Open Science and Reproducibility in Education Science

This report represents the work of

William Christopher Baxter and Robert Gyurcsan

Submitted to the Faculty of

Worcester Polytechnic Institute

As evidence of completion of a degree requirement for the

Degree of Bachelor of Science.

Date Submitted: May 3, 2023

Professor Neil Heffernan

Professor Stacy Shaw

Abstract

Reproducibility – the ability to replicate and verify the results of a study – is key to the integrity of the scientific community. Our team assessed papers in three conferences in education technology to determine their adherence to open science practices and determine their reproducibility. This was done by scouring every paper for openly provided data and materials, which could be used to replicate and verify the results of each paper. Our findings indicate low reproducibility levels in the field of education technology, and we propose methods that can improve such metrics.

Acknowledgements

We would like to thank Dr. Neil Heffernan and Dr. Stacy Shaw for their help and support throughout this project. We would also like to thank Aaron Haim for his help advising on this project. Finally, we would like to thank Worcester Polytechnic Institute for the opportunity to work on this project. The skills and experience gained on this project will help us in our future work.

Executive Summary

The purpose of this project was to evaluate the reproducibility and prevalence of open science practices in the education technology community. Open science is a movement of research practices whereby the materials, research, and data used in a study are readily available for other scientists for the sake of verification and replication of results. Open science practices are necessary to ensure the accountability and integrity of the scientific community, as phenomena presented in studies ought to be reproducible no matter who conducts the study. We reviewed papers in the conference proceedings of Artificial Intelligence and Education, Educational Data Mining, and Learning @ Scale, cataloging each paper's adherence to open science practices and recording the results in a Qualtrics survey and a Google document. Papers which were potentially reproducible were flagged for the project lead to replicate, who would then determine whether a study is reproducible by running the code and data provided. We determined that for the two most recent years of L@S and EDM, only **two** out of the 301 papers in the proceedings had reproducible results.

There were two phases to the project. In the first phase, conducted in B term, we reviewed the two most recent years of EDM and L@S. After the completion of our survey, our project lead reviewed papers that we determined had openly provided data and enough materials to potentially replicate the results. The second phase was conducted in C and D term. Learning from the data we acquired, our project lead expanded the metrics which we reviewed the papers on, allowing for more specificity and depth. We also expanded the scope of the project to include AIED in addition to L@S and EDM, and increased the range of conferences to the most recent

four years instead of the most recent two. As this second survey phase is still ongoing, the project lead has not yet reviewed and determined the reproducibility of the current data, but currently only 5.69% of the studies surveyed have both open-source code and data, allowing them to be checked for reproducibility by the project lead.

We adopted a standard review process for analyzing papers. We would first briefly read over the paper, and then we would open an online survey. We would then comb through the study again and enter in what open science practices we found in the survey, while also entering what practices we did not find where applicable. The survey we used when reviewing our papers was conducted via the Qualtrics platform. Despite the survey expanding between the phases of the project, the main fields of the survey persist between variations. The primary fields for the survey were open methodology, open data, and open materials. Open methodology gauged whether a study's methodology could be accessed, and if it was not available to the public, if it could be obtained through WPI's network. Open data assessed whether the data the study used was provided clearly and in full in the study. Open materials gauged whether the source code the study used to generate the results was provided, and if not, whether some supplementary materials the researchers used were given instead. All three of these fields must be fulfilled for a study to be potentially reproducible. True reproducibility, though, is verified by the project lead, who runs the provided source code with the given data. If the results of the code match up with the results given in the study, the paper was considered reproducible.

The results of our project indicate there is clear room for the education technology community to improve in open science practices. We recommend that all education technology researchers, where possible, provide their data and source code in an easily accessible manner. However, we understand that due to the sensitive nature and legality of personally identifiable

data of students, this may not always be possible. We also recommend, in the event of the continuation of this project by other students, that the 2019 and 2020 conferences of EDM, L@S, and AIED be reviewed and redone to verify the integrity of the data. In addition, we recommend that the upcoming conferences of 2023 be surveyed. Lastly, we recommend that the 2017 and 2018 conferences of EDM, L@S, and AIED ought to be surveyed and have their results compared to current reproducibility data to chart the evolution of open science practices.

Authorship

Introduction

Primary Author- Robert Gyurcsan

Revisions- All

Background

Primary Author- Robert Gyurcsan

Revisions- All

Methodology

Primary Author- Chris Baxter

Revisions- All

Limitations

Primary Author- Chris Baxter

Revisions- All

Findings

Primary Author- Robert Gyurcsan

Revisions- All

Conclusions and Recommendations

Primary Author- Chris Baxter

Revisions- All

Table of Contents

List of Figures

1. Introduction

Open science has become more and more prominent in the field of psychology research. The fundamental goal of Open Science is to make research, data, and materials available to everyone, not just other scientists. The availability of open access to data, methods, and platforms could provide better collaboration in research. This research involves the analysis of how papers in four different conferences used Open Science principles in their research. The push toward open science was catalyzed by many researchers wanting to be able to more easily collaborate and build off each other's research. It is thought that access to data, methods, and materials will make research more transparent and reproducible. The open science movement uses six principles aimed at making all research readily accessible to everyone. The six main principles are open education, open access, open methodology, open data, open materials, and open peer review. Four of these principles are what this project is primarily focused on. They are open access, open methodology, open data, and open materials.

There have been many studies conducted to investigate the effects of open science on a paper but not many to examine how conferences suggest the use of Open Science and how researchers have adapted. There has not been much research as to the progress of researchers utilizing open science principles in their papers. The goal of this project is to examine the progress that has been made over several years with respect to the openness of methodologies, data, materials, access, and the ability for one's work to be reproduced.

2. Background

Science research, in general, for a long time was, and much of it still is, very closed off. Researchers wanted to be able to build off colleagues' work more easily and as such wanted to have access to exact results. Around 2010, researchers began to examine and use the term Open Science more commonly (Spellman et al., 2017). It was around this time that Open Science research started to become an increasingly important aspect of psychological research. This movement encouraged researchers to provide access to all their work to everyone as well as documentation as to how they did their research. The movement aims at increasing transparency and availability of information in the interest of furthering research.

Open Science can be represented by six commonly accepted principles that researchers are encouraged to build into their work. The principles are open education, open access, open methodology, open data, open materials, and open peer review. For the purposes of this project, we will only examine open access, open methodology, open data, and open materials. This project will also examine the reproducibility of researchers' experiments. These principles were primarily chosen because they are the most common principles and most well-known and understood in the science community.

2.1 Open Access

Open access is the principle that papers are available to all readers without having to pay. Research is mainly funded by the public and as such it does not make sense to have the public pay again to read the results of the research. This means that after the public pays taxes which sometimes go to fund certain research, they should again pay to be able to access findings and data that the research yielded. Open access includes initial drafts of papers or preprints that are available on repositories that authors may want to release for commenting. The Budapest Open

Access Initiative, in general, states that open access is any work that is available online that authors release to the world without expecting payment (Bailey, 2006). A fundamental aspect of a work being open access is the work being online. The initiative states that the primary source of open access materials is that of peer reviewed journals because the authors of such papers are not paid. There are two primary forms of open access. Green open access is done by self-archiving papers on the internet. Gold open access is provided by publishing papers in conferences or journals (Gargouri et al., 2012).

2.2 Open Data

Open data is the principle that all data, raw and processed, is available and can be accessed by others without restriction. This does not just include the successful data, but the failed experiments as well. There is some bias towards only making data that produces significant results available. Open data, in our project, is the specific dataset used in the experiment, including the processed data. Again, the argument of who funds the collection of the data. The public should be able to view data they fund to collect. The question does come whether some data can be open. In some cases, data cannot be openly shared to due the sensitivity of the data. Often in research of this nature, including psychology and education research, there is data that is personal information or information that cannot be shared with the public.

2.3 Open Methodology

Open methodology is the principle that a researcher's experiment is described in enough detail such that it can be reproduced by another researcher. This is a major aspect of whether an experiment can be replicated. Often this is supplied in sufficient detail via the methods section of a paper, however it is also a possibility that there may be other methods used detailed elsewhere

such as an online repository. This is to say, there are methodologies present in the text of the papers as well as methodologies located elsewhere. This is at the heart of open science because it is how other researchers can learn from previous experiments. An open methodology section is written in enough detail such that all datasets and materials are mentioned as well as how they were applied. The open data and open materials sections are often detailed in an open methodology.

2.4 Open Materials

Open materials are the principle that anything used to analyze the data is available. Materials include source code as well as anything else used to analyze the raw dataset. Source code must be able to be read as well as used. This pillar refers to open methodology. If a material is mentioned in the methodology, then it should also be provided via a link or similar method. Materials should also be commented on so the researcher attempting to reproduce the experiment can understand what each section of the code does. For the purposes of this project, we only care about source code, however, details on all materials used paint a better picture of the experiment.

2.5 Reproducibility

The primary objective of this project is to determine the reproducibility of experiments. An experiment being reproducible means that the same results can be obtained using the same data as well as any materials. The paper must be open access to view the experiment at all. The methodology must be open to understand how the experiment was performed. The raw data must be available because without it there is no data to reproduce. The materials must be open because you must have a method to manipulate the data with. A paper may be reproducible if only some of the principles are applied.

2.6 Conferences

Learning at Scale

The Learning at Scale Conference deals with collecting research on a large scale about technology-aided learning environments. This field tries to examine how computer-aided learning can be tailored to an individual's needs. This examines the rise in online courses, mobile learning applications, and similar applications.

Educational Data Mining

The International Educational Data Mining Society aims at supporting research in the field of Educational Data Mining. This is a newer field that involves analyzing the data that is collected in educational settings. This data comes primarily from computer-aided learning environments and administrative data from educational institutions.

Learning Analytics and Knowledge

The International Conference on Learning Analytics and Knowledge is aimed at researching the intersection of learning and analytics. This conference brings together researchers in education fields, data scientists, software developers, and policymakers in the government. This conference is held with the Association for Computing Machinery.

