

Open Data is defined as the dataset(s) used within a research project are accessible by all [20, 22]. Datasets that were openly accessible typically require a license or mention of being within the public domain. As dataset licenses were typically ambiguous or complex to properly understand [13, 29], however, a dataset would be marked as open if there was a link, or a link to another paper with a link, to the dataset regardless of if a license is available. A dataset could also be marked as on request if the paper explicitly mentions that the dataset could be requested from the author. If the paper did not use a dataset, then the field was marked as non-applicable. In addition, the documentation of the dataset was included as a separate field, marking it as full if the fields of the dataset was mapped to an associated description and partial if at least one field was mapped in some location.

Open Materials is defined as technologies – such as open source software [11, 28], freeware, or non-restrictive services – are accessible to all. For this work, a paper had open materials if a link, or a link to another paper with a link, to all materials and the source the authors used were provided. If the paper did not make the source code public but mentioned using at least one public material, then the paper would be marked as having 'partial' open materials. If the paper did not use a material, then the field was marked as non-applicable. This field could be expanded to encompass commercial products; however, there was no prior literature as of the writing of this paper, so while mentioned in the explanations document, they were excluded from the marking on the survey. Three additional subfields of open materials were recorded within the survey: the documentation which provided understanding on how to use the materials and source [5], the README which contained information on the source along with some setup instructions [15], and the license which marked whether the source can be used openly [6, 28, 32].

A **preregistration** for a paper describes the process or methodology to be conducted before the actual research takes place typically to prevent p-hacking or hypothesizing after the results were known [24, 25, 34]. If a preregistration needed to be updated, then a new preregistration could be created to preserve the development of the research project. If a paper had a link to some location hosting the preregistration (e.g., Open Science Framework[7], AsPredicted[8]), then the paper was marked as such. If a preregistration was deemed to be unnecessary, such as for a theoretical or argumentative paper, the field was marked as non-applicable.

Expanding upon past efforts, we additionally looked at preprints. A **preprint** is a paper that is published in an independent repository prior to formal peer review or publication in a conference or journal [3, 12]. Preprints have numerous uses such as quick availability, trackable development cycles, and a DOI. If a paper had a link to some location hosting the preprint (e.g., ArXiv[9], EdArXiv[10]), then the paper is marked as such.

3.1.1 Undergraduate Interpretation. Similar to the EDM work [9], the peer review was conducted by two trained undergraduate research assistants, known as 'Reviewers', supervised by the lead of this work, known as the 'Meta-Reviewer'. To mitigate potential errors during the review process, each reviewer was randomly assigned ten papers reviewed by the other reviewer and provided a second review. If the reviewers disagreed or were unsure on a particular paper, the meta-reviewer was responsible for providing feedback and the correct status assigned to the paper. As an addition step, the meta-reviewer also provided a quick double-check to find any major errors in the provided reviews.

3.2 Author Survey

To accomplish RQ2, authors of the papers being reviewed were surveyed to provide input on their papers [8, 9]. For each paper published within the two L@S conferences, an email was sent to the corresponding author[11]. Authors with multiple papers published in the proceedings were sent a single email containing each paper they should complete the survey for instead[12]. To encourage authors to respond to the survey, a separate email was sent to all participants to provide a quick overview on what the survey was and where it would be sent from. The survey responses were publicly released and linked by the DOI for the paper as stated in our International Review Board (IRB) study. Any author information provided was removed from the released dataset. The survey itself was sent on November 29th, 2022 and continues to collect responses. This work reported on responses collected up to January 3rd, 2023.

The survey requested the name and email of the author along with the DOI of the associated paper. The content of the survey was separated into six subsections:

- Data,
- Materials,
- Preregistration,
- Preprint,
- Reproducibility, and
- Resource Degradation.

[5]Learn more about open access through ACMDL on their website: https://www.acm.org/publications/openaccess

[6]In the United States, this typically included all publications from the United States National Institutes of Health (NIH): https://publicaccess.nih.gov/

[7]https://osf.io/registries

[8]https://aspredicted.org/

[9]https://arxiv.org/

[10]https://edarxiv.org/

[11]The first author was assumed to be the corresponding author if none was mentioned.

[12]This email survey was conducted in parallel with two separate research projects for other conferences to mitigate the issues mentioned above. The other research projects will be reported at their associated conferences.

The data section collected information on the dataset, its documentation, and its license. The author reported on whether:

- the dataset was publicly available,
- the dataset could be shared on request,
- the dataset could not be shared, or
- an 'other' option with a text box if the answer did not fit the previous categories.

If the dataset was not publicly accessible, the author was asked to explain their reasoning behind the choice. If the dataset could be provided, whether publicly or on request, the author was asked to provide a link to its location. The link was removed from the public dataset if on request in case sensitive information was shared. All questions are shown to the author of the paper for full transparency.

