# Vis Repligogy: Towards a Culture of Facilitating Replication Studies in Visualization Pedagogy and Research

Uzma Haque Syeda, Laura South, Justin Raynor, Liudas Panavas, David Saffo, Thomas Morriss, Cody Dunne, and Michelle A. Borkin

Khoury College of Computer Sciences, Northeastern University, Boston, Massachusetts, USA

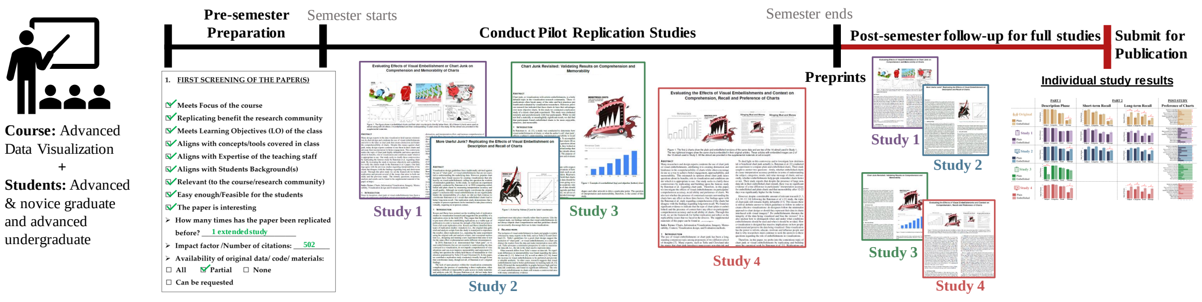


**Figure 1:** *Proof-of-concept implementation of our Vis Repligogy Framework implemented with Model 2.*


## Abstract

*In this paper, we present the Vis Repligogy framework that enables conducting replication studies in the class. Replication studies are crucial to strengthening the data visualization field and ensuring its foundations are solid and methods accurate. Although visualization researchers acknowledge the epistemological significance of replications and their necessity to establish trust and reliability, the field has made little progress to support the publication of such studies and, importantly, provide methods to the community to encourage replications. Therefore, we contribute Vis Repligogy, a novel framework to systematically incorporate replications within visualization course curricula that not only teaches students replication and evaluation methodologies but also results in executed replication studies to validate prior work. To validate the feasibility of the framework, we present case studies of two graduate data visualization courses that implemented it. These courses resulted in a total of five replication studies. Finally, we reflect on our experience implementing the Vis Repligogy framework to provide useful recommendations for future use. We envision this framework will encourage instructors to conduct replications in their courses, help facilitate more replications in visualization pedagogy and in research, and support a culture shift towards reproducible research. Supplemental materials of this paper are available at* `https://osf.io/ncb6d/`.



## CCS Concepts

• ***Human-centered computing*** → ***Visualization theory and methods;*** *Visualization pedagogy;*


## 1. Introduction

Data visualization is a young and growing field with many unresolved, debated, and untested theories, methods, and guidelines that are yet to be tested, validated, and strengthened. The rapidly growing popularity of the field has instigated many institutions to offer visualization courses, resulting in more students and therefore, more visualization researchers. Historically, visualization researchers rarely question or evaluate previously published literature (e.g., [KH18, MD20]). Kosara postulates that the field is "an empire built on sand" and that more replication studies are needed to strengthen its foundations and advance visualization towards being a real science [Kos16]. A **replication study** is a method to evaluate, validate, and expand the findings of a prior study by repeating it using the original, modified, or alternative approach [HSBAGS14, WEHE20]. Replication studies are the gold standard for investigating the credibility, rigor, and generalizability of previously published research [FS12, GL18]. These studies are said to be "at the heart of any science" [Lam90] as they enable building a reliable body of cumulative knowledge in any research paradigm (e.g., [BID*14, ACDF*13]).

Although researchers widely acknowledge the epistemological significance of replications, these studies remain underappreciated and disincentivized as unoriginal and not contributing anything new in most research fields (e.g., [SM18, HMS16]). Replications are perceived to carry less weight than novel research in most publication venues and are, therefore, extremely hard to publish (e.g., [Kos16, KH18, WCRC14]). This pattern has led to a *replication crisis* in physics [BI15], psychology [Col15], and HCI (e.g., [EH18, Kom22, WEHE20]) where the community fails to replicate prior studies independently. Failure to address this crisis means a field cannot trust its theories, methods, and assumptions.

There have been some efforts to make studies more replicable through encouragement to share research artifacts like data and source code



DOI: 10.2312/eved.20241054

[Kom22], pre-registrations [CGD18], and registered reports and conducting post-publication peer reviews [Pet18]. However, despite these efforts, the visualization research community has made little progress in supporting the publication of replications and providing replication methods to the community. Some initiatives to motivate researchers to conduct more replication studies, such as RepliCHI [WCRC14, WRCC13], and the CHI 2018 SIG on Transparency and Openness Promotion guidelines [CP18] have either been discontinued or not progressed since their incarnation. Therefore, despite advocacy for more replication studies in HCI and data visualization (e.g., [WRCC13, HSBAGS14, Kos16, KH18]), researchers are not motivated and are reluctant to conduct these studies to survive in the "publish or perish" research culture [Hor15]. This lack of replications also affects the motivation for open science practices, as researchers see no value in putting the time and effort into sharing research artifacts that will most likely not be used (e.g., [WEHE20, EH18]).

To encourage more replications, the community needs more methods, resources, and incentives available to researchers to conduct these studies. We also need to educate students, i.e., future researchers, on how to conduct replication studies. The growing popularity of the data visualization field [DB18] has resulted in more institutions offering data visualization courses and, therefore, more learners [Wol15]. We have a unique opportunity to conduct replication studies within the classroom, as it will not only benefit the visualization community by validating theory and methods through replications but also benefit the students by giving them valuable research skills and appreciation for replications. Teaching replication studies in the classroom is not a new concept. It is practiced in psychology and cognitive science to encourage more replications (e.g., [dLAL*19, HSA*18, WBB*19]). Such efforts never permeated the field of data visualization and no endeavors have been made to incorporate these efforts from other fields into the field of visualization. Therefore, our research question is: *"How can we leverage visualization learners to conduct and report more replication studies to validate theory?"*

To address this research question and to take the first step toward a culture in the visualization field where replications are prevalent and appreciated, we contribute **Vis Repligogy: a framework to incorporate replication studies within visualization course curricula**. Besides getting hands-on practice with conducting replication studies and research in general, students learn the importance of replication studies and transparency in research through open-science practices. The aim of the framework is to motivate and teach learners and researchers to conduct replication studies—but not necessarily produce publishable work. As conducting replications in the classroom has been shown to benefit both students and the community in fields that overlap with data visualization (e.g., [HSA*18, WBB*19, H*14, SGL*14]), we anticipate similar benefits through incorporation of the Vis Repligogy framework into visualization courses. We present two case studies as proof of concept—two graduate courses that utilized the Vis Repligogy framework and had students conduct replication studies to validate prior work. Finally, we reflect on our experience of implementing the framework and provide additional actionable recommendations to support and encourage instructors to conduct replications in their courses. In summary, in this paper we contribute the **Vis Repligogy framework**, including **four implementation models** and a **detailed step-by-step implementation process** for systematically incorporating replication studies within visualization course curricula. We also contribute **reflections** on implementations of Vis Repligogy, and **recommendations** for future implementations of the framework.