Artificial Intelligence in Education

The International Artificial Intelligence in Education Society brings together researchers from the computer science, education, and psychology fields. The society promotes research for adaptive learning environments for all students. This society is also part of a society that promotes computer-supported learning.

3. Methodology

To assess reproducibility and open science practices in the learning science community, we scrutinized the research articles, short papers, and posters recorded in the proceedings for the recent conferences AIED, EDM, and L@S. In order to obtain an appropriate assessment of the reproducibility of the papers in the conferences, the papers had to be cataloged using a uniform method that balanced thoroughness and punctuality. We utilized a dual-pronged approach, recording basic information about the papers in a Qualtrics survey, while also documenting and justifying these survey entries in a supplemental google doc (Qualtrics, 2015; Google, n.d.). Documentation for all the conferences was collected using Qualtrics surveys, which contained multiple fields pertaining to different qualities the examined papers exhibited that contribute to reproducibility analysis. These qualities included but were not limited to: the paper's digital object identifier (DOI), open methodology, materials availability, and data accessibility. Despite the survey being modified over time to record more information to suit the needs of the project, the general meanings of the survey fields did not change, and rather only expanded upon the options that could be recorded when analyzing different papers. Supplemental documentation is provided via google docs for every reviewed paper explaining decisions and marks made for that paper, going into depth about what materials were provided in the paper, if any, linking source code, data sources and what the data is, licenses the project and its components are managed under, the presences of preprints of the paper on the internet, and other fields. Link degradation was also noted as well in the cases where applicable. Time to review papers was limited to a maximum of 15 minutes per paper primarily due to logistics. In the event that a paper is on track to exceed this limit, it would be marked as such in the applicable survey field. This timer is separate from the timer used by the project lead who verifies reproducibility of flagged papers.

The timeframe for which conferences were surveyed was limited to a maximum of the four most recent years. Initially the scope of the IQP limited the conferences to the first two most recent, but after the survey of all of the required papers during the first term, the survey was expanded in detail, adding additional content to each of the major fields and necessitating a new survey of the previously analyzed conferences. This provided twofold benefits: not only could the new entries be compared to the previous entries, allowing a check in the validity of previous data, but the 2019 and 2020 data can be compared with more recent data to show potential changes in the use of open science practices with time. The analysis of the conferences to the four most recent years was done for primarily two reasons. First, practices of open science have been gaining traction and expanding since the early 2010's, and reviewing papers further beyond this local timeframe may potentially misrepresent modern open science and reproducibility practices. Second, older papers are far more likely to have difficulties in reproducibility: links often degrade over time, data and source code are lost, old packages and libraries were designed to work with older equipment, and software and may run differently than anticipated or not even be usable at all.

3.1. Open Methodology

Open methodology is an umbrella term under open science which covers whether the methodology of the paper is openly accessible. Methodology is defined as a comprehensive overview of how a paper's research is conducted, providing structure to the study's setup and procedures. All papers in the EDM, AIED, and L@S conferences contain sections dedicated to the methodology concerning data collection and analysis in their paper. Whether this methodology is reproducible or not, however, is up for assessment on a per-paper basis. Alas, due to the 15-minute time limit for each paper, verifying the validity of the methodology is

difficult to do alongside verification of the validity of the results of the study. As such, open methodology for the sake of this paper will be defined on how accessible the methodology is - whether it can be read freely or if access to the study is gatekept. There is also a field for no open methodology, however this field was unused as no studies in the reviewed conferences were completely inaccessible.

L@S papers are hosted on the ACMDL website, where they are outfitted with metadata tags describing their state of availability. When papers do not have tags, they typically can only be accessed via an institution or some other method of gatekept access. However, if papers have the tags “Open Access” or “Public Access”, the papers are freely available. The difference between these tags is that “Public Access” is a special type of open access where the paper must be completely accessible to the public within a year of publication. Regardless of which of the two tags the paper has, so long as it has one, the paper is openly accessible. This tag is documented under the “Open Methodology” section of the survey. In the event that the paper is neither of these tags but is still accessible through an institution or other means but is not publicly accessible, the paper will be marked as “Available”. All of EDM and AIED’s papers are open access and are marked accordingly.

After the first term of the project, the survey was expanded to include a text box next to the Public Access and Open Access fields of open methodology. These fields are where the license for the paper would be recorded, if any was provided by the researchers. Open methodology does not extend its definition to content licenses, but rather only to the paper itself.

3.2. Open Data

Open Data is a term used to define the presence of one or more openly accessible datasets which are clearly provided without restrictions on its use. These datasets are either released into

the public domain or with a license or terms of use. Documentation in this study about open data will mark whether a study contains open data, a link to open data, or a link to an openly available study with the dataset. This field is marked as non-applicable if the paper does not use or collect data.

Due to the often sensitive and or identifiable nature of educational data, not all data can be released publicly. While some papers do not present their data at all due to this, on occasion there are authors which will provide the data in the event one requests it. The survey is equipped with an "On Request" field for this specific case. However, this requires the authors to express it outright; if there is no explicit statement of data being available on request by the author in the paper, the documentation will be marked as having no open data.

There is a separate field to mark whether a study has documentation available for the dataset. Data documentation is defined as the individual mapping of field names in the dataset to their respective values, along with a description of what the values mean. Full documentation will be marked if the paper has full documentation of the data or links to it. Partial documentation will be marked if there is documentation in at least one field, indicating the presence of a greater dataset that is not included or able to be included. A single mapping or a table of some data documentation is enough for a definition of partial documentation. Tables that include the frequencies of data do not count towards data documentation, as it is not a direct code mapping of what a data field is and how it could be used in the provided code. The fields will be marked not applicable if the study does not use or collect data. A data license field was also included in later editions of the survey, but it is not mandatory for data to have a license to be open. With partial data documentation, the dataset does not require a license, nor terms of use, nor does it need to be openly released, it just needs to have a mention of the documentation for

the data present in the study. In the event that data is not provided by the study but the paper did use data for their results, the paper's open data is marked as none, with the documentation and license fields being marked as non-applicable unless partial documentation is provided in the study.

3.3. Open Materials

Open materials is a term defined as resources that are accessible and usable without severely mitigating restrictions. A study requires open source code to be reproducible. Commonly open materials is synonymous with open source in the field of software development, but in the case for analyzing papers under the lens of open science the scope will be expanded: open materials is not only defined as an open source code, but in the event that source code is not present, open source software can still suffice for partial open materials. For example, this includes base models like BERT, processors like word2vec, or what environment the study is being run in like R or scikit-learn (*BERT*, n.d.; TensorFlow Staff, 2022; R Core Team, 2022; (scikit-learn, 2019). In such an event where there is no source code but the researchers include some materials that can be publicly accessed, that paper will still be marked as "No Source, Materials Available".

Additional subfields are present to clarify further upon open materials for the Qualtrics form. Documentation, like in open data, is necessary for proper understanding on how to use provided materials. For the purposes of this report, source code documentation qualifies as full if the code is readable and has comments, in addition to the presence of a README. Partial source code documentation is also an option, in the event that there is a code or algorithm snippet in the study itself. Given that it is possible for a study to have a README but no easily readable code, there is a separate field for a README that exists despite overlap with documentation. A

README is required for full source documentation, but it does not entail it. Lastly, a license field is also present, which can be marked if the materials possess a permissible software license allowing the user to openly use it. All of these fields can be marked non-applicable if standard non-applicability terms apply. Source documentation, README, and license are marked as non-applicable if there is no source code provided nor any source documentation in the paper.

Paid materials and commercial products are exempt from being defined as open materials due to the complex nature of the literature as to whether they are open or not. As of writing this paper there is no literature explicitly stating whether or not paid materials constitute open materials. Because of this, paid materials are not mentioned in the supplemental documentation unless they are the only materials provided, to which an explicit mention is made in the documentation stating that the only materials used are paid products, and the Qualtrics survey is marked as having no open materials. If paywalled materials are used in conjunction with other open materials, the study will still be marked as having open materials. Regardless, so long as source code is present and open data is provided, allowing for the replication of results, paywalled materials usually do not impact the reproducibility of the study unless in extenuating circumstances. Such instances are made note of and left to the project lead to evaluate for themselves.

3.4 Preregistration

Preregistration is the process in which researchers outline the steps necessary for conducting the study prior to execution and data collection. This process is registered to an independent registry, such as the Open Science Foundation, where it is hosted and openly accessible (OSF Staff, 2023). These preregistrations preserve original methodologies and can be added on to over time as the study deems appropriate. The methodology of these preregistrations

cannot be overwritten. Common preregistration practices dictate that the preregistration is cited within the body of the paper and stating what parts of the research were preregistered. If a link to a preregistration is provided or if the study is found as preregistered to the OSF foundation website, the paper will be marked accordingly in the Qualtrics survey and the associated documentation. If a preregistration exists but lacks in content, a note will be made in the explanations document stating what is lacking on the preregistration. If the paper is a theoretical or an argumentative paper or has other aspects which would disqualify the need for preregistration, the field will be marked as non-applicable.

3.5 Preprints

Preprints are papers posted to online platforms prior to peer review and publication in academic journals, typically posted in dedicated servers and archives for hosting preprints (PLOS, n.d.; Columbia University Irvine Medical Staff, 2021). These papers, when available, can potentially vary from the original published papers. If a preprint is located on a hosting website such as EdArXiv or ArXiv, the necessary field would be filled in the survey, and the link would be provided in both the survey and the supplemental document (EdArXiv Staff, 2023; Cornell University Staff, 2019). Differences between the conference paper and its preprint are recorded in the supplemental documentation. Information provided in preprints can contribute to the reproducibility of a paper, and in any instance such information is required to reproduce the study, it is marked as such on the supplemental document. If the preprint is absent or dated after the conference (thus making it not a preprint) the field is marked as “No”, but nevertheless is still recorded in the survey document. If the paper is a work in progress, an abstract, or demonstration, this field is marked as non-applicable.