The materials section collected information on the materials, source, source documentation, and source license. The questions in this section are the same as those within the data section with the data keywords replaced with material ones.

The preregistration section collected information on the preregistration, if applicable, of a paper. The author reported on whether:

- the preregistration was public,
- the preregistration was private,
- no preregistration existed, or
- an 'other' option with a text box if the answer did not fit the previous categories.

If the preregistration could be provided, whether publicly on privately, the author was asked to provide a link to its location. If no preregistration was created, the author was asked to explain their reasoning behind the choice.

The preprint section collected information on the preprint of a paper. The author reported on whether a preprint was created. If the answer could not easily be a yes or no, an 'other' option was provided with a text box. If the preprint was created, the author was asked to provide a link to its location. If no preprint was created, the author was asked to explain their reasoning behind the choice.

The reproducibility section collected information on instructions and materials needed to reproduce the results in the paper and potentially replicate the research in the future. The author was asked to provide any methodologies not included within the paper along with the reasoning for exclusion. In addition, the author was asked to provide any setup instructions – including file locations, software versions, setup scripts, etc. – needed to execute the source with the dataset along with the reasoning for exclusion.

The resource degradation section collected information on any resources which no longer existed as the provided location in the paper. The authors were asked to review and update any degraded resources. If any resources were degraded, the author was asked if they knew the reasonings behind the degradation.

3.3 Reproducibility

Reproducibility is defined as obtaining the exact outcomes reporter when given a static input (e.g., dataset, text file, configuration settings, etc.) and a deterministic methodology (e.g., source code, third-party software, etc.) [23, 26, 27]. To accomplish RQ3, we assumed that a paper is potentially reproducible if a paper had a publicly available or author provided dataset and source [9]. If the exact results reported in the paper could be obtained by executing the dataset and source, the paper would be considered reproducible. If the dataset or the source was not provided or found within the 15-minute timeframe of the peer survey, then the paper would be marked as not reproducible. If the paper did not use a dataset or source code, then the paper was marked as non-applicable. If the reproducibility test did not produce the exact results but was within the confidence interval reported, then the explanations document would mark the reproduction as partial.

Each paper was given a maximum limit of 6 hours to reproduce the reported results. Actions that continued executing after the 6 hour limit were automatically stopped unless the researcher conducting the reproduction, known as the tester, believed that the action could be completed within an additional hour. A timing website[13] was used to manually track how long the reproducibility test took. Additionally, the timer was broken down into three subsections – setup, execution, and debugging – to gain a better understanding of how the tester allocated time to the project. Any breaks taken by the tester during the reproduction stopped all timers and actions and recorded in the explanations document.

The setup time tracked the time taken prior to the first execution period. Any downloads and environment setup instructions were included in this time period. The execution time tracked the time taken during an action provided by the author during the execution of their program. This time was the total time on all runs of the execution, including those that failed. The debug time tracked the time taken during an action taken by the tester between executions of the program. Any issues or fixes made were reported in the explanations document. The times reported were a benchmark based on the tester's machine, so the results would most likely differ for different testers. As such, any observations made would be related to the actions taken during the associated time subsection.

All reproducibility tests ran on a big data machine within the authors' lab. The machine was known to run numerous calculations quickly and supported common performance enhancers like the CUDA Toolkit[14]. Additionally, it ran on an Unix-based operating system with a Bash shell which most researcher-provided scripts were typically executed on. The specifications of the machine are listed in Appendix A.

3.3.1 Python. If the environment needed to reproduce the source used Python[15], then the following steps were taken:

(1) If a specific version of Python was specified, download and select the version of Python.
(2) Create a empty virtual environment using 'venv'[16] and activate it.
(3) Follow any setup steps specified by the analysis.
(4) If the analysis is in a Python (.py) file:
  (a) Run the file using the 'python' command.
(5) If the analysis is in a Python Notebook (.ipynb):
  (a) Install 'ipykernel' and 'notebook' using the 'pip' command.[17]

[13] This work used https://stopwatch.online-timers.com/multiple-stopwatches
[14] https://developer.nvidia.com/cuda-toolkit
[15] https://www.python.org/
[16] This is the recommended way for Python 3; however, there are other methods to do so.
[17] If the path is improperly configured, the command may need to be prefixed with 'python -m'.

(b) Open the notebook and specify the kernel used as the one within the virtual environment.
(c) Run the notebook.

3.3.2 R. If the environment needed to reproduce the source used R[18], then the following steps were taken:

(1) If a specific version of R was specified, download and select the version of R.
(2) Create a new project using RStudio[19] or another IDE that can use 'packrat'[20][21].
(3) Follow any setup steps specified by the analysis.
(4) Run the R script.