## 2. Related Works

### 2.1. Replication Studies in HCI and Data Visualization

Replications are a key method for evaluating the credibility and validity of previous research (e.g., [FS12, GL18]). Replications are classified under different terminologies. In this paper, we adhere to the terminology consistent with Schmidt [Sch16] and Brandt et al. [BID*14] where *exact* replications are conducted by the original authors [Sch16], *close* replications are independent replications by other researchers aiming to replicate the original study as closely as possible [BID*14], and *conceptual* replications build upon the original research [Sch16]. The practice of replicating previous studies is uncommon in the field of HCI [WCRC14], as researchers are incentivized to publish novel research [HSBAGS14, WCRC14]. It is documented that the field of HCI contains very few replications [HSBAGS14, GB08]. Data visualization is similar, with a publication culture that makes it difficult to publish replication studies and thus makes replications rare [KH18, Kos16]. Sukumar and Metoyer [SM18] point out that most visualization replications are published in non-visualization venues and highlight the dire need for more replications in the field. To address this need, we contribute Vis Repligogy for rapidly conducting replication studies in a classroom and provide alternative incentives, i.e., pedagogy, to encourage more studies in visualization.

### 2.2. Teaching Replication Studies in the Classroom

Conducting replication studies in the classroom, including studies of publishable quality, is an existing concept in the fields of psychology (e.g., [WBB*19, HSA*18, dLAA*19, HSD*19]), cognitive science [dLAL*19], and economics [H*14]. For example, de Leeuw et al. [dLAL*19] integrated replications into a cognitive science research methods course for undergraduates and reported many benefits, including increased student interest, development of research skills, and appreciation for sound research practices. Wagge et al. [WBB*19] reports a meta-analysis of nine student replications of a single study using the Collaborative Replications and Education Project (CREP). CREP aims to produce high-quality direct replications by training psychology students and having them conduct replication projects in their classrooms that are then crowdsourced for meta-analyses. Course replication curricula include replications as a lab component [HSA*18, HSD*19], for programming and statistics [HSD*19], best practices of research [HSD*19], and HCI methods [Wil13]. In the field of psychology, there is a growing interest in teaching and conducting replication studies in undergraduate and graduate courses [FS12, WBL*19, SGL*14] and many have highlighted and demonstrated the benefits this practice has on both pedagogy and the research community (e.g., [SGL*14, FS12]). To our knowledge, no prior work or pedagogical methods for teaching replication studies in visualization courses exist. Frameworks like CREP requires a lot of resources and logistical support, something that is not well suited as a starting point for the visualization community that is struggling with conducting replications in the first place. To encourage more replication studies in visualization, we contribute the Vis Repligogy Framework, which, unlike similar frameworks in other fields [HSA*18], is more robust and provides a detailed step-by-step guide to conduct replications in the classroom.

## 3. Vis Repligogy Framework

This section presents the Vis Repligogy Framework for conducting replications as part of course curricula. The framework consists of a set of **implementation models** (shown in Fig. 2) and a **process model** (Fig. 3).

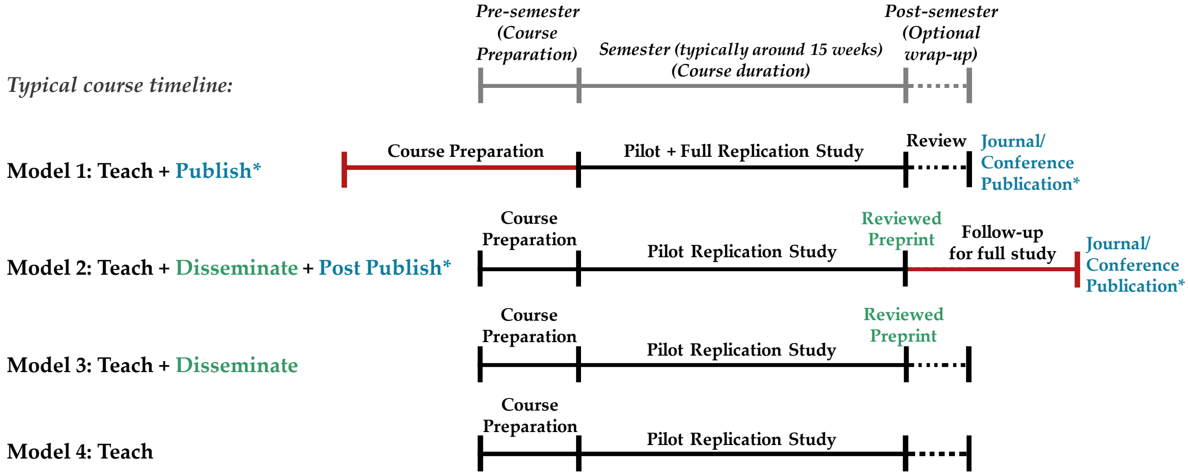


**Figure 2:** *Vis Repligogy Framework Implementation Models for integrating replications in a visualization course. Figure (not to scale) shows the course timeline for each of the four models. The asterisks * beside publications indicate that we cannot guarantee this outcome.*

### 3.1. Implementation Models

We contribute four models (Fig. 2) of the Vis Repligogy framework. Each model incorporates replication studies within a visualization course curriculum with different rigors, durations, benefits, and end goals. Instructors can choose the appropriate model to implement in their class based on their goal, class nature, number of students, and feasibility of implementation.