3.6 Reproducibility

The ultimate goal of the project is to find papers that could potentially have replicable results. However, the charge of running the code for results verification is not given to the members of the IQP, rather the project lead, and as such we never outright state whether a paper is reproducible. Instead, the field for reproducibility has additional options beyond full reproducibility, partial reproducibility, and none. In the event a paper has open data, accessible source code, and open methodology, a paper can be marked as "Not Tested" which flags the reviewer, the IQP lead, to look over the study personally and verify reproducibility. If a study is missing any of the requirements listed above, the paper cannot be reproduced, and is marked as such and an explanation is provided in the supplemental documentation. This includes if a study has data and source code available upon request, but a special flag is made in the supplemental documentation stating that the study is not reproducible as of a given moment, but could be upon contacting the researchers. In the event that a study is more theoretical in nature and has no results to reproduce, the reproducibility field can be marked as not applicable like in the other fields.

4. Limitations

The methodology of this paper implies some limitations on the results of the project which ought to be acknowledged. The first significant limitation imposed by the methodology is that of the time limit constraint per paper. The time limit exists primarily for logistical reasons. Many reviewed papers use machine learning models, datasets, and methodologies developed in other papers, which are cited and referenced. This cuts down on the researchers' methodology size, however tracking the original sources of these materials for raw data and code is time consuming. Fortunately, most of the EDM community utilizes common datasets such as ASSISTments, Bridge To Algebra, and others when not utilizing experimentally derived data

from their own sources. These common datasets are stored on their own websites or data hosting sites such as PSLC Datashop, where they can be readily obtained. However, this is not always the case for open materials when the source code is not provided. Referenced studies often do not have readily available code, libraries, or materials available, and often reference another source for their methodology. This hypothetically results in long chains of referenced studies, ultimately leading to a foundational study which may or may not have usable materials. Imposing the time limit, while potentially risking missing material used in a study separated by multiple degrees from the original, allows us to move on to other surveys. Without the time limit, cataloging the conferences would take far longer, and thus for the sake of having a larger data pool this constraint was implemented.

Another noteworthy limitation concerns the search of preprints. Preprints were sourced only from ExArXiv and ArXiv. Other repositories exist such as Google Scholar, however, given the prospect that indexing on personal websites may not be detected by the search algorithms, we decided that we would only look for preprints on the two aforementioned ArXiv sites. It is possible that we missed some preprints due to this limitation, however, there have been no instances where these two websites did not have an archived print of a paper when the conference paper itself provides a link to the preprint.

A more ambiguous limitation is how our understanding, as reviewers, of what materials to look for evolved and changed over time. For the first instance of the survey there were instances where we would include data collection materials as part of open materials. This changed later on as the focus of the project was towards reproducing results, not reproducing the data collection experiments. In a similar vein, as we became more accustomed to the definition of open materials, there were instances where some materials we listed were not, and some

materials we did not list were. For example, even though the names of models are often provided in methodologies as to how researchers processed their data, unless the specific environment they were used in was provided, that model is not actually an open material. Without knowing what environment was used for the model, the exact open material is unknown, as an algorithm could differ in how it is coded whether it is coded in R or in Scikit-Learn. Fortunately, unless packages are separate for algorithms in certain environments, most environments have all the necessary models an experiment can use. While this was not taken into account when we were first learning how to catalog reproducibility metrics, later on these algorithms eventually were condensed under their environments when listed in open materials. This results in a discrepancy between the first survey datasets and the more recent survey dataset - less open materials are taken into account.

In addition, instances where data and materials were available as “on request” were marked as such in first instances of the survey, but after the survey was overhauled, we would complete “on request” surveys and make accounts to access data where applicable. There were multiple occasions where studies that were on request had bots that gave you the data within minutes of completing the survey, allowing the data and materials to be checked for documentation. Naturally the data and materials would still be marked as on request, but the fields regarding data and source documentation would be filled out as well. However, despite Kaggle and Datashop requiring accounts to access data, these websites are so commonly used by studies that they were eventually changed to not be considered “On Request” but rather “Open Data”.

5. Findings

As detailed in the methodology chapter, the data from this project was collected in a series of surveys. The results from these surveys can be found in the appendix as Figures 1 to 29. In this chapter we will present the relevant data found from the surveys as it relates to the goals of our project. The findings will be examined as they relate to each of the principles of open science we chose to analyze in the four conferences chosen. Please refer to tables 1 and 2 where the findings are tabulated from phase 1 and phase 2 of the project. The findings are in agreement with the data in the subsequent sections in this chapter.

5.1 Open Methodology

As seen in Figure 11. all the methodologies were open access which indicates they were accessible to the reader. In Figure 1. 6.86% of the papers were only labelled as available. In Figure 20. 82.8% of the papers were available, 15.05% were open access and 2.15% were public access. These findings are very good as they show that most if not all of the methodologies are accessible by everyone. As stated in the methodology, this field is not as relevant when comparing conferences or papers due to the fact that all the papers examined were open and accessible to everyone and as such no further comments are needed. Had there been papers or methodologies that were only accessible via an institution, there could be more analysis done as to what types of papers those were and why they were not freely accessible to everyone.

5.2 Open Data

As seen in Figure 2. 75.99% of the papers examined did not have open data. Only 15.10% of papers examined had open data. As stated in the methodology, some data cannot be released to the public for several reasons. However, approximately 76% of the papers not having accessible data contradict much of what the open science principles are about. As seen in Figure 12. Approximately 79% of the papers examined were recorded as not being open. Figure 21. also

has approximately 79% of the papers examined recorded being not open. As stated, some of this can be explained by the data simply not being able to be released due to privacy concerns. The other portion of these datasets is due to the datasets simply not being linked to the published papers.

As stated in the methodology, the other portion of the data analysis is data documentation. As seen in Figure 3. approximately 74% of the datasets examined were determined to have no documentation. The remaining 26% are comprised of datasets having partial and full documentation as well as other less relevant categories for analysis. Figure 13. shows very similar percentages. Approximately 72% of the datasets examined did not have documentation, however 14% had partial documentation and 12% had full documentation. Finally, in Figure 22., 72% is seen as having no documentation however 20% is labelled as N/A. This very much aligns with what is evident in the open data findings. Approximately 75% did not have open data as well as data documentation.

Finally, as stated in the methodology, a dataset license field was added in later iterations of the survey. As seen in Figure 7., many of the papers were recorded as N/A. This was due to the dataset not being available for everyone to access. Only 3.35% of the approximately 16% that had open data had dataset licenses provided. This data shows that this is not a common practice among researchers, however it could become more common as time goes on.

5.3 Open Materials

Materials are anything that was used in the experiment or to process data. Having these materials available to all furthers progress in the field as well as allows researchers to complete similar research without starting from scratch. As seen in Figure 4. approximately 52% of the papers had no materials available reported. Many of these papers likely mentioned the use of

materials in the methodology sections of the papers, however, did not attach or link to any specific material. Only about 10% of papers had sources available. As seen in Figure 14. there were approximately 64% of papers that did not have open materials. However, about 15% did have open materials. In Figure 23. the data is broken up slightly differently. A very similar number can be seen as to how many papers had no open materials. The full and partial open materials together are close to the same percentage as the papers marked as open materials for Figure 14.

Like the open data section, the open material section also had an option for materials documentation. As seen in Figure 5. most of the papers that provided materials or source code also provided at least partial documentation. Similarly, in Figure 15., a large percentage of no documentation can be seen. In Figure 15., the data is broken up into fewer categories therefore it is difficult to compare. Around 30% have either partial or full documentation for materials however this is more broken down and does not specify source or source code specifically. Here the documentation could simply be referring to how the material was used in the experiment. In Figure 24., we see similar numbers where around 21% is labelled as N/A due to the lack of open materials. If no open materials are provided, the answer to whether the materials are documented must not be asked.

Often if there is source code available, README is provided to explain the broader use of the materials provided. Figure 6. indicates that most of the time when there is source code provided via a digital repository, README is provided. In Figures 16 and 25. as they are broken up differently with the N/A option not used or used much less it is much harder to identify the situations present. The team assumes that much of the “no” options are the result of no source code provided. As Figure 6. shows us, almost always as source code is provided, README is as

well. We can assume that the “yes” option correlates closely with the amount of source codes present among the materials used in the research.

Also, a more recent addition to the survey is the source license question. Figure 10. shows us that half of the time when source code is provided, a license is not provided in the repository. Figure 26. is presented in a slightly different way. The N/A option appears to be used much less in which case we can assume some of the “no” option mean that there is no license because there is no open materials provided. The “yes” option is used much less than the ”no” option which does not agree with the data in Figure 10. which would indicate the surveys approaching the question slightly differently.

5.4 Reproducibility

The reproducibility was determined by the graduate student leading the project as detailed in the methodology. As seen in Figure 9., approximately 97% were either not reproducible or not tested. This means that some key factors were and still are missing in these papers to not allow the results to be replicated. As seen in Figure 19., less than 1% was reproducible. Figure 29. also confirms this with no papers showing in the “yes” result in the survey. All three survey results yield the same findings that nearly no papers were reproducible. Possible methods to improve this are discussed in the conclusions and recommendations section later in the paper.

6. Conclusions and Recommendations

Over the past three terms our team has surveyed the conferences of EDM, AIED, and L@S to identify studies that have potentially reproducible results. Through the use of a Qualtrics survey, we reviewed each paper within these conferences and assessed the accessibility of their materials and data. Having reproducible results is integral to practices of open science, and as

modern open science practices are adopted over time, it is important to assess the spread of the adoption of these practices for the sake of a more genuine and transparent scientific community.

For the first editions of the survey, we found that for the past two years of the conferences EDM and L@S that only 2 studies out of 301 were reproducible, or approximately 0.664% reproducibility rate. Additionally, 33 studies from these conferences had open data, or 42 if we include the papers with data available on request. This translates to a 10.9% open data rate, respectively 14.0% if we include data available upon request. 36 of the studies have full open materials, resulting in a 12.0% open materials rate. Only 3 studies were preregistered, for a preregistration rate of 0.998%.