3.4 Statement Compliance Review

In contrast to previously reviewed conferences, L@S provided an 'Open Science Statement' within its Call for Papers (CFP) which, as stated, promotes the adoption of open science practices and robust reproducibilty and follow-ups. L@S had this statement in its CFP since 2018[22] and was continually updated since then for greater clarity. In 2021, or the 8th conference, the open science statement was the following:

"Authors are encouraged to conduct their scientific inquiry using emerging best practices in open science. Authors are encouraged to preregister their study design, hypotheses, and analysis plans, and publish these using platforms such as OSF.io or AsPredicted.org. Whenever possible, feasible, and ethical, authors are encouraged to make their data, materials, and scripts openly available for inspection, replication, and follow-up analysis. The best way to share these materials is to use an established platform like OSF.io."[23]

In 2022, the statement was appended with a section on preprints:

"Authors are also encouraged to post pre-prints of their submissions with preprint hosting sites such as EdArXiv.org or on their own sites. If accepted, any preprint version should be updated to refer readers to the journal version as the document of record. The Learning@Scale steering committee supports open dissemination of knowledge as it can lead to productive exchanges, as well as earlier and greater citation of published work (See The Effect of Open Access [10])."[24]

As mentioned in the statement, the CFP encouraged open data, open materials, preregistration, and preprints. To accomplish RQ4, this work was compared to the results obtained from previous works for other conferences on the adoption of open science and robustness of reproducibility. We additionally assumed that any author who published a paper to the conference had read the Call For Papers (CFP) as the important dates for submission and acceptance were below the statement. As there was no direct translation between the different subfields of education technology, we compared the percentage of the number of papers which adopted the practice. For reference, we also provided the exact numbers along with the percentage. The results are interleaved throughout Section 4.

4 RESULTS

4.1 Peer Review

As shown in Table 1, across the 72 papers published in the 9th proceedings and the 21 papers published in the 8th proceedings, there were 41 research articles and 52 short papers. Five of the short papers were misclassified by ACMDL and were instead Work-in-Progress papers. Out of the published papers, 2 were publicly accessible, 14 were openly accessible, and the remaining 77 were available to view with membership access. Similar to the LAK work [7], no posters were published in the L@S proceedings, so they were not reviewed.

Only 1 paper, which was openly accessible, had used a dataset that was already or made openly available and was fully documented. Although some of the authors on this work were involved with the paper in question, the choice to make their data open was before this work was conceptualized. Regardless, this was much lower than the approximately 70 papers at previous conferences which had open data [7–9]. However, 19 papers, or 20%, did not use a dataset within their paper and was marked as non-applicable, or 'N/A', which was much higher than the approximately 10 papers at the other three conferences combined.

Five papers, or 5%, used materials and made the source openly available, while 10% of papers provided at least one openly accessible material. This was consistent with the results from LAK, while AIED and EDM were much higher at around 9% and 15%, respectively. Out of the 5 papers, 2 had full documentation, 4 contained a README on the source, and 3 had a permissive license. Once again, 18 papers, or 19%, did not use any materials within their paper and was marked as N/A. Totaling all the previous conferences together once again was approximately 10 papers, denoting a considerable increase.

None of the papers had a preregistration linked to it. However, 17 papers, or 18%, were marked as non-applicable as the paper was more theoretical or argumentative in nature. This was common in previous conferences, where at most 3 papers had a preregistration at EDM, with only 1 paper at LAK and 0 at AIED.

We only located 5 papers had a preprint published before the conference, all of which were on preprint servers (e.g., arXiv, EdArXiv). The preprints themselves were either the exact same as those published within the conference proceedings or contained more in-depth information. Comparing the preprint's date submission to the 'Important Dates' section in L@S 2022[25], the preprints were published after the acceptance notification was received[26]. For the preprints that contained the exact same information as the published copy, they would either had been considered post-prints, as they were published with the reviewer's comments, or the publisher's version, as they were the exact document published by

[18] https://cran.r-project.org/
[19] https://posit.co/products/open-source/rstudio/
[20] https://cran.r-project.org/package=packrat
[21] 'packrat' is the most commonly used option for managing R dependencies. It is not the only method.
[22] This statement might be inaccurate. The addition of the open science statement could have been provided in the 2017, or 4th conference, CFP; however, the associated website could not be found to confirm this.
[23] Pulled on January 30th, 2023 from https://learningatscale.hpi.de/index.php/call-for-papers/ls/author-guidelines/index.html
[24] Pulled on January 30th, 2023 from https://learningatscale.acm.org/las2022/call-for-papers/
[25] https://learningatscale.acm.org/las2022/call-for-papers/
[26] The L@S 2021 website hosted by the *European MOOCs Stakeholders Summit* (EMOOCs) no longer exist, meaning this sentence might be slightly inaccurate. However, looking at previous L@S conferences, this is unlikely.