**Typical semester timeline:** Before delineating the four implementation models, we define some of the terminologies used throughout the paper. Based on a 3-semester-per-year typical American university system (i.e., Fall, Spring, and Summer terms), a typical semester duration (shown in Fig. 2) is approximately 15 to 18 weeks and consists of around 48 hours total class time. The pre-semester duration of a course is the time required to prepare and plan a course. The amount of time needed for this preparation will largely depend on the expertise and experience of the instructor and will take days to weeks. Besides an instructor's expertise on the course topics, the pre-semester duration will also vary depending on whether they have previously taught the course, implemented a course with Vis Repligogy, read the papers to select for a study/studies to replicate, planned the assignments previously so as to provide adequate theoretical and statistical support to students in their replication projects, and determined what type of IRB protocol to use. IRBs, or Institutional Review Boards, or more generally known as Ethics Review Boards examine research projects to make sure they abide by relevant laws, uphold generally recognized ethical standards, adhere to institutional norms, and protect study subjects as required. An optional post-semester wrap-up typically takes around a week but can take up to months if follow-up studies are involved (as in Model 2 in Fig. 2).

**Model 1: Teach + Publish*** is aimed at teaching and conducting publication-quality replication studies. In this model, the students under the supervision of the teaching staff, conduct both a pilot study (if applicable) and a full replication study within the duration of the course. The pilot (small-scale preliminary study) is optional and should take up no more than one-third of the semester and the remainder of the time should be allocated to run the full study. The end goal of the course project is submission to a journal or conference publication which means that the student projects need to be reviewed thoroughly by the instructor and the teaching assistant(s). Publication, however, is not a guarantee and this is indicated by the asterisks* in Fig. 2. Compared to a typical course timeline (shown in Fig. 2), this model has a longer course preparation duration to provide the instructor with adequate time to prepare to run full replication studies during the semester. Course preparation time can be used to prepare necessary assignments, materials, and resources to support the students in their replication project(s) and this will largely depend on the focus of the course and what needs to be taught/covered. It should also be used to apply for full IRB study protocol approvals and to make arrangements to initiate the study participant recruitment process(es).

**Model 2: Teach + Disseminate + Post Publish*** is aimed at teaching, disseminating, and then following up post-semester to conduct the full study towards the goal of publication. It should be noted that only the high-quality replication pilots from the given course should be disseminated and pursued for full study execution post-semester. The difference between this model and a typical course timeline is the longer post-semester duration for full-publication activities. During the semester, this model only aims to run a pilot/preliminary replication study and disseminate only the **reviewed** preprint(s) of the high-quality replication project(s). With this model, instructors will inform the students during the semester of the opportunity to conduct a full follow-up of the replication project(s). Interested students can participate in this follow-up full study and work towards potential publication. We acknowledge that running full studies with students post-semester may not be viable in many cases. Students may not be available to commit the time needed for the full study or may simply not be interested. However, if students are interested but do not have the time to commit to the full study, it should be noted that it is not necessary that the follow-up study take place immediately after the semester ends. The follow-up study could be conducted anytime after the semester ends which provides more flexibility to the students' busy schedules. If applicable and possible, these post semester follow-up can also be funded by the instructor to provide external motivation to the students. In cases where no student is available to conduct the follow-up studies, the instructor can simply migrate to Model 3 of the framework.

**Model 3: Teach + Disseminate** is aimed at teaching and dissemi-


Proceedings published by Eurographics - The European Association for Computer Graphics.

nating **reviewed** (and possibly winnowed) high-quality pilot replication studies. This model is the same as Model 2 with the exception of not having a follow-up post-semester for full publication. The end goal of this model is just to disseminate **reviewed and possibly winnowed preprint(s)** with relevant keywords. Therefore, Model 3 can be fully implemented within a typical course timeline.

**Model 4: Teach** aims only to teach students about replications and how to execute them. This model, like Model 3, can be implemented within a typical course timeline. The pedagogy in this and the other models can include how to conduct replication studies, the values of practicing open science, and overall best practices of ethical and sound study design and research practices. The students gain hands-on experience in research and appreciate the value of replicable studies [SGL*14, FS12]. Although this model can be applied to other study disciplines, data visualization courses should scope the course materials to visualization-related topics and research that are relevant to the focus of their course.

Beyond the immediate deliverables of each model and pedagogical learning objectives, the framework has additional **motivation and impact**. Teaching replications in class have been reported to benefit both the students and the research community (e.g., [SGL*14, FS12]). Models 1 and 2 enable students to potentially publish and also train them in academic writing. All of the models have the potential to directly benefit the research community by either publishing replications (Models 1 and 2) or disseminating reviewed high-quality preprints of pilot replications (Models 2 and 3) to its corpus of knowledge. These models (Models 1, 2, and 3) also provide intrinsic motivation to the students since the stakes of their project(s) go beyond just a course grade. Although Model 4 ranks the lowest in terms of the impact on the research community, we assert based on our own experience and the experience of others in psychology, that this model will indirectly benefit the research community by teaching and instilling the importance of replications, and the value of sound research practices (such as open sharing of research artifacts), amongst the next generation of researchers who will go on to become more cognizant of these issues (e.g., [SGL*14, FS12]).

### 3.2. Framework Implementation Process

To make the framework models actionable, here we contribute a detailed implementation process with step-by-step guidelines (Fig. 3) on how an instructor can implement the Vis Repligogy Framework in their course. The framework implementation process is based on our experience, reflection, and lessons learned and is inspired by the potential models for replication in the classroom by Hawkins et al. 2018 [HSA*18]. Instructors can use the provided process guidelines to make informed decisions about how to conduct the replications in their classes.

**Step 1: Choose a Level of Course** – The first step is to determine the level of the course from the following options: (1) A non-introductory/advanced course, or (2) An introductory course. Either course type can be focused on a specific visualization topic (e.g., "evaluations") or a general topic (e.g., "information visualization").

**Step 2: Choose Level of Students** – The level of the course selected will determine the level of the students in the course. We define the student experience levels as follows: Experienced/advanced graduate students (i.e., at least 2 years of post-graduate education and research experience), novice graduate students (i.e., in their first or second year with little or no research experience), advanced undergraduate students (i.e., within the last one or two years of their degree program with possibly some research experience), and novice undergraduates (i.e., minimal to no prior experience in computer science typical within the first one or two years of the degree program). A graduate student is someone with an undergraduate qualification and can either be a Masters or PhD student. As presented in Fig. 3, these student levels are grouped together in combinations to line up appropriately with the different framework models.

Note that for an advanced course, instructors can choose a combination of advanced/novice graduates and undergraduates (Step 2 in Fig. 3), although choosing advanced/novice undergraduates is highly unlikely. We recommend instructors carefully assess the background (students' degree programs and research experience) and expertise (relevant technical and soft skills) of advanced/novice undergraduates if they wish to include them in such courses or override prerequisites for special cases. For an introductory course, a combination of novice graduates and novice/advanced undergraduates can be chosen. In general, the course prerequisites (i.e., minimum required skills or previously completed courses required to take the course) can serve as an effective filter to enroll students with the minimum level of skills/background needed for the course (e.g., previous statistics or experimental methods course).