The later editions of the survey are still underway, and as a result the project lead has not conducted tests on the potentially reproducible papers. Instead, current data only states whether a study is testable, meaning they have open methodology, open-source code, and open data. Out of the 541 current entries in the survey, there are 24 papers that have all three provided or with source and data available upon request, resulting in a 4.44% testability rate. This is the current value but is likely to change post-submission. Given phase 1's L@S entries, it is expected to decrease.

Due to the sensitive nature of data related to students, it is impossible for all studies regarding educational research to be directly reproducible. As such, we have no expectations that every study will be able to align exactly with open science practices, but these practices ought to be implemented wherever they can be. Regardless, these studies are an illuminating view into research transparency and reproducibility in education science and show that there is much more room to grow as for the proliferation of open science.

We recommend that for future research into this topic that the next conferences released in 2023 are assessed just as we have done. For the sake of verification, we also recommend the resurvey of random segments of the 2019 and 2020 conferences to ensure the integrity of our results. In addition, we also recommend the survey of the 2017 and 2018 conferences to see potential changes in the frequency of open science practices in comparison to more recent conferences.

References

AIED Staff. (n.d.). *International AIED Society*. Iaied.org. https://iaied.org/about

Allen, C., & Mehler, D. M. A. (2019). Open science challenges, benefits and tips in early career and beyond. *PLOS Biology*, *17*(5), e3000246. https://doi.org/10.1371/journal.pbio.3000246

BERT. (n.d.). Huggingface.co. https://huggingface.co/docs/transformers/model_doc/bert

Charles W. Bailey, Jr., "What Is Open Access?, " in Open Access: Key Strategic, Technicaland Economic Aspects, ed. Neil Jacobs (Oxford: Chandos Publishing, 2006), 13-26.https://www.worldcat.org/title/254997540

Columbia University Irvine Medical Staff. (2021, January 31). *What is a preprint? Is it the right publishing choice for you? | Augustus C. Long Health Sciences Library*. Library.cumc.columbia.edu. https://library.cumc.columbia.edu/kb/what_is_preprint

Cornell University Staff. (2019). *arXiv.org e-Print archive*. Arxiv.org. https://arxiv.org/

EdArXiv Staff. (2023). *EdArXiv Preprints*. Edarxiv.org; Center for Open Science. https://edarxiv.org/

EDM Staff. (n.d.). *educationaldatamining.org*. Educationaldatamining.org. https://educationaldatamining.org/

Gargouri, Y., Larivière, V., Gingras, Y., Carr, L., & Harnad, S. (2012). Green and Gold Open Access Percentages and Growth, by Discipline. *ArXiv*. arXiv. https://doi.org/10.48550/arXiv.1206.3664

Google. (n.d.). *Google Docs: Free Online Documents for Personal Use*. Www.google.com. https://docs.google.com/document

L@S Staff. (2023). *L@S2023 – Learning@Scale 2023*. Learning@Scale. https://learningatscale.acm.org/las2023/

LAK Staff. (2011). *International Conference on Learning Analytics & Knowledge (LAK) - Society for Learning Analytics Research (SoLAR)*. Society for Learning Analytics Research (SoLAR). https://www.solaresearch.org/events/lak/

OSF Staff. (2023). *OSF*. Osf.io; Center for Open Science. https://osf.io/registries

PLOS. (n.d.). *Preprints*. PLOS. https://plos.org/open-science/preprints/

Qualtrics. (2015). *The Leading Research & Experience Software | Qualtrics*. Qualtrics. https://www.qualtrics.com/

R Core Team. (2022). *R: The R Project for Statistical Computing*. R-Project.org. https://www.r-project.org/

scikit-learn. (2019). *scikit-learn: machine learning in Python*. Scikit-Learn.org. https://scikit-learn.org/stable/

SciPy. (2020). *SciPy.org — SciPy.org*. Scipy.org. https://scipy.org/

Spellman, B., Gilbert, E., & Corker, K. S. (2017). Open science: What, why, and how.

TensorFlow Staff. (2022, December 15). *Word2Vec | TensorFlow Core*. TensorFlow. https://www.tensorflow.org/tutorials/text/word2vec

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

Watson, M. (2015). When will “open science” become simply “science”?. *Genome Biology*, *16*(1). https://doi.org/10.1186/s13059-015-0669-2

Appendix

Table 1. This is a table of the data collected from phase 1 of the project, only consisting of data from EDM and L@S.

Conference	Total Papers	Open Methodology	Open Data	Open Materials	Reproducible
EDM	208	208	41	31	2
L@S	93	16	1	5	0
Total	301	224	42	36	2

Table 2. This is a table of the data collected from phase 2 of the project, consisting of data from AIED, EDM, and L@S.

Conference	Total Papers	Open Methodology	Open Data	Open Source	Source and Data
AIED	127	0 (All available)	5	6	1
EDM	414	414	82	46	23
L@S	0	0	0	0	0
Total	541	414	87	52	24

Open Methodology

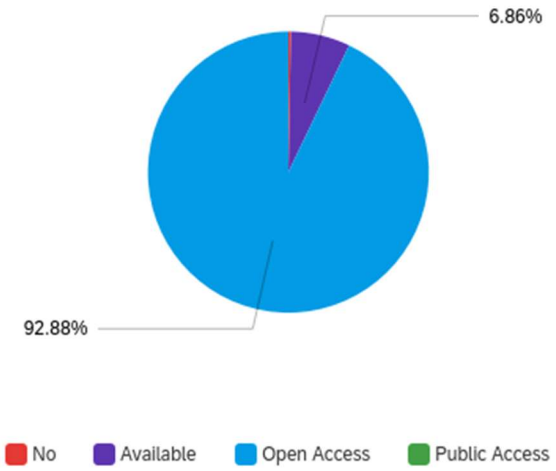

Figure 1. This chart indicates the percentage of papers in the Open Science Peer Review Survey regarding the openness of the methodology. All papers fall into the categories of No, Available, Open Access, or Public Access.

Open Data

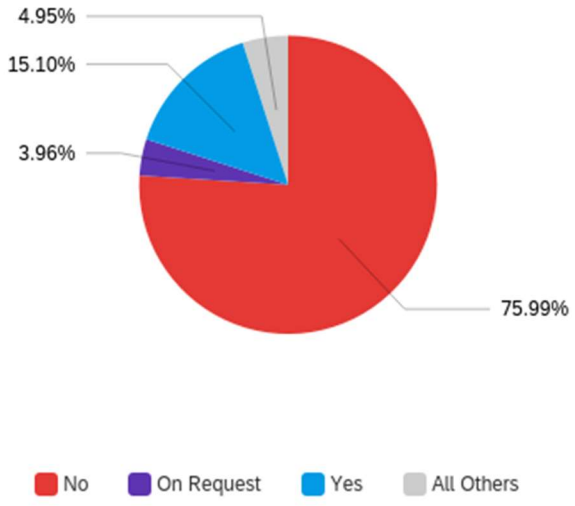

Figure 2. This chart indicates the percentage of papers in the Open Science Peer Review Survey regarding the openness of the data. All papers fall into the categories of No, On Request, Yes, Exceeded Time Limit, Meta-Reviewer Input Needed and N/A.

Data Documentation

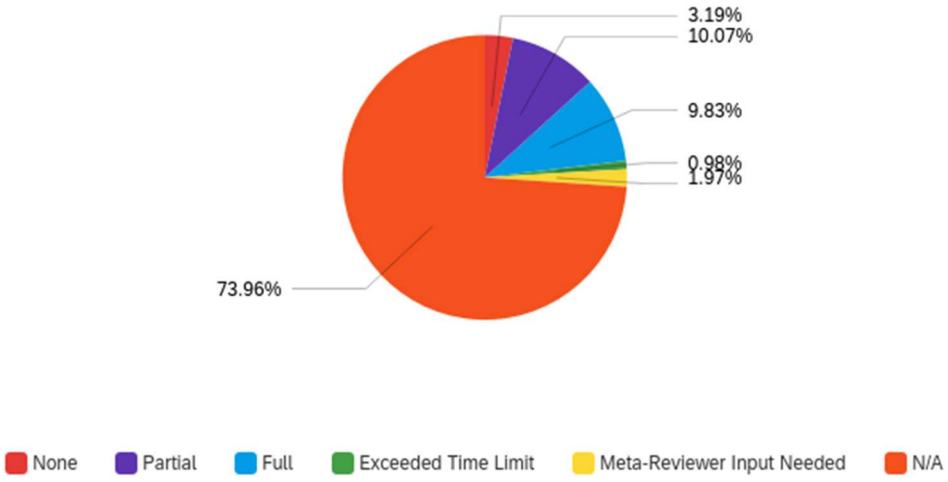

Figure 3. This chart indicates the percentage of papers in the Open Science Peer Review Survey regarding the quality of the data documentation. All papers fall into the categories of None, Partial, Full, Exceeded Time Limit, Meta-Reviewer Input Needed and N/A.

Open Materials

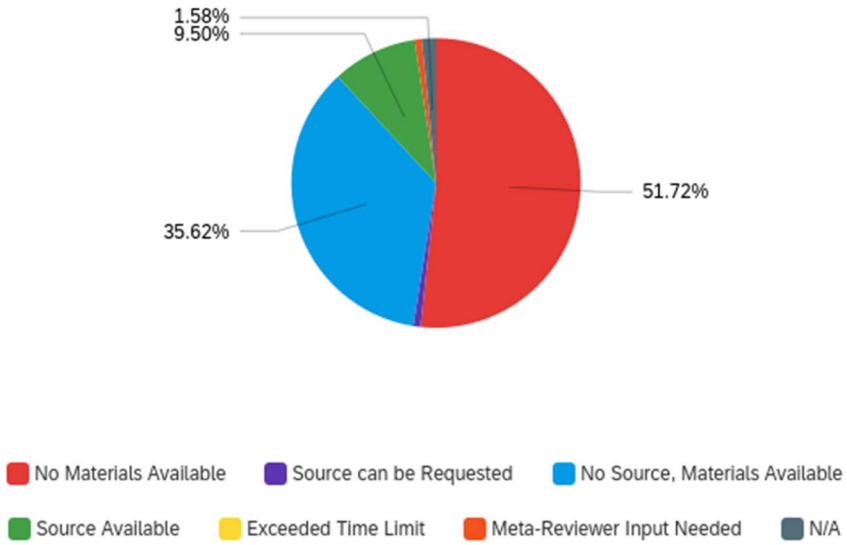

Figure 4. This chart indicates the percentage of papers in the Open Science Peer Review Survey regarding the openness of the materials. All papers fall into the categories of No Materials Available, Source can be Requested, No Source but Materials Available, Source Available, Exceeded Time Limit, Meta-Reviewer Input Needed and N/A.