		Public Access	Open Access	Available	Total
Number of Papers		2	14	77	93
Conference	8^{th}	2	3	16	21
	9^{th}	0	11	61	72
Misclassified		0	1	4	5
Data Availability	Open	0	1	0	1
	On Request	0	0	0	0
	No	2	11	60	73
	N/A	0	2	17	19
Data Documentation	Full	1	0	0	1
Material Availability	Full	0	0	5	5
	Partial	1	2	6	9
	On Request	0	0	2	2
	No	1	10	48	59
	N/A	0	2	16	18
Material Documentation	Full	0	0	2	2
	Partial	0	0	3	3
README		0	0	4	4
License		0	0	3	3
Preregistration	Yes	0	0	0	0
	No	2	12	62	76
	N/A	0	2	15	17
Preprint	Yes	0	0	5	5
	No	2	12	59	73
	N/A	0	2	13	15

Table 1: A representation of the metadata collected from full papers, short papers, and poster papers broken up by open methodology published within the proceedings of the 9th and 8th L@S conferences.

ACM. For the sake of this work, preprints, post-prints, and the publisher's version will all be referred to as preprints.

Within the peer review of LAK [7], AIED [8], and EDM [9], no data was collected on preprints. However, the responses of authors from the surveys of AIED and EDM contained about a 2% adoption rate, each reporting 5 papers had an associated preprint. However, only 3 responses from AIED and 4 from EDM provided a link to the associated preprint.

Finally, no links were degraded or within the reviewed papers. While this was considerably better compared to the 3%-5% of papers from previous conferences, the result was not directly comparable due to the large number of papers which did not use data or any materials.

4.2 Author Survey

Of the 93 surveys sent within the one month period, only 7, or 8% of the articles, provided a complete response. Five, or 5% of the surveys, did not reach their destination in a timely fashion: 1 received an auto response email about a delay in reading the email and 4 emails were no longer available or locatable on the email server.

Authors from AIED and EDM provided 17 and 13 responses within the chosen timeframe, respectively. While the participant pool was much larger, it only accounted for 6% of the results at each conference. As the response rate was low, any conclusions drawn from the survey results was likely to reflect the attitudes and reasoning of a subgroup of the L@S community. We talk about this limitation in more detail in Section 6.

Out of the 7 responses:

- 1 reported that their dataset was publicly available,
- 1 reported that their dataset could be requesteted,
- 3 reported that their dataset could not be shared, and
- 2 reported that they did not use a dataset.

For the 2 public and on request responses, the papers did not mention the dataset could be shared. Responses from AIED and EDM could not be compared as the public and private datsets only represented a single author in each scenario.

Between the 4 on request and cannot share responses, 1 mentioned that the dataset needed to be reviewed before releasing, 2 mentioned that the dataset was still being used as part of an ongoing study and might be released later, and 1 mentioned that the dataset contains sensitive information which could not be released, due to restrictions set by their human ethics committee. AIED and EDM were similar and different in their opinions as they were unable to release their dataset because of sensitive information or did not have the rights to the dataset.

For materials:

- 2 reported that their materials and source were publicly available,
- 2 reported that their materials and source could not be shared,

- 2 reported that they did not use any materials, and
- 1 mentioned that their materials and source were to become public at a later date.

Both public responses did not mention the materials in the paper. Out of the other 5 papers, 2 mentioned that the materials were still being used as part of an ongoing study and might be released later while the remaining 3 mentioned that it was non-applicable to their paper. Previous conferences had different reasons: they did not have the required rights to release the materials, the materials contained or referenced sensitive information, or there was a lack of time or motivation to make the materials public.

Towards reproducibility, 2 mentioned that additional information was necessary to reproduce their work. The information did not make any existing papers more reproducible or add another paper which could be tested for reproducibility. This was similar to AIED and EDM.

All 7 survey responses mentioned that they did not create a preregistration. Three believed that one was not necessary due to the type of study, 1 did not remember the option existed, 1 was unfamiliar with how to create or use a preregistration, 1 did not know what a preregistration was, and 1 provided no reasoning.

The majority of the authors who did not create a preregistration (13/28) at AIED and EDM did not know what a preregistration was. Excluding those who were unfamiliar with preregistration, the next largest group (10/28) did not believe that a preregistration was necessary, similar to the L@S results.

One author reported that they did create a preprint and provided a link while the remaining 6 did not. Out of the 6 who did not create a preprint, 5 did not believe a preprint was necessary while 1 provided no reasoning.