**Step 3: Choose Appropriate Model** – The next step is to identify the appropriate implementation model (Fig. 2). **If advanced graduate students** are selected in Step 2, then Models 1 or 2 are recommended. Although Models 3 and 4 could also be selected with an advanced student population, we recommend aspiring to contribute more replication studies to the visualization community by selecting Models 1 or 2.

**If a combination of novice/advanced graduates and advanced undergraduates is selected**, then either Model 2 or 3 are the recommended choices. Model 4 technically can be chosen but is not recommended here for the same reasons explained previously (shown with a thin line in Fig. 3). We caution against implementation of Model 1 as novice graduates and advanced undergraduates may not have the research skills or experience to successfully conduct a replication within a semester although there is evidence of such instances in the field of psychology [WBB*19, FS12].

**If a combination of novice graduate and novice/advanced undergraduates or just undergraduates is chosen**, then we recommend choosing Model 4. For undergraduate students, the arrow in Fig. 3 is dashed because although a common practice in psychology, we want to exercise caution since this concept is new in visualization pedagogy. Models 1 and 2 are not provided as options here because novice students will likely find it too overwhelming to conduct full or even pilot replications with high stakes of disseminating or possibly publishing them. However, with a combination of novice graduates and novice/advanced undergraduates, Model 3 may be attempted with caution.

**Step 4: Choose Mode of Paper Selection** – In this step, the instructor selects a mode of paper selection. Students can work on the replication projects individually or in groups depending on the total number of students, and the complexity of the replication project(s). In the **first method**, **students/groups select a replication project based on their interest and the instructor approves** based on the focus and learning objectives of the class and also the feasibility, suitability, and need for replication for that particular work to the scientific community. This method is most suitable for Model 1 with advanced graduate students. It may also work for Models 2 and 3, with a mix of novice and advanced graduates and undergraduates, but the novices may find

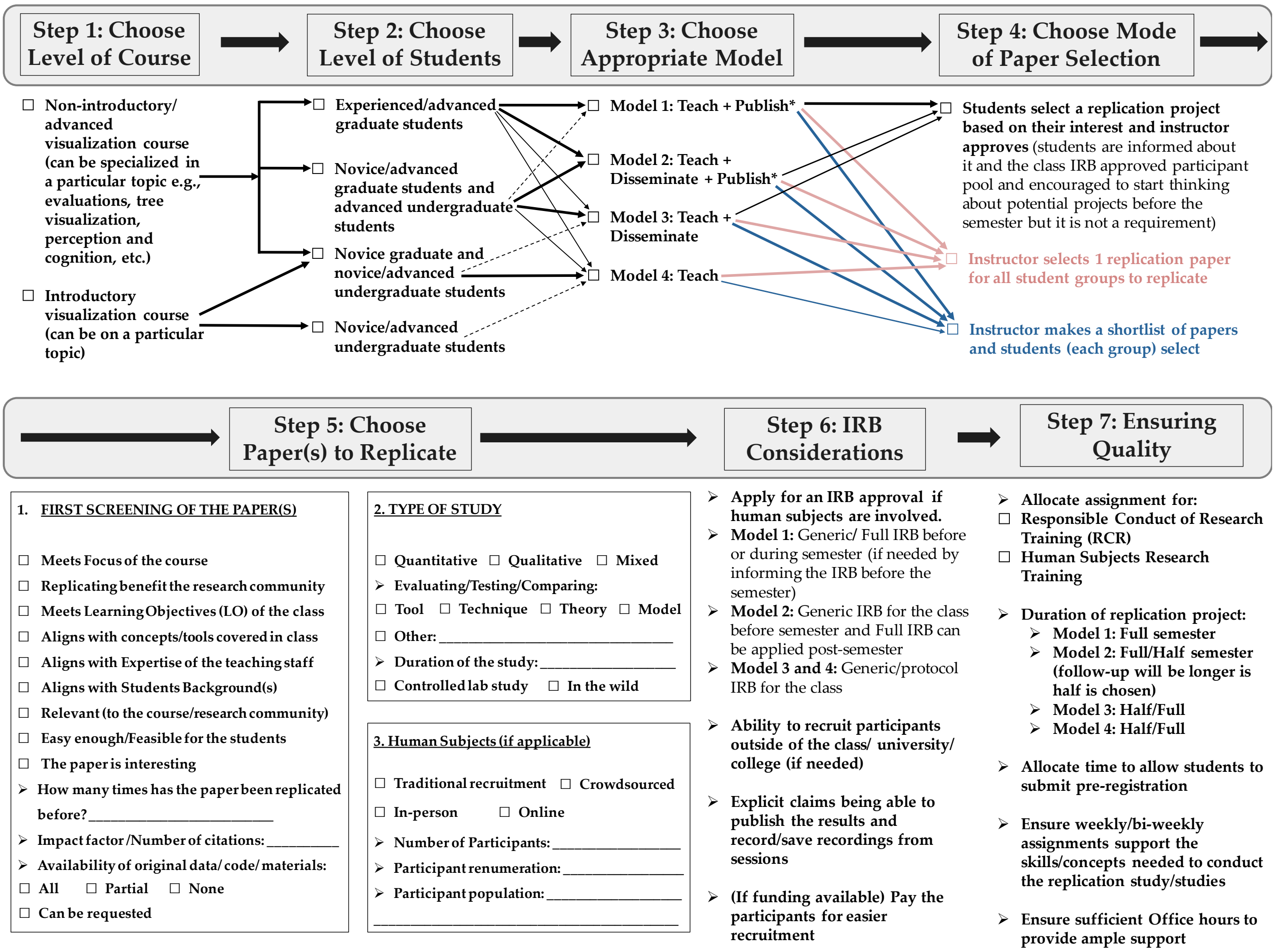


**Figure 3:** *Vis Repligogy Framework Process Model as a step-by-step guide. Solid arrows indicate the recommendations, thinner solid arrows indicate that it is acceptable but not ideal, and dashed arrows are not recommended. The color coding in Step 4 is to improve readability.*

it difficult to choose a study on their own (thinner arrow in Fig. 3). This method enables students to pick a study they are motivated to conduct and have more ownership. Additionally, if each student/group replicates a different study or conducts different replications of the same study, a greater quantity of replications are completed with great contribution potential to the community.