Source Documentation

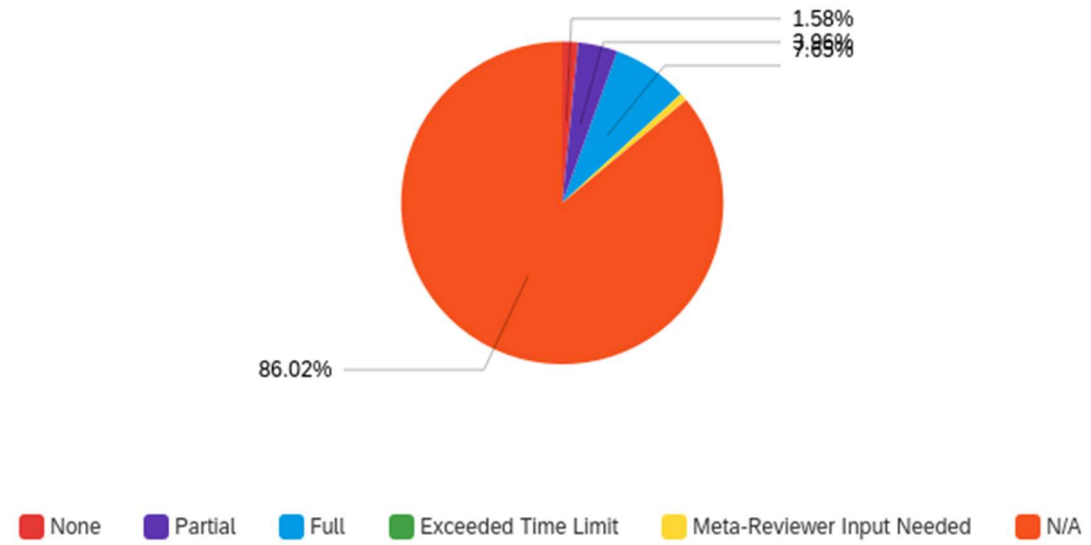

Figure 5. This chart indicates the percentage of papers in the Open Science Peer Review Survey regarding the quality of the source documentation. All papers fall into the categories of None, Partial, Full, Exceeded Time Limit, Meta-Reviewer Input Needed and N/A.

Source README

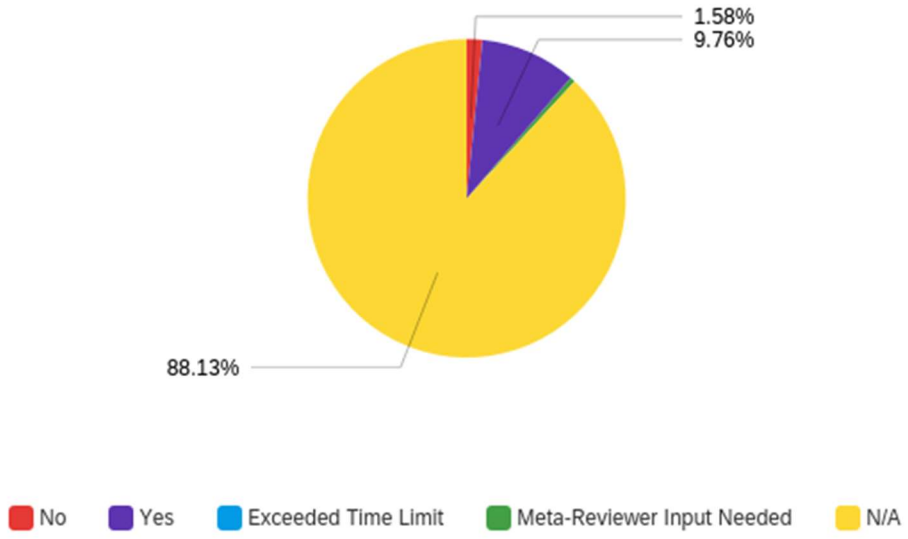

Figure 6. This chart indicates the percentage of papers in the Open Science Peer Review Survey regarding the existence of a README. All papers fall into the categories of No, Yes, Exceeded Time Limit, Meta-Reviewer Input Needed and N/A.

Data License

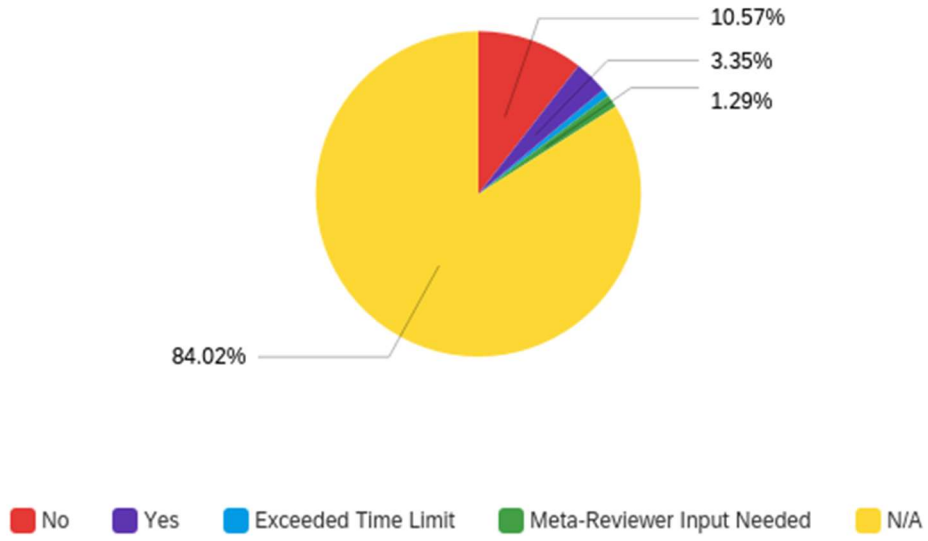

Figure 7. This chart indicates the percentage of papers in the Open Science Peer Review Survey regarding the existence of a data license. All papers fall into the categories of No, Yes, Exceeded Time Limit, Meta-Reviewer Input Needed and N/A.

Preregistration

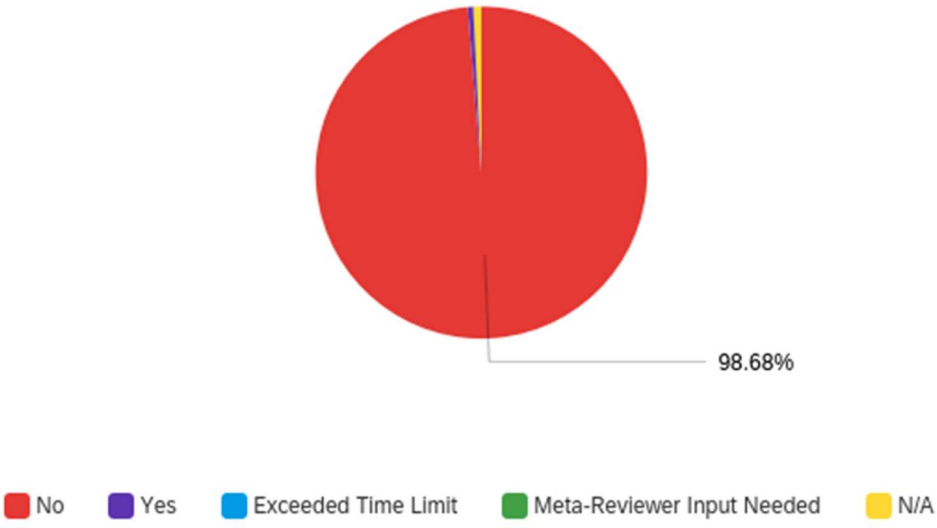

Figure 8. This chart indicates the percentage of papers in the Open Science Peer Review Survey regarding the existence of a preregistration. All papers fall into the categories of No, Yes, Exceeded Time Limit, Meta-Reviewer Input Needed and N/A.

Reproducible

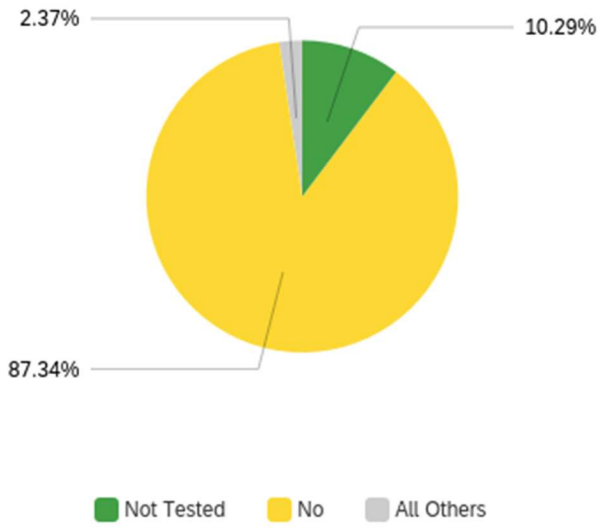

Figure 9. This chart indicates the percentage of papers in the Open Science Peer Review Survey regarding reproducibility. All papers fall into the categories of Partially, Fully, Not Tested, No and N/A.