Similar to the preregistration results, the majority of authors who did not create a preprint (7/20) were either unfamiliar with preprints or did not know of them altogether. The next largest group (6/20) did not believe a preprint was necessary due to the eventual publication of the work.

No survey responses reported any degraded resources. However, 1 did provide a link to an OSF site for their paper. This was similar to EDM. Two responses at AIED had mentioned that some of their links were potentially degraded, but they did not provide any alternative links to the resources.

4.3 Reproducibility

No papers provided both open data and open code, preventing us from attempting any reproducibility efforts (The one paper that did have the full dataset did not release the associated materials or source). Additionally, the author survey responses did not provide enough information to reproduce the papers.

5 DISCUSSION

5.1 'Statement on Open Science' Comparison

When reviewing the previous conferences as a whole, LAK [7], AIED [8], and EDM [9] had made their data and materials more openly available than L@S. While degradation of resources was lower in L@S, this did not account for the 20% of papers which were marked as non-applicable for having open data or open materials. Additionally, the degraded resource links referred to datasets and materials, which was not provided within most of the L@S papers. As such, there was not enough information to make any meaningful comparisons within the two practices. This was further evidenced through the author surveys, where only one paper for datasets and two papers for materials could be made public.

Reviewing the cases where the dataset and materials could not be requested, the works published at L@S, at least within the response subgroup, mentioned that the resources were still part of an ongoing study. The resources could be requested in the future after the contributions in the work were completed; however, the authors did not comment on any future plans.

The reproducibility of the papers were not affected by the responses collected from the authors throughout the previous conferences and this one. It was unlikely to extract anything useful from a general question about reproducibility to the authors. A replication of this work after a few years would likely reveal different results depending on what work would be completed and whether they had moved on from their exploratory or theoretical phase of development.

5.1.1 Preregistration and Preprints. Preregistrations and preprints were simplified to binary outcomes, so we could perform some comparison between this conferences, AIED, and EDM. Within the previous conferences, the use of preregistrations was approximately equal, even including the survey responses. Preprints, on the other hand, were used by AIED and EDM. However, viewing the survey responses for the opinions that expressed that they did not create a preregistration or preprint showed that the majority of authors did not know what a preregistration (13/28) or preprint (7/20) was. Excluding those without the knowledge, however, authors at AIED, EDM, and L@S agreed that preregistrations (16/33) and preprints (12/26) were not necessary.

For preregistrations, some of the authors responded that preregistrations were typically used for empirical studies or confirmatory analyses, which they were not conducting. They believed the primary format of a preregistration is geared toward confirmatory analyses; however, this is a common misunderstanding. Preregistrations can be used for documenting hypotheses for exploratory analyses, exclusion or inclusion of data rules, defining dependent variables of interest, and/or even documenting that authors have no a priori hypotheses. Simply put, preregistrations can range in specificity from a simple statement about what authors are interested in looking at to a full analysis plan of an entire study.

For preprints, some of the authors responded that there was no point in uploading a preprint if it would just be released in the conference proceedings. However, not all published papers are open access directly on the publisher's website. Depending on the conference, one might have to pay a fee to make the paper open on the publisher's website (ACMDL for LAK, L@S), or the proceedings might not be public whatsoever (Springer for AIED). As such, although the preprint might contain the contents of the paper prior to submission, it would be openly accessible for a reader to view, typically through the author's website, institutional repository, or on a preprint server like arXiv or EdArXiv. One must be careful to make sure that the uploaded version of a paper was, in fact, a preprint.

6 LIMITATIONS AND FUTURE WORK

Most of the limitations and mitigations from previous works [7–9] for RQ1 were still applicable to this paper. For the peer review, this included the subjective opinions of the reviewers which were mitigated using an explanations document containing the reasons behind our choices. For the author survey, this included the nonexistent fallback strategy for emails that failed to reach their destination and ambiguous instructions in the survey itself.

The survey responses rate from authors was low, only encompassing 7 of the 93 papers. However, this was still higher than the ratio at AIED and EDM, around 6% for each conference. Additionally, it was likely that the respondents of the surveys would had published papers across each of these conferences as they are roughly within the same discipline of education technology. As such, any conclusions would likely reflect only a small subgroup of the education technology community. Future work would like need to collect a larger number of responses to conclude anything meaningful about a conference, much less a discipline. Additionally, one might compare a less related field like economic or biology to see whether these results would be consistent across fields with similar, less, or more levels of openness.

We were unlikely to locate any preprints that were not stored within a preprint server. Personal websites and institutional repositories were unlikely to be noticed by an indexing search engine. Additionally, depending on the institution and personal website, it would be impossible to determine when the preprint was archived, making the preprint status of the paper before conference publication unclear.

No papers publicly provided their data and materials, so we could not attempt to reproduce any of the reported results. This was still the case after including the responses from the survey. As our preregistration stipulated which years would be reviewed, we reported the result as is within our paper. Future work should conduct this review over a broader timeline to verify previous years.