The **second method** is where the **instructor selects one replication paper for all students/groups to replicate**. This method is suitable for all 4 models. For Models 1, 2, and 3, multiple replications of the same study can result in cumulative evaluation and possible meta-analyses, although results from Model 3 will not be as rigorous as Models 1 and 2. The choice of replicating only one study also provides degrees of freedom to have different types and levels of replications (exact, conceptual, or close [BID*14] or a mix of those), and different experimental procedures to rigorously evaluate the study selected. This method also allows for more manageable grading and support for students, especially for larger classes or complex studies, as everyone will replicate the same study, and collective instructions will benefit all groups and save time. This method has great potential benefit to the community as it can run a single study multiple times, and possibly in multiple ways.

In the **third method**, the **instructor makes a shortlist of the potential replication papers** based on the learning objectives and focus of the class and the students/groups select based on their interests. We caution against using this with Model 4 as novice students might find it difficult to make an informed decision regarding which study to choose. This method results in a greater number of different replication studies and gives both the instructor more control as well as some sense of independent choice for the students.

**Step 5: Choose Paper(s) to Replicate** – When determining paper(s) to replicate in the class, considerations to balance research and pedagogy need to be kept in mind. This is indicated in Fig. 3 as the "First Screening of the papers" in Step 5. If a paper checks-off the first 9 checkboxes in Fig. 3, only then can it be considered to be replicated in the class. These criteria include meeting the course's focus and learning objectives, aligning with the expertise and background of the teaching staff and students, feasibility of replication within the constraints of the class, relevancy for the class and the community as a whole, and whether replicating the paper will benefit and inform the research community.

Other considerations in the first screening (shown in Step 5 of Figure 3) are left at the discretion of the instructor. Based on the focus and learning objectives of the course, the need and feasibility of replication, instructors can choose old seminal or new cutting-edge papers to replicate. As a starting point, they can refer to the research questions pointed out by Robert Kosara [Kos16] or the 16 replication studies listed by Sukumar and Metoyer [SM18]. Also, it solely depends on the instructor what skills they want the students to gain, and based on that they can choose a paper with all, partial, or no open materials. For example, if the instructor wants the students to easily conduct an exact/close replication then they should select papers with all of the study materials available. If the objective is to teach students research independence and critical thinking as well as instill the values of open science for replicable studies, then the instructor could choose paper(s) with unavailable or partially available materials. Student level and expertise should also be kept in mind when determining the trade-off between instructor control versus student independence. The type of paper chosen, its complexity, and the amount of its research artifacts openly available will influence these objectives directly and should be handled with caution, especially when the student population includes novices.

After the first screening, the instructor determines the type of study in the second part of Step 5. They select and check off the type of analyses, what is being evaluated, the total duration, and the environment of the study. If the study involves human subjects, then they will move on to the third part of Step 5. This checklist inquires about the environment of the study and the number and type of participants. Determining these will make it easier for the instructor to assess the fit and feasibility of executing the study in the class, and will aid in devising an appropriate IRB application in Step 6. If the chosen paper is not a good fit, this step can be iterated with alternative options.

**Step 6: IRB/Ethical Considerations** – Studies involving human subjects require IRB approval. Although IRB standards may differ across universities, we recommend applying for IRB approval well in advance during the pre-semester period, especially for models that aim to conduct full replication studies as there might not be enough time to get IRB approval within the duration of the semester. What type of IRB protocol should be applied will largely depend on the model selection in Step 3 and the paper selection method in Step 4. For all models, instructors should carefully consider for their IRB protocol the appropriate anticipated participant pool and the explicit declaration that the data collected might be saved and published.

**If students select replication studies on their own** and if Model 1, 2, or 3 is chosen then a standard generic and non-specific IRB for the class can be submitted before the semester. The instructor can then resubmit, if necessary, specific study amendments to the generic IRB at the very start of the semester once students select replication paper(s). For Model 2, which conducts the full replication study after the semester ends, further IRB amendments specific to the study can be submitted post-semester.

**If the instructor chooses one study for all or makes a shortlist for students to choose from**, then for Model 1 a full IRB protocol specific to the selected study should be submitted well ahead of the course (pre-semester). For Models 2, 3 and 4 a generic protocol IRB for the class applied pre-semester should suffice. However, for Model 2 alone, an amendment/full IRB should be submitted post-semester to run full studies. Note that the IRB protocol is easier if the instructor pre-selects a paper/pool of papers to be replicated.

**Step 7: Ensuring Quality** - To ensure the highest quality of work from the students, stringent quality control measures are necessary as mitigation of possible risk factors associated with student-led replications. We highly recommend ensuring that the students complete a Responsible Conduct of Research training course and, if human subjects are involved, a Human Subjects Research Training course to minimize concerns about responsible and ethical research. We also recommend choosing an appropriate timeline for the project to ensure the students have sufficient time to conduct the replication study(ies), i.e., a full semester for Model 1 and half/full semester for the other models. Along these lines, instructors should allocate time for pre-registrations, and assignments should be prepared to support the students' project timeline in a way that helps the students with the development or enhancement of skills needed to conduct the replication research. Planning out frequent (e.g., weekly or bi-weekly) replication project milestones to evaluate students' progress is also a good strategy to ensure students are spending sufficient time on the assignments and not procrastinating. It also allows the teaching staff to monitor students' progress and provide necessary feedback early on/ periodically. Additionally, sufficient office hours outside of class needs to be allocated to provide ample support for the students.

The instructor will be ready to plan and conduct replications in their class after completing the seven steps of the implementation process. We acknowledge that individual contexts and circumstances may differ, and therefore, mapping planned assignments and activities to the capability, capacity, expectations, and context of the course is an essential part of the instructor's role while implementing the Vis Repligogy framework. Allocating enough time for this and being thoughtful and flexible will also enhance the chances of successful implementation. Finally, we also highly recommend consulting and using the Replication Recipe by Brant et al. 2014 [BID*14] which outlines standard criteria for conducting high-quality close replications.

## 4. Proof-of-Concept Implementations

In order to validate the Vis Repligogy framework, we successfully implemented the framework using Models 1 and 2 in two different advanced graduate data visualization courses. In this section we discuss these two case studies and their outcomes. In both case study implementations, the papers selected for replication reinforced the courses' learning objectives of introducing replications and the importance of open science and aligned with the expertise of the teaching staff and the students' backgrounds. Additionally, neither paper has a published replication thus the potential for the students' work to contribute to the visualization community. Both papers included data and procedures that were partially or fully unavailable which reinforced to students the importance of open science. In Case Study 1, the instructor purposefully chose to have a paper with missing information in support of the course's learning objectives including critical thinking and problem solving skills. In Case Study 2, the student chose the paper due its significance in the field as well as opportunity to generate the missing procedures and data for the community. In both cases the papers enabled the students to explore the vast space of study design and analysis possibilities for an effective replication.