Source License

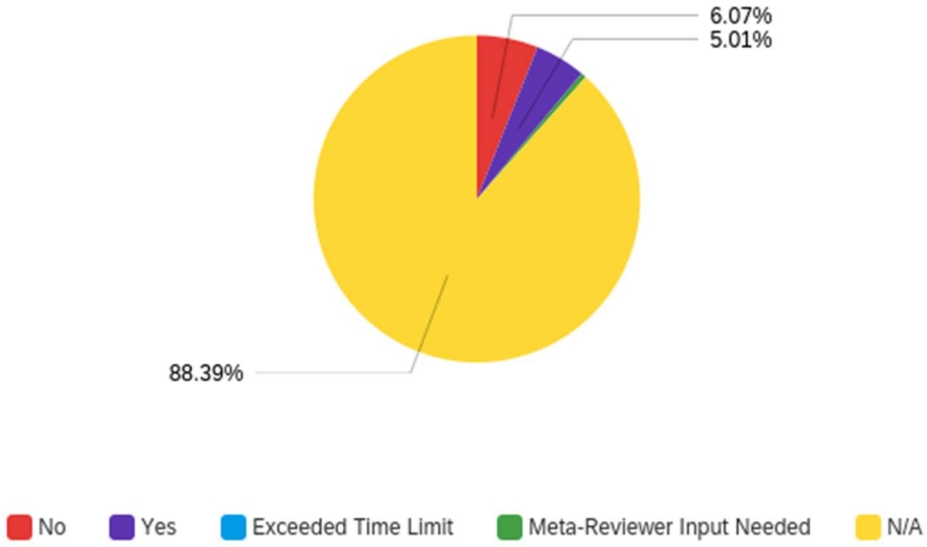

Figure 10. This chart indicates the percentage of papers in the Open Science Peer Review Survey regarding the existence of a source license. All papers fall into the categories of No, Yes, Exceeded Time Limit, Meta-Reviewer Input Needed and N/A.

Preprints

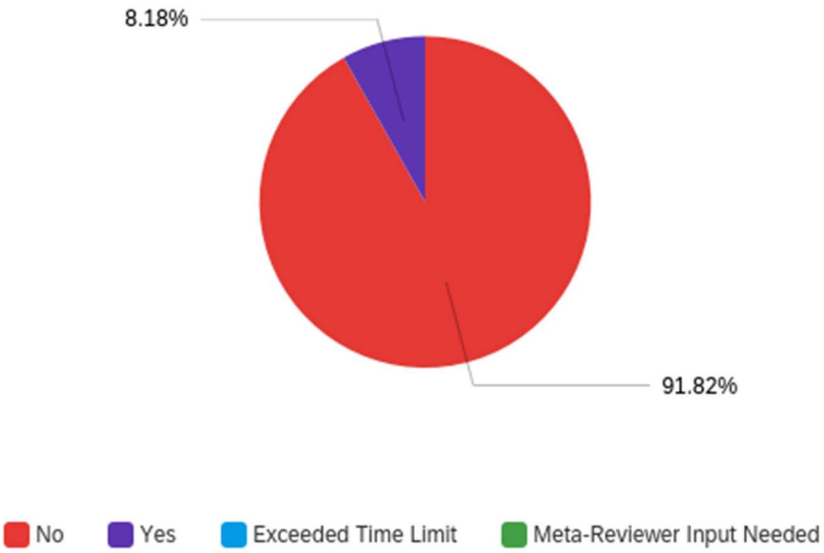

Figure 10. This chart indicates the percentage of papers in the Open Science Peer Review Survey regarding the existence of a preprint. All papers fall into the categories of No, Yes, Exceeded Time Limit and Meta-Reviewer Input Needed.

Open Methodology

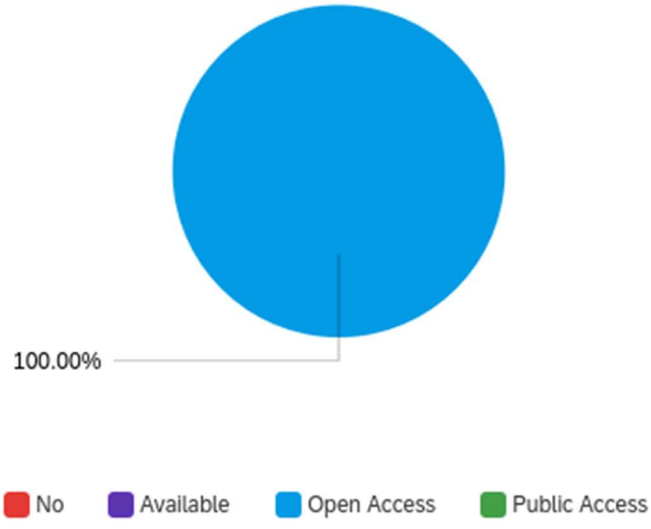

Figure 11. This chart indicates the percentage of papers in the EDM OS Survey regarding the openness of the methodology. All papers fall into the categories of No, Available, Open Access, or Public Access.

Open Data

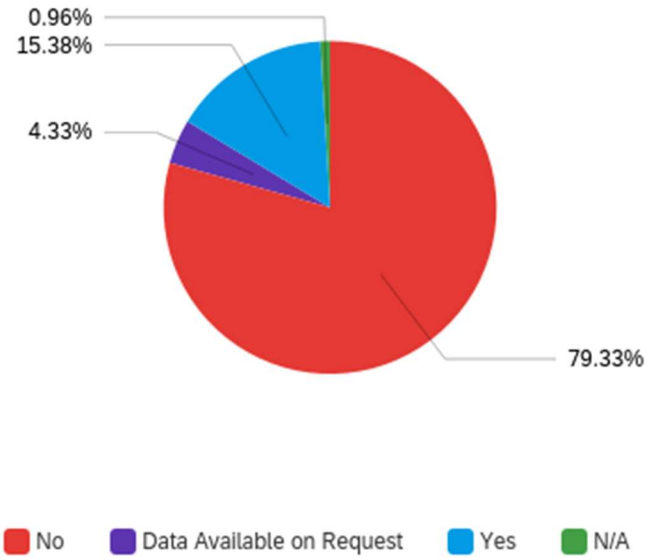

Figure 12. This chart indicates the percentage of papers in the EDM OS Survey regarding the openness of the data. All papers fall into the categories of No, On Request, Yes and N/A.

Data Documentation

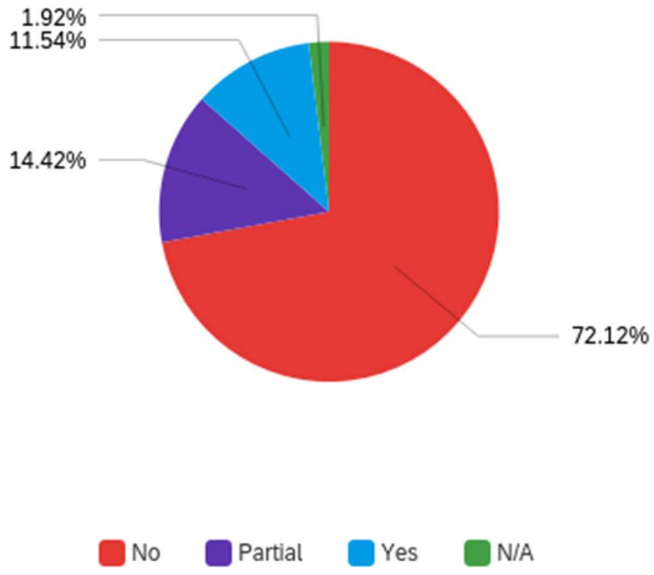

Figure 13. This chart indicates the percentage of papers in the EDM OS Survey regarding the quality of the data documentation. All papers fall into the categories of No, Partial, Yes and N/A.

Open Materials

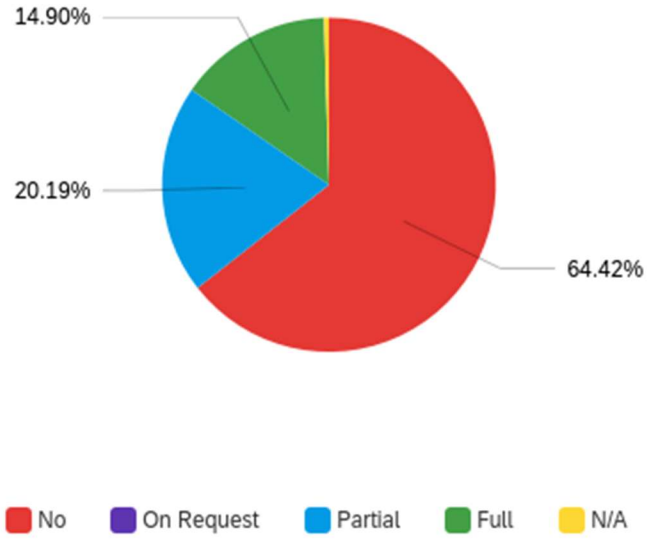

Figure 14. This chart indicates the percentage of papers in the EDM Survey regarding the openness of the materials. All papers fall into the categories of No, On Request, Partial, Full and N/A.

Materials Documentation

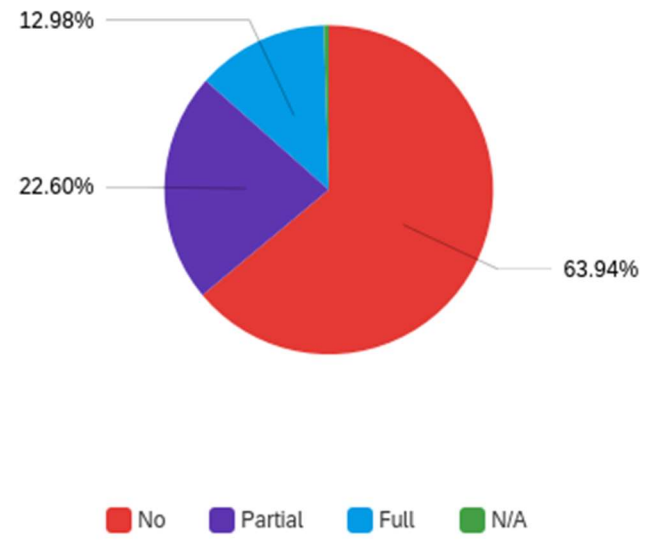

Figure 15. This chart indicates the percentage of papers in the EDM OS Survey regarding the quality of the materials documentation. All papers fall into the categories of No, Partial, Full and N/A.