Additionally, the survey responses did not specifically ask whether the authors viewed the open science statement nor the CFP for L@S. It was reasonable to assume that the author had viewed the CFP as the submission link is typically accessible on that page. Additionally, the important dates and the submission link were generally summarized at the bottom of the CFP, which the author would have needed to scroll past the open science statement to reach. As such, we reasoned that the survey responses could be used to analyze the adoption of open science practices and their reasonings not to.

7 CONCLUSION

Compared to papers published in the proceedings of LAK, AIED, and EDM, the overall adoption of open science practices was much lower in L@S papers with 1% using an openly available dataset and 5% using openly available materials. However, when comparing the responses of authors from AIED and EDM, researchers at L@S were more familiar with open science practices, considering the low sample size and the ratio between survey responses to published papers.

Not all papers need to or can meet all open science practices because of circumstances. For example, numerous papers within L@S were marked as non-applicable for open data or materials as they were theoretical in nature on how to design a platform. If a dataset was available, it might contain personally identifiable information or the author might not own the rights to the dataset itself, making it difficult or impossible to distribute. Similarly to the LAK work [7], this work failed to meet the open data requirements. The raw dataset was papers from the ACMDL site, which were made accessible through my institution. Even if the paper was open access, we would likely not have the right to redistribute the papers with the rest of the collected metadata.

This work was meant to review and compare how L@S's 'Statement on Open Science' affects the adoption of open science and robust reproducibility practices within research papers. While the chosen dataset might have not provided an accurate view of L@S as a whole, we were able to get a better understanding of some of the reasons for or against the adoption of various open science practices. In the future, we hoped to get a clearer understanding of L@S while encouraging the Learning @ Scale community to share their practices and have a larger discussion of the topic. Open science is not a catch-all to make all research perfect, but hopefully, it would provide a way to communicate and collaborate on work more effectively.

ACKNOWLEDGMENTS

We would like to thank the NSF (e.g., 2118725, 2118904, 1950683, 1917808, 1931523, 1940236, 1917713, 1903304, 1822830, 1759229, 1724889, 1636782, & 1535428), IES (e.g., R305N210049, R305D210031, R305A170137, R305A170243, R305A180401, R305A120125, & R305R220012), GAANN (e.g., P200A180088 & P200A150306), EIR (U411B190024 & S411B210024), ONR (N00014-18-1-2768), and Schmidt Futures. None of the opinions expressed here are that of the funders. We are funded under an NHI grant (R44GM146483) with Teachly as a SBIR. We would also like to extend our thanks to Judy Kay and the blinded reviewers who provided their suggestions towards improving this paper.

REFERENCES

[1] Karen M Albert. 2006. Open access: implications for scholarly publishing and medical libraries. *Journal of the Medical Library Association* 94, 3 (2006), 253. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1525322/

[2] Monya Baker. 2016. 1,500 scientists lift the lid on reproducibility. *Nature* 533, 7604 (01 May 2016), 452–454. https://doi.org/10.1038/533452a

[3] Philip E. Bourne, Jessica K. Polka, Ronald D. Vale, and Robert Kiley. 2017. Ten simple rules to consider regarding preprint submission. *PLOS Computational Biology* 13, 5 (05 2017), 1–6. https://doi.org/10.1371/journal.pcbi.1005473

[4] Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* 349, 6251 (2015), aac4716. https://doi.org/10.1126/science.aac4716 arXiv:https://www.science.org/doi/pdf/10.1126/science.aac4716

[5] Barthélémy Dagenais and Martin P. Robillard. 2010. Creating and Evolving Developer Documentation: Understanding the Decisions of Open Source Contributors. In *Proceedings of the Eighteenth ACM SIGSOFT International Symposium on Foundations of Software Engineering* (Santa Fe, New Mexico, USA) *(FSE '10)*. Association for Computing Machinery, New York, NY, USA, 127–136. https://doi.org/10.1145/1882291.1882312

[6] Arnoud Engelfriet. 2009. Choosing an open source license. *IEEE software* 27, 1 (2009), 48–49. https://doi.org/10.1109/MS.2010.5

[7] Aaron Haim, Stacy Shaw, and Neil Heffernan. 2023. How to Open Science: A Principle and Reproducibility Review of the Learning Analytics and Knowledge Conference. In *LAK23: 13th International Learning Analytics and Knowledge Conference* (Arlington, TX, USA) *(LAK2023)*. Association for Computing Machinery, New York, NY, USA, 156–164. https://doi.org/10.1145/3576050.3576071