### 4.1. Case Study 1: Replications using Model 2

The Vis Repligogy framework was implemented in an advanced data visualization seminar course focused on evaluation studies in visualization research. The course, CS 7295 "Special Topics in Data Visualization: Evaluation Methodology, Techniques, and Applications", was taught in Fall 2020. The student population (n=7) consisted of three novice and three advanced graduate students, and one advanced undergraduate student. Based on the course and student type, Model 2 of the framework was selected (Fig. 3). The replication study assignment was conducted in the first half of the semester, and the instructor selected the seminal "chartjunk" paper by Bateman et al. [BMG*10] for the three student groups (n=3, n=2, and n=2) to closely replicate. This paper investigates the effects of visual embellishments on comprehension accuracy, memorability, and preference of charts. For the second half of the semester, the instructor gave the students the option to either pick for themselves a published evaluation to replicate, or to design and execute a novel evaluation study. One student group (n=2) chose to design and run a novel conceptual replication of the Bateman et al. [BMG*10] paper based on their interests and reflection of the lessons learned from the prior three replications conducted in the first half of the semester. The study conducted in the class was a pilot study for future post-semester follow-up work. Post-semester, six out of the seven students expanded the four pilot studies conducted in the course into four full replications with as many participants as the original study. The results of these studies validated the work by Bateman et al. [BMG*10] and provided further evidence that visual embellishments enhance memorability. These replication studies are currently being prepared for submission for publication in a peer reviewed journal. The process of this case study is illustrated in Fig. 1 and the pilot project reports, the pre-registrations for the full studies, including the results are available in the Supplemental Materials of the paper. As the timelines and procedures for Models 2, 3, and 4 are almost identical with the exception of the final goal, we assert that this Case Study 1 implementation also provides validation for Models 3 and 4 (Fig. 2) of the framework.

### 4.2. Case Study 2: Replication using Model 1

The Vis Repligogy framework was implemented in another advanced data visualization course focused on advanced topics in visualization not included in the college's standard introductory visualization course. Case Study 2 was conducted in CS 7375 "Seminar in Human-Computer Interaction: Advanced Topics in Data Visualization", was taught in Fall 2023. The student population (n=10) consisted of eight novice and two advanced graduate students. Assessing the course and student type and goal for all student final project papers to lead to publications, Model 1 of the Vis Repligogy framework was selected (Fig. 3). In this course students again worked in small groups and, unlike the previous case study, the course project was conducted over the full duration of the semester. Although the importance of replications and open science were part of the course curriculum, the completion of a replication study was not required but rather one option to satisfy the final project requirement. (The other two project options were to complete a design study, or write a visualization survey paper.) One student group (n=2) opted to conduct a replication study of the computational evaluation by Matuszewski et al. [MSM99] that compares heuristics for crossing minimization in layered networks. One student in the group was an advanced student with expertise in network visualization. The students implemented all heuristics examined by Matuszewski and recreated the datasets used as the originals were all lost. The study results were mostly consistent with Matuszewski et al. [MSM99], even when extended to larger networks. This replication is currently being prepared for submission to a peer reviewed journal. The project report of this study is provided in the Supplemental Materials at `https://osf.io/ncb6d/`.

## 5. Reflections and Recommendations

In this section, we provide additional reflection and recommendations for instructors on how to effectively implement and use the framework. We assert that these seemingly simple suggestions and considerations can help streamline the implementation experience.

**Best practices for teaching replication studies in a visualization course:** Based on our experience, a number of pedagogical choices stood out as helping the students conduct their studies. First, *completion of Responsible Conduct of Research Training (RCR) and Human Subjects Research Training* as assignments prior to conducting the studies compelled the students to learn about ethical and responsible research involving human participants and made them eligible to conduct research involving human participants for future projects or courses. Second, *conducting pre-registration* [CDBG20] before the pilot studies as another assignment helped the students to internalize the effort and necessary details needed to complete such registrations. It also reinforced the value of making proper study design and analysis plans ahead of time to make their research process more transparent. Therefore, allocating time for these training and pre-registrations in the form of assignments will not only benefit the students but will also aid the teaching staff in monitoring students' progress and provide useful feedback and guidance on study design and analysis plans early on and make the process smoother.

**Determining the appropriate replication project duration:** For Model 1, the replication-based project should be the core of the class and span the entire semester (around 15 to 18 weeks) since a publication-quality journal/paper is the target. For Model 2, either a half or full semester can be dedicated to the replication project. In our experience with Model 2, we allocated the first half of the semester solely to replication projects, and the latter half of the semester was open to other types of projects, including replications. Reflecting back on our experience, we recommend allocating the second half of the semester to the replication project(s) as it will allow the instructor to map out assignments and lectures to teach, enhance, and/or help build necessary skills among students to increase their chances of success with the replication projects. For Models 3 and 4, it is left to the instructor's discretion how long they want the replication project to take up during the semester. A full semester should be allocated for larger classes or for studies that are more complex and need additional time. For simpler studies or studies that can be broken down into smaller parts, a half-semester should suffice. However, if possible, allocating the whole semester to conduct the replications as a semester-long project will inevitably allow more time to map out necessary assignments and enable the students to gain contextual knowledge, practice, experience, and more confidence in the topics of the course.

**Adjusting course content to optimize replication quality:** We found in our implementation that weekly and bi-weekly assignments throughout the semester proved advantageous to ensure the students were equipped with the necessary knowledge to tackle each assignment/replication milestone towards completion of the project. A mix of in-class practice, assignments, and project milestones can be mapped out to adequately support the students. Caution must be executed so as to not inundate the students with too many assignments and ensure teaching staff have sufficient time to provide feedback to students. Additionally, breaking the project down into manageable portions (e.g., pre-registration, analysis plan, conducting the experiment, conducting the analysis, writing the report/paper, and submitting supplementary materials) in the form of assignments also helped check for any academic integrity violations (e.g., cheating, plagiarism, etc.). We also recommend as an instructor would do in any university course, to review assignments for academic integrity and utilize plagiarism-checking tools. Platforms such as OSF


Proceedings published by Eurographics - The European Association for Computer Graphics.

can be utilized to submit pre-registrations and software such as Turnitin can be utilized to check for plagiarism in each assignment. Having students abide by open-science practices such as submitting analysis codes and data also shield them against possible academic misconduct. We also found that conducting regular in-class discussions about the replication projects was extremely helpful in clarifying confusion and sharing ideas and insights, and contributed positively to the success of the replication experimental procedures. Additionally, for larger classes, the pre-requisites of the course can be made more stringent in order to acquire students with a particular background or familiarity with some skillsets in order to adjust the course contents to focus more on the topics relevant to the replications and less on introductory concepts.