README

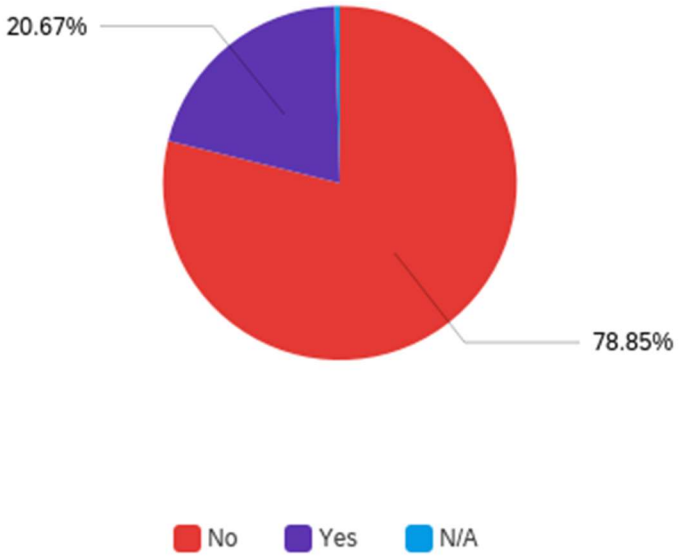

Figure 16. This chart indicates the percentage of papers in the EDM OS Survey regarding the existence of a README. All papers fall into the categories of No, Yes and N/A.

License

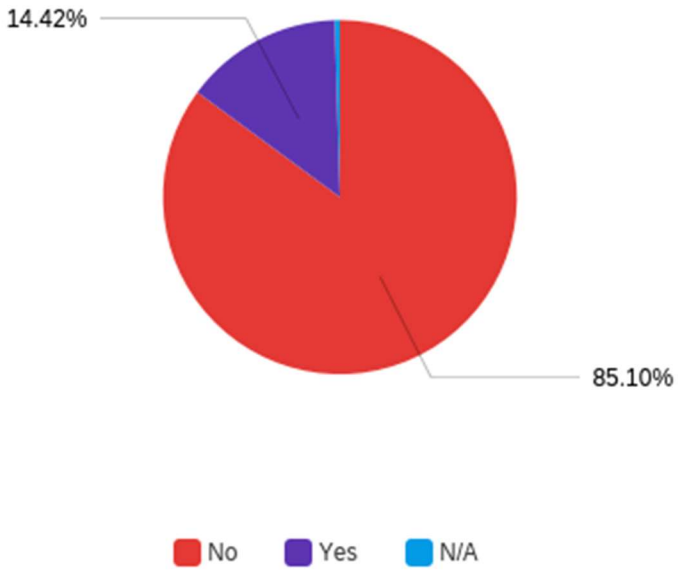

Figure 17. This chart indicates the percentage of papers in the EDM OS Survey regarding the existence of a source license. All papers fall into the categories of No, Yes and N/A.

Preregistration

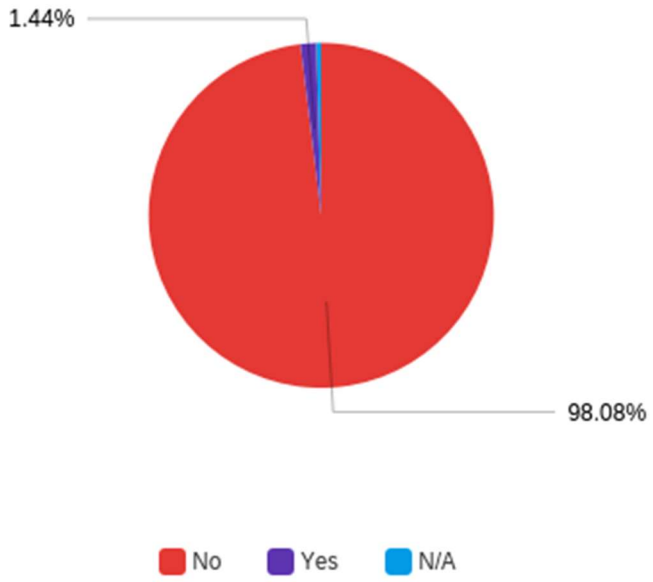

Figure 18. This chart indicates the percentage of papers in the EDM OS Survey regarding the existence of a preregistration. All papers fall into the categories of No, Yes and N/A.

Reproducible

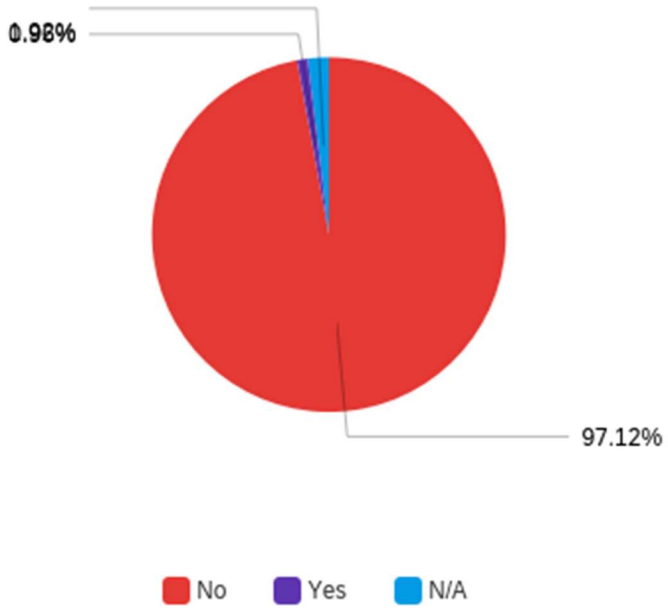

Figure 19. This chart indicates the percentage of papers in the EDM OS Survey regarding reproducibility. All papers fall into the categories of Yes, No and N/A.

Open Methodology

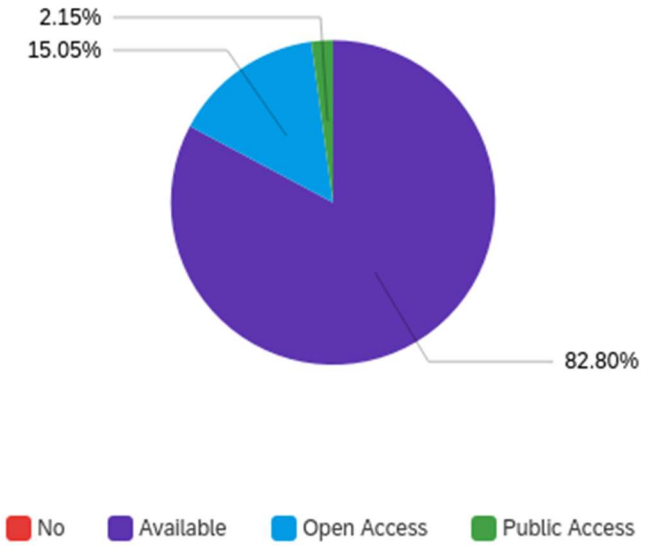

Figure 20. This chart indicates the percentage of papers in the Learning At Scale OS Survey regarding the openness of the methodology. All papers fall into the categories of No, Available, Open Access, or Public Access.

Open Data

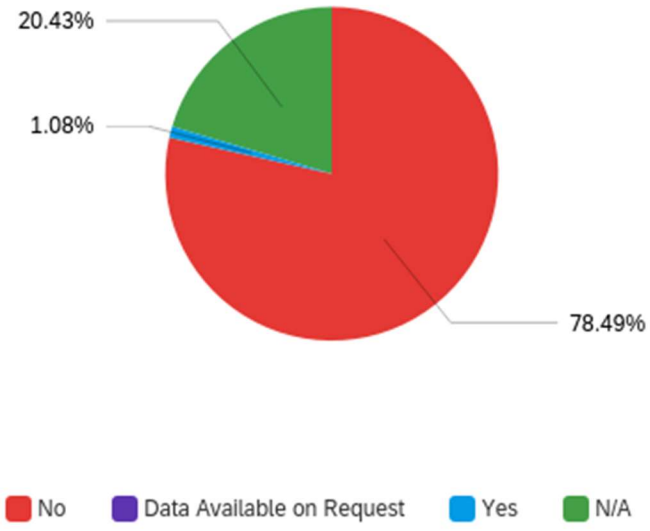

Figure 21. This chart indicates the percentage of papers in the Learning at Scale OS Survey regarding the openness of the data. All papers fall into the categories of No, On Request, Yes and N/A.

Data Documentation

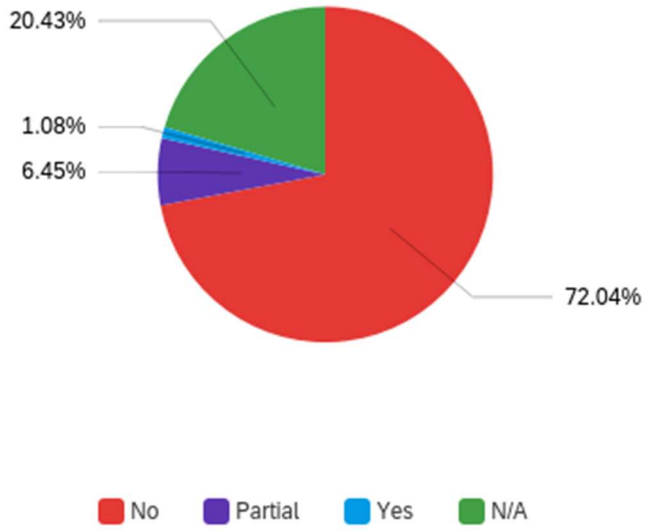

Figure 22. This chart indicates the percentage of papers in the Learning at Scale OS Survey regarding the quality of the data documentation. All papers fall into the categories of No, Partial, Yes and N/A.