[8] Aaron Haim, Stacy T Shaw, and III Heffernan, Neil T. 2023. How to Open Science: A Reproducibility Author Survey of the Artificial Intelligence in Education Conference. https://doi.org/10.35542/osf.io/xkmfw

[9] Aaron Haim, Stacy T Shaw, and III Heffernan, Neil T. 2023. How to Open Science: Debugging Reproducibility within the Educational Data Mining Conference. https://doi.org/10.35542/osf.io/4hn8q

[10] Steve Hitchcock. 2004. *The effect of open access and downloads ('hits') on citation impact: a bibliography of studies.* Project Report. University of Southampton. https://eprints.soton.ac.uk/354006/

[11] Johndan Johnson-Eilola. 2002. Open Source Basics: Definitions, Models, and Questions. In *Proceedings of the 20th Annual International Conference on Computer Documentation* (Toronto, Ontario, Canada) *(SIGDOC '02)*. Association for Computing Machinery, New York, NY, USA, 79–83. https://doi.org/10.1145/584955.584967

[12] Jocelyn Kaiser. 2017. The preprint dilemma. *Science* 357, 6358 (2017), 1344–1349. https://doi.org/10.1126/science.357.6358.1344 arXiv:https://www.science.org/doi/pdf/10.1126/science.357.6358.1344

[13] Mashael Khayyat and Frank Bannister. 2015. Open data licensing: more than meets the eye. *Information Polity* 20, 4 (2015), 231–252. https://doi.org/10.3233/IP-150357

[14] K.D. Knorr. 1981. *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science.* Elsevier Science Limited, Los Angeles, CA, USA. https://books.google.com/books?id=HF8cAQAAIAAJ

[15] Miika Koskela, Inka Simola, and Kostas Stefanidis. 2018. Open Source Software Recommendations Using Github. In *Digital Libraries for Open Knowledge*, Eva Méndez, Fabio Crestani, Cristina Ribeiro, Gabriel David, and João Correia Lopes (Eds.). Springer International Publishing, Cham, 279–285. https://doi.org/10.1007/978-3-030-00066-0{_}24

[16] Peter Kraker, Derick Leony, Wolfgang Reinhardt, and Günter Beham. 2011. The case for an open science in technology enhanced learning. *International Journal of Technology Enhanced Learning* 3, 6 (2011), 643–654. https://doi.org/10.1504/IJTEL.2011.045454 arXiv:https://www.inderscienceonline.com/doi/pdf/10.1504/IJTEL.2011.045454

[17] B. Latour, S. Woolgar, and J. Salk. 1986. *Laboratory Life: The Construction of Scientific Facts.* Princeton University Press, Princeton, NJ, USA. https://books.google.com/books?id=XTcjm0flPdYC

[18] Learning at Scale 2021. *L@S '21: Proceedings of the Eighth ACM Conference on Learning @ Scale* (Virtual Event, Germany). Learning at Scale, Association for Computing Machinery, New York, NY, USA.

[19] Learning at Scale 2022. *L@S '22: Proceedings of the Ninth ACM Conference on Learning @ Scale* (New York City, NY, USA). Learning at Scale, Association for Computing Machinery, New York, NY, USA.

[20] Jennifer C. Molloy. 2011. The Open Knowledge Foundation: Open Data Means Better Science. *PLOS Biology* 9, 12 (12 2011), 1–4. https://doi.org/10.1371/journal.pbio.1001195

[21] Benjamin Motz, Christopher Brooks, Joshua Quick, Yoav Bergner, Geraldine Gray, Charles Lang, Warren Li, and Fernando Marmolejo-Ramos. 2022. A Baseline Measure of Open Research Practices in Learning Analytics. https://doi.org/10.35542/osf.io/325d7

[22] Peter Murray-Rust. 2008. Open data in science. *Nature Precedings* 1, 1 (18 Jan 2008), 1. https://doi.org/10.1038/npre.2008.1526.1

[23] E.M. National Academies of Sciences, P.G. Affairs, E.M.P.P. Committee on Science, B.R.D. Information, D.E.P. Sciences, C.A.T. Statistics, B.M.S. Analytics, D.E.L. Studies, N.R.S. Board, D.B.S.S. Education, et al. 2019. *Reproducibility and Replicability in Science.* National Academies Press, Washington, D.C., USA. https://books.google.com/books?id=6T-3DwAAQBAJ

[24] Brian A. Nosek, Emorie D. Beck, Lorne Campbell, Jessica K. Flake, Tom E. Hardwicke, David T. Mellor, Anna E. van 't Veer, and Simine Vazire. 2019. Preregistration Is Hard, And Worthwhile. *Trends in Cognitive Sciences* 23, 10 (01 Oct 2019), 815–818. https://doi.org/10.1016/j.tics.2019.07.009