**Choosing and migrating between models:** This framework is designed from the instructor's point of view to pick a model knowing what type of class they will be teaching. Even if an instructor begins with one model (Fig. 2), they can still change their mind during or after the semester. For example, an instructor may choose to change models if the expertise of the enrolled students differs from what was expected, the students wish to pursue publication after a preprint, or the course timeline changes. *Migrating between Models 2, 3, and 4 is possible* as their timelines are essentially the same—with the exception of the post-semester step for Model 2. For example, an instructor starting with Model 4 can later migrate to Model 3 if they deem the preliminary pilot replications are of sufficient quality to disseminate. For courses with willing and able students, the instructor can migrate from Model 3 to 2 if the students agree to work post-semester to conduct a full study for publication. The reverse is also possible, in which the instructor switches from Model 2 to 3 when it is unfeasible to run a post-semester full study. If, in Models 1–3, the pilot studies do not meet the dissemination standards, the instructor can switch to Model 4. Note that it is *not possible or recommended to migrate between certain models*. Switching to Model 1 is not possible as it requires longer pre-semester preparation. Although migrating between models is possible, choosing one and not changing will help scope the focus of the course and make the process more streamlined. For larger classes, or with more complex replication studies, migrating to a lower model such as Model 3 or 4 will lower the stakes of the course and will make it more manageable while still acquiring the pedagogical benefits from conducting replications.

**Selecting a paper will depend heavily on the objectives of the course and the model chosen:** *For Models 3 and 4, we recommend selecting papers that include enough details to replicate.* As the primary goal of these models is pedagogy, and the students are likely novices, the selection of papers with incomplete or little implementation information may frustrate and confuse students. Models 1 and 2, offer more freedom to choose papers with full, partial, or no available information as these are more applicable for advanced students. If the instructor decides to select papers with partial or no replicable information, *adequate support and technical help from the teaching staff* needs to be ensured (especially for Models 3 and 4). From the perspective of *learning objectives*, instructors can also decide what type of papers to choose. If the learning objectives are geared toward conducting replications in general, then choosing a paper with partial or full information is recommended as this will help students learn from other researchers' examples. However, if the objectives are to build critical thinking skills and to train students to explore the vast decision space of study design, then papers with partial or no replicable information are well suited. If the course consists of mostly novice students, then the need to have easier and cleaner replications is important and this might limit the number of available studies. In this case, instead of selecting multiple replication projects, conducting conceptual replications of the same study can be more reasonable.

**Limitations and Future Work:** As Vis Repligogy was developed and piloted in small graduate courses, future studies need to investigate how well the framework scales to larger classes in accordance with the recommendations and guidance provided in this work. Further validation, and additions to the framework as needed, in larger courses will broaden the generalizability of the framework. How to use the framework to run replications in a pre-coordinated way across multiple semesters over time can also be investigated. The Vis Repligogy framework is dependent on the expertise and knowledge of the course instructor. Currently, whether it is the course instructor curating and choosing a paper for students to replicate or the students choosing their own paper, it is their burden to find possible replications of a potential selected paper. The process of finding replication studies is currently a major challenge. First, looking for studies that have been replicated is difficult as there is no specific consensus regarding the accepted terminology used for replication studies, and this makes determining the appropriate query term for such studies challenging [HSBAGS14]. Replication studies are rarely indicated as such in a paper's keywords or title. Instead, they are often wrapped and embedded in novel contributions to increase the chances of publication. Whether it is conducted in class or in research, we advocate that *the visualization community needs to have more streamlined publication standards and standardized terminology for replication studies*. Currently, the most systematic way to search for replications is to browse a set of publication outlets [QR19]. Moreover, most visualization replications are published in non-visualization venues [SM18] which adds to the complexity of the problem. Therefore, *we recommend that the visualization community creates indices of papers* that have been replicated (or not) and are good candidates for future replications. Finally, the framework may be applicable to other fields but currently has been only tested and validated for visualization courses and we leave to future work its generalizability to other fields.

## 6. Conclusion

We contribute the Vis Repligogy framework to incorporate replication studies into visualization course curricula. The framework enables instructors to systematically prepare and conduct replication studies in their course and guides instructors on how to successfully run these studies with students. As a proof of concept of the framework, we also present two case studies where students conducted replications in graduate visualization courses. We also reflect on our implementation of the Vis Repligogy framework and provide recommendations for others who are interested in conducting replications in their course. Through this work, we aim to pave a path towards a culture where more efforts are made for students to partake in visualization research, and help build a cohort of informed and responsible visualization researchers for the future. We hope this work will enable a path forward for the visualization community on how to conduct replication studies within the classroom, and to continue validating and developing theories and methods to bolster the foundations of the field.

**Acknowledgments:** The authors wish to thank all the teaching staff and students of both courses and everyone who provided invaluable feedback on this research, including Steve Haroz and John Alexis Guerra-Gómez. This work was supported by the Khoury College of Computer Sciences, Northeastern University.

## References


[ACDF*13] Asendorpf J. B., Conner M., De Fruyt F., De Houwer J., Denissen J. J., Fiedler K., Fiedler S., Funder D. C., Kliegl R., Nosek B. A., et al.: Replication is more than hitting the lottery twice. *European Journal of Personality 27*, 2 (2013), 138–138. 1

[BI15] Begley C. G., Ioannidis J. P.: Reproducibility in science: improving the standard for basic and preclinical research. *Circulation research 116*, 1 (2015), 116–126. 1

[BID*14] Brandt M. J., IJzerman H., Dijksterhuis A., Farach F. J., Geller J., Giner-Sorolla R., Grange J. A., Perugini M., Spies J. R., van't Veer A.: The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology 50* (2014), 217–224. 1, 2, 5, 6

[BMG*10] Bateman S., Mandryk R. L., Gutwin C., Genest A., McDine D., Brooks C.: Useful junk? the effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2010), pp. 2573–2582. 7

[CDBG20] Cockburn A., Dragicevic P., Besançon L., Gutwin C.: Threats of a replication crisis in empirical computer science. *Communications of the ACM 63*, 8 (2020), 70–79. 7

[CGD18] Cockburn A., Gutwin C., Dix A.: Hark no more: On the preregistration of chi experiments. In *Proceedings of the 2018 chi conference on human factors in computing systems* (2018), pp. 1–12. 2

[Col15] Collaboration O. S.: Estimating the reproducibility of psychological science. *Science 349*, 6251 (2015), aac4716. 1

[CP18] Chuang L. L., Pfeil U.: Transparency and openness promotion guidelines for HCI. In *Extended abstracts of the 2018 CHI conference on human factors in computing systems* (2018), pp. 1–4. 2