Open Materials

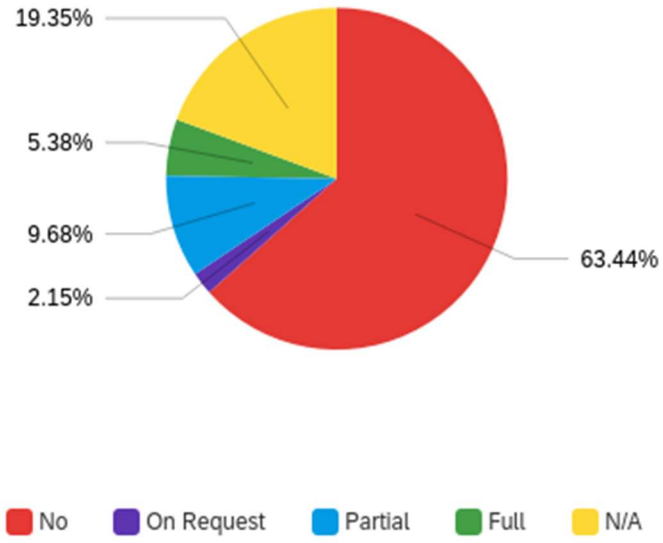

Figure 23. This chart indicates the percentage of papers in the Learning at Scale OS Survey regarding the openness of the materials. All papers fall into the categories of No, On Request, Partial, Full and N/A.

Material Documentation

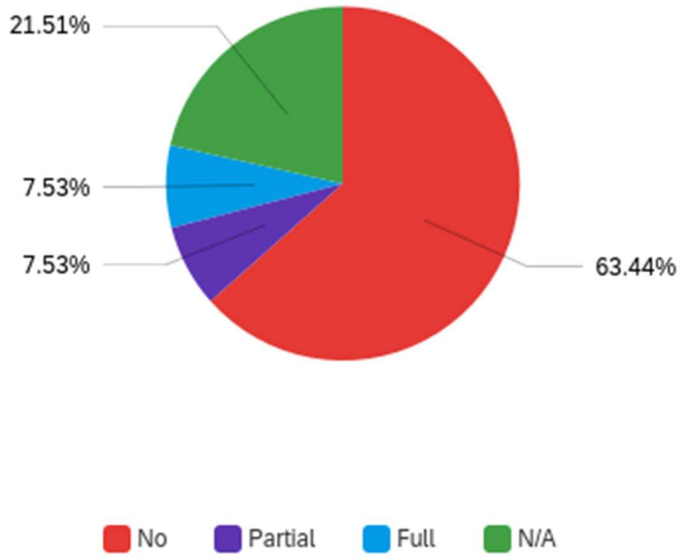

Figure 24. This chart indicates the percentage of papers in the Learning at Scale OS Survey regarding the quality of the materials documentation. All papers fall into the categories of No, Partial, Full and N/A.

README

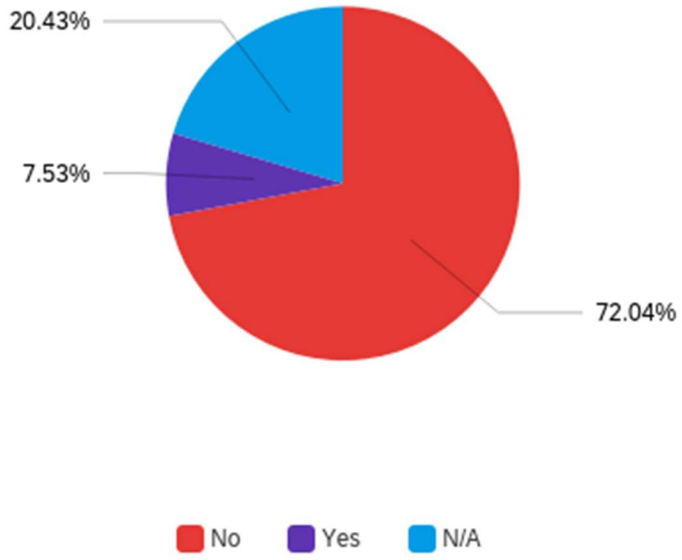

Figure 25. This chart indicates the percentage of papers in the Learning at Scale OS Survey regarding the existence of README. All papers fall into the categories of No, Yes and N/A.

Source License

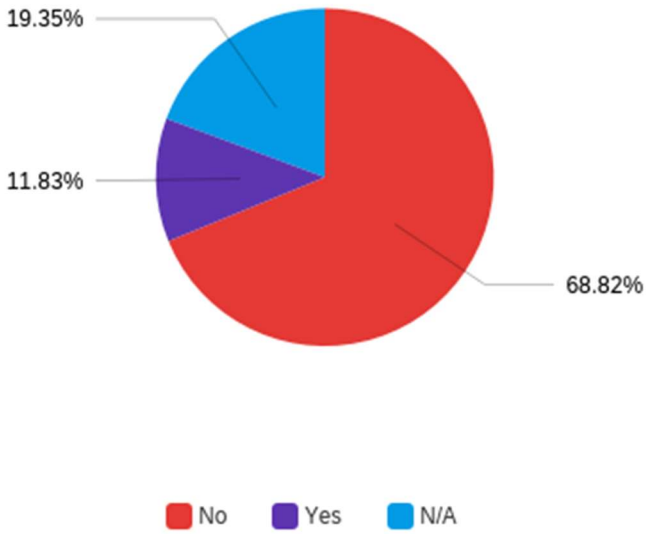

Figure 26. This chart indicates the percentage of papers in the Learning at Scale OS Survey regarding the existence of a source license. All papers fall into the categories of No, Yes and N/A.

Preregistration

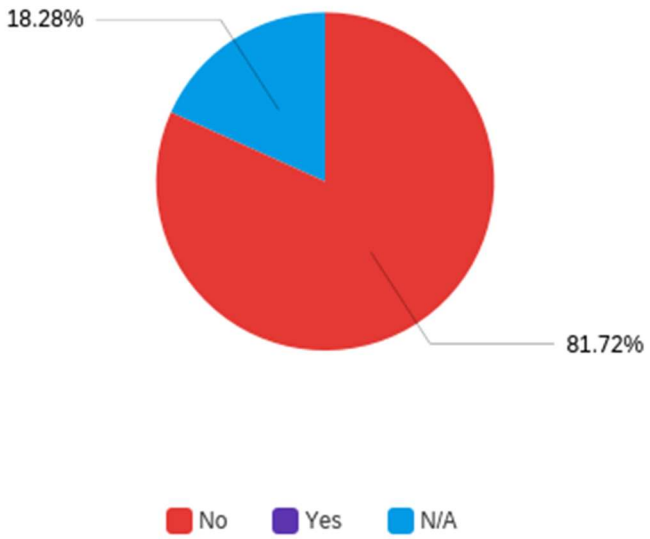

Figure 27. This chart indicates the percentage of papers in the Learning at Scale OS Survey regarding the existence of a preregistration. All papers fall into the categories of No, Yes and N/A.

Preprint

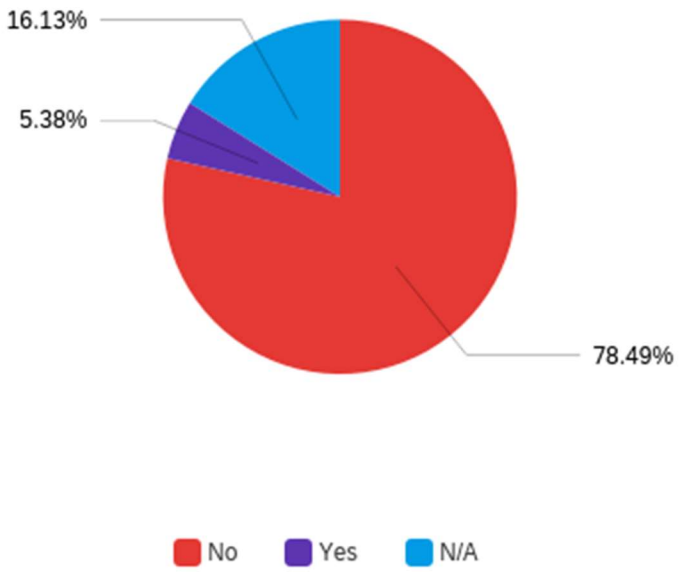

Figure 28. This chart indicates the percentage of papers in the Learning at Scale OS Survey regarding the existence of a preprint. All papers fall into the categories of No, Yes and N/A.

Reproducible

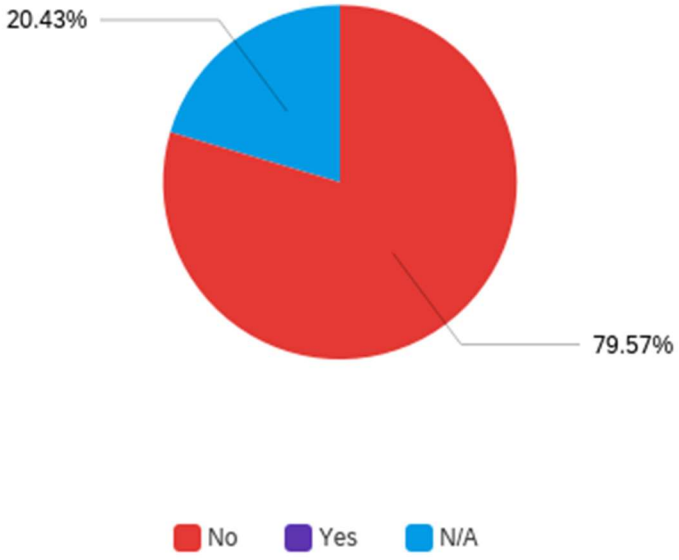

Figure 29. This chart indicates the percentage of papers in the Learning at Scale OS Survey regarding the reproducibility. All papers fall into the categories of Yes, No and N/A.