[25] Brian A. Nosek, Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. 2018. The preregistration revolution. *Proceedings of the National Academy of Sciences* 115, 11 (2018), 2600–2606. https://doi.org/10.1073/pnas.1708274114 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1708274114

[26] Brian A. Nosek, Tom E. Hardwicke, Hannah Moshontz, Aurélien Allard, Katherine S. Corker, Anna Dreber, Fiona Fidler, Joe Hilgard, Melissa Kline Struhl, Michèle B. Nuijten, Julia M. Rohrer, Felipe Romero, Anne M. Scheel, Laura D. Scherer, Felix D. Schönbrodt, and Simine Vazire. 2022. Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology* 73, 1 (2022), 719–748. https://doi.org/10.1146/annurev-psych-020821-114157 arXiv:https://doi.org/10.1146/annurev-psych-020821-114157 PMID: 34665669.

[27] Prasad Patil, Roger D. Peng, and Jeffrey T. Leek. 2016. A statistical definition for reproducibility and replicability. *bioRxiv* 1, 1 (2016), 1–1. https://doi.org/10.1101/066803 arXiv:https://www.biorxiv.org/content/early/2016/07/29/066803.full.pdf

[28] Bruce Perens et al. 1999. The open source definition. *Open sources: voices from the open source revolution* 1 (1999), 171–188. https://www.researchgate.net/publication/200027107{_}Perens{_}Open{_}Source{_}Definition{_}LG{_}26

[29] Gopi Krishnan Rajbahadur, Erika Tuck, Li Zi, Zhang Wei, Dayi Lin, Boyuan Chen, Zhen Ming Jiang, and Daniel Morales German. 2021. Can I use this publicly available dataset to build commercial AI software? Most likely not. *CoRR* abs/2111.02374 (2021), 1–1. arXiv:2111.02374 https://doi.org/10.48550/arXiv.2111.02374

[30] Diana Rhoten and Walter W. Powell. 2007. The Frontiers of Intellectual Property: Expanded Protection versus New Models of Open Science. *Annual Review of Law and Social Science* 3, 1 (2007), 345–373. https://doi.org/10.1146/annurev.lawsocsci.3.081806.112900 arXiv:https://doi.org/10.1146/annurev.lawsocsci.3.081806.112900

[31] Ido Roll, Daniel M. Russell, and Dragan Gašević. 2018. Learning at Scale. *International Journal of Artificial Intelligence in Education* 28, 4 (01 Dec 2018), 471–477. https://doi.org/10.1007/s40593-018-0170-7

[32] Hendrik Schoettle. 2019. Open Source License Compliance-Why and How? *Computer* 52, 08 (aug 2019), 63–67. https://doi.org/10.1109/MC.2019.2915690

[33] Barbara A. Spellman. 2015. A Short (Personal) Future History of Revolution 2.0. *Perspectives on Psychological Science* 10, 6 (2015), 886–899. https://doi.org/10.1177/1745691615609918 arXiv:https://doi.org/10.1177/1745691615609918 PMID: 26581743.

[34] Anna Elisabeth van 't Veer and Roger Giner-Sorolla. 2016. Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology* 67 (2016), 2–12. https://doi.org/10.1016/j.jesp.2016.03.004 Special Issue: Confirmatory.

[35] Ruben Vicente-Saez and Clara Martinez-Fuentes. 2018. Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research* 88 (2018), 428–436. https://doi.org/10.1016/j.jbusres.2017.12.043

A COMPUTER SPECIFICATIONS

A.1 Hardware Components

- AMD Ryzen Threadripper 2950X[27]
- NVIDIA GeForce RTX 3090[28]
- Corsair VENGEANCE LPX 128GB (4 x 32GB) DDR4 DRAM 2133MHz C18 Memory Kit
- WD Blue SN550 NVMe SSD (WDC WDS200T2B0C-00PXH0)[29]

A.2 Software Components

Some of the software components are considered the default if no specific version was specified in Section 3.3.

- Ubuntu 20.04.5 LTS[30]
- Linux Kernel 5.15.0-53-generic
- GNU bash 5.0.17(1)-release (x86_64-pc-linux-gnu)
- Python 3.8.10[31]
- R version 4.2.2 Patched (2022-11-10 r83330)[32]

[27] https://www.amd.com/en/product/7926
[28] https://www.nvidia.com/en-us/geforce/graphics-cards/30-series/rtx-3090-3090ti/
[29] https://documents.westerndigital.com/content/dam/doc-library/en_us/assets/public/western-digital/product/internal-drives/wd-blue-nvme-ssd/product-brief-wd-blue-sn550-nvme-ssd.pdf
[30] https://releases.ubuntu.com/focal/
[31] https://www.python.org/downloads/release/python-3810/
[32] https://cran.r-project.org/bin/linux/ubuntu/