[DB18] D'Ignazio C., Bhargava R.: Creative data literacy: A constructionist approach to teaching information visualization. doi:https://hdl.handle.net/1721.1/123473. 2

[dLAA*19] de Leeuw J., Andrews J., Altman Z., Andrews R., Appleby R., Bonnano J., DeStefano I., Doyle-Samay E., Faruqui A., Griesmer C., et al.: Similar event-related potentials to structural violations in music and language: A replication of patel, gibson, ratner, besson, & holcomb (1998). *Meta-Psychology 3* (2019). 2

[dLAL*19] de Leeuw J. R., Andrews J., Livingston K., Franke M., Hartshorne J., Hawkins R., Wagge J.: Using replication studies to teach research methods in cognitive science. *Perspectives on Psychological Science 7*, 6 (2019), 600–604. 2

[EH18] Echtler F., Häussler M.: Open source, open science, and the replication crisis in hci. In *Extended abstracts of the 2018 CHI conference on human factors in computing systems* (2018), pp. 1–8. 1, 2

[FS12] Frank M. C., Saxe R.: Teaching replication. *Perspectives on Psychological Science 7*, 6 (2012), 600–604. 1, 2, 4

[GB08] Greenberg S., Buxton B.: Usability evaluation considered harmful (some of the time). In *Proceedings of the SIGCHI conference on Human factors in computing systems* (2008), pp. 111–120. 2

[GL18] Gast D. L., Ledford J. R.: Replication. *Single case research methodology* (2018), 77–95. 1, 2

[H*14] Höffler J. H., et al.: Teaching replication in quantitative empirical economics. In *European Economic Association & Econometric Society 2014 Parallel Meetings (Toulouse), August* (2014), vol. 28. 2

[HMS16] Hüffmeier J., Mazei J., Schultze T.: Reconceptualizing replication as a sequence of different studies: A replication typology. *Journal of Experimental Social Psychology 66* (2016), 81–92. 1

[Hor15] Hornbæk K.: We must be more wrong in HCI research. *Interactions 22*, 6 (2015), 20–21. 2

[HSA*18] Hawkins R. X., Smith E. N., Au C., Arias J. M., Catapano R., Hermann E., Keil M., Lampinen A., Raposo S., Reynolds J., et al.: Improving the replicability of psychological science through pedagogy. *Advances in Methods and Practices in Psychological Science 1*, 1 (2018), 7–18. 2, 4

[HSBAGS14] Hornbæk K., Sander S. S., Bargas-Avila J. A., Grue Simonsen J.: Is once enough? on the extent and content of replications in human-computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2014), pp. 3523–3532. 1, 2, 8

[HSD*19] Hartshorne J. K., Skorb L., Dietz S. L., Garcia C. R., Iozzo G. L., Lamirato K. E., Ledoux J. R., Mu J., Murdock K. N., Ravid J., et al.: The meta-science of adult statistical word segmentation: Part 1. *Collabra: Psychology 5*, 1 (2019). 2

[KH18] Kosara R., Haroz S.: Skipping the replication crisis in visualization: Threats to study validity and how to address them: Position paper. In *2018 IEEE Evaluation and Beyond-Methodological Approaches for Visualization (BELIV)* (2018), IEEE, pp. 102–107. 1, 2

[Kom22] Komninos A.: The value of open data in hci: A case report from mobile text entry research. *Multimodal Technologies and Interaction 6*, 9 (2022), 71. 1, 2

[Kos16] Kosara R.: An empire built on sand: Reexamining what we think we know about visualization. In *Proceedings of the sixth workshop on beyond time and errors on novel evaluation methods for visualization* (2016), pp. 162–168. 1, 2, 6

[Lam90] Lamal P. A.: On the importance of replication. *Journal of Social Behavior and Personality 5*, 4 (1990), 31. 1

[MD20] Meyer M., Dykes J.: Criteria for rigor in visualization design study. *IEEE Transactions on Visualization and Computer Graphics 26*, 1 (2020), 87–97. 1

[MSM99] Matuszewski C., Schönfeld R., Molitor P.: Using sifting for k-layer straightline crossing minimization. In *Graph Drawing: 7th International Symposium, GD'99 Štiřín Castle, Czech Republic September 15–19, 1999 Proceedings 7* (1999), Springer, pp. 217–224. 7

[Pet18] Peterson G. I.: Postpublication peer review: A crucial tool. *Science 359*, 6381 (2018), 1225–1226. 2

[QR19] Quadri G. J., Rosen P.: You can't publish replication studies (and how to anyways). *IEEE VIS Workshop on Vis X Vision, 2019.* (2019). 8

[Sch16] Schmidt S.: Shall we really do it again? the powerful concept of replication is neglected in the social sciences. doi:10.1037/a0015108. 2

[SGL*14] Standing L. G., Grenier M., Lane E. A., Roberts M. S., Sykes S. J.: Using replication projects in teaching research methods. *Psychology Teaching Review 20*, 1 (2014), 96–104. 2, 4

[SM18] Sukumar P. T., Metoyer R.: Towards designing unbiased replication studies in information visualization. In *2018 IEEE Evaluation and Beyond-Methodological Approaches for Visualization (BELIV)* (2018), IEEE, pp. 93–101. 1, 2, 6, 8

[WBB*19] Wagge J. R., Baciu C., Banas K., Nadler J. T., Schwarz S., Weisberg Y., IJzerman H., Legate N., Grahe J.: A demonstration of the collaborative replication and education project: Replication attempts of the red-romance effect. *Collabra: Psychology 5*, 1 (2019). 2, 4

[WBL*19] Wagge J. R., Brandt M. J., Lazarevic L. B., Legate N., Christopherson C., Wiggins B., Grahe J. E.: Publishing research with undergraduate students via replication work: The collaborative replications and education project. *Frontiers in psychology* (2019), 247. 2

[WCRC14] Wilson M. L., Chi E. H., Reeves S., Coyle D.: Replichi: the workshop ii. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. 2014, pp. 33–36. 1, 2

[WEHE20] Wacharamanotham C., Eisenring L., Haroz S., Echtler F.: Transparency of chi research artifacts: Results of a self-reported survey. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), pp. 1–14. 1, 2

[Wil13] Wilson M. L.: Teaching hci methods: Replicating a study of collaborative search. In *RepliCHI* (2013), pp. 39–43. 2

[Wol15] Wolfe J.: Teaching students to focus on the data in data visualization. *Journal of Business and Technical Communication 29*, 3 (2015), 344–359. doi:10.1177/1050651915573944. 2

[WRCC13] Wilson M. L., Resnick P., Coyle D., Chi E. H.: Replichi: the workshop. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 2013, pp. 3159–3162. 2