

Bringing the Gold Standard into the Classroom: Replication in University Teaching¹

NICOLE JANZ

University of Cambridge

Reproducibility is held to be the gold standard for scientific research. The credibility of published work depends on being able to replicate the results. However, there are few incentives to conduct replication studies in political science. Replications are difficult to conduct, time-consuming, and hard to publish because of a presumed lack of originality. This article sees a solution in a profound change in graduate teaching. Universities should introduce replications as class assignments in methods training or invest in new stand-alone replication workshops to establish a culture of replication and reproducibility. This article will first discuss the benefits of conducting replications. The main part will focus on concrete steps in integrating replication in the classroom, from selecting a paper to final manuscripts. Drawing on the author's own teaching experience as well as that of others, particular emphasis will be on the pitfalls and challenges of letting students replicate work, as well as potential criticism. Only if universities nurture a reproducibility and replication culture can we ensure that the gold standard of reliable, credible, and valid research is not just an empty phrase.

Keywords: reproducibility, replication, research transparency, open data, methods training

Reproducibility is the gold standard for scientific research. The legitimacy of published work depends on whether we can replicate the analysis and reach the same results. Therefore, authors must provide information on how exactly they collect and analyze data. Without such transparency, scholars cannot fully understand the value of results and create new knowledge (King 1995). Replication studies can serve as a vehicle to hold the original author “accountable” for their work, thereby acting as a “deterrent” for “irresponsible behavior” (Ishiyama 2014:79). While most political scientists agree on the benefits of reproducibility and replication, there is still no consensus about how to implement these principles in practice. Two main problems are that (i) not all researchers work transparently and (ii) there are few professional rewards for those who cross-check previous work through replication.

¹An earlier version of this article was presented at the ISA Annual Convention, Toronto, 2014. I would like to thank the students of the Cambridge Replication Workshop and the numerous scholars who answered the students' replication queries. The workshop was supported by the Economic and Social Research Council (ES/L003120/1) and the Social Sciences Research Methods Centre at the University of Cambridge. Thanks are also due to the ISP editors and the anonymous reviewers.

Original authors do not always provide sufficient data and analysis details or may not archive them so that others can understand each step in the original analysis (Lupia and Elman 2014). One reason is that the field still lacks clear guidelines on how research information should be shared. A recent study found that only 18 of 120 political science journals have a replication policy (Gherghina and Katsanidou 2013). In addition, working transparently involves maintaining detailed logs of data collection and variable transformations as well as of the analysis itself. Teaching commitments and the pressure to publish often leave little incentive to invest the necessary time. Fear of losing reputation should a replication attempt fail might also explain a reluctance to allow access to replication data (Lupia and Elman 2014).

If data are not made available, scholars cannot evaluate and cross-check published work. But even if all research were transparent, there would be little incentive to conduct replications. A common criterion in the peer review process is the presentation of new, original research, which marginalizes the re-analysis of published work (Carsey 2014). There is little motivation for scholars to conduct a replication study when the prospect of publication is low. Therefore, much of the knowledge we trust today remains unchecked.

This article argues that the twin challenges of irreproducibility and the scarcity of researchers willing to be replicators can be alleviated through a change in graduate training. By teaching transparency tools and encouraging students to replicate existing work, the gold standard for scientific research can be implemented more efficiently than before. Keeping logs and depositing data publicly will develop a transparency routine for students' future careers. If they conduct replication studies as part of their methods training, they will not only understand methods better but also learn firsthand, by trying to re-analyze published work, when an analysis is really reproducible and when it is not.

There have been several calls from inside the discipline and other fields² to implement replication in teaching. King stated that it is "an extremely useful pedagogical tool," even if the lack of data availability can make the task difficult (King 1995:445). He proposed new policies for graduate studies, such as requiring students to submit replication data for their PhD dissertations, and letting students replicate published articles if data sets in political science become more widely available (King 2006). While some scholars cautioned that generations of students would become "data vultures" (Gibson 1995:475), focusing on finding errors, the general consensus among scholars and research funders³ is that replication is necessary. Its implementation partly depends on integrating principles of transparency as "central elements" into graduate teaching (Carsey 2014:74), and there is a "growing trend" of assigning students the task of conducting replication studies in methods training (Carsey 2014:74).

Still, there is no common understanding of how universities should include replication in different types of courses and workshops. Therefore, this article contributes to both literature and practice by examining the issue systematically and suggesting how universities and teachers can integrate principles of data access and transparency as central elements into graduate teaching. After a definition of the key terms, the article will discuss the benefits of replication studies for graduate students. The main section will describe concrete steps in how to integrate reproducibility and replication in the classroom in different types of courses, including details of a replication process: selecting a paper, obtaining data, re-

²Calls for integrating replications in the class room have also been made in psychology (Frank and Saxe 2012; Koole and Lakens 2012) and economics (Hoeffler 2013).

³The National Science Foundation now requires a data management plan (National Science Foundation 2011). The American Political Science Association published revised ethics guidelines emphasizing the need for full data access and research transparency (Lupia and Elman 2014).

analyzing, adding value to the re-analysis, cross-checking among students, and publishing the results. Drawing from the author's own experience of teaching a replication workshop⁴ at the University of Cambridge as well as that of others, particular emphasis will be on the pitfalls and challenges in such teaching as well as potential criticism of letting students replicate work. The conclusion will describe further steps necessary in the field.

Reproducibility vs. Replication

There may be ambiguity in the meaning of terms such as replication and reproducibility.⁵ What kinds of information must authors provide? Is replication simply re-analysis of an existing article? And what counts as a "failed" replication?

Reproducibility

An author can improve reproducibility by providing information to "understand, evaluate, and build upon a prior work" (King 1995:444). Newly drafted ethics guidelines by the American Political Science Association (APSA) emphasize that researchers must provide (i) data access, (ii) details of how they collected the data, and (iii) details of the analysis that led to their conclusions (Lupia and Elman 2014). In practice, this means that the author should provide supplemental documents such as data files and software codes (for example, STATA do-files or Rscripts). It should be clear where the original sources of data can be found and how variables were transformed (Dafoe 2014).⁶ These data can be made available in repositories such as the Institution for Social and Policy Studies (ISPS) at Yale University, the Dataverse Network at Harvard University, and the Inter-university Consortium for Political and Social Research (ICPSR); on journal-specific Web sites and archives; or on the original author's webpage (Dafoe 2014).⁷ If privacy, proprietary issues, and other nondisclosure agreements prevent full data access, this should be noted in the paper (Carsey 2014; Lupia and Alter 2014).

Replication

A replication is the process by which a published article's hypotheses and findings are re-analyzed to confirm or challenge the results. How exactly a replication study should be conducted, however, is still an "open question" (Carsey 2014:73), and it is important for the integration of replication into teaching to provide clarity. There are three main questions: (i) Should the same, similar, or newly collected data be used? (ii) How closely should one follow the original models? and (iii) How far should the new results deviate from the original work before claiming that the replication "failed"?

For many scholars, a first and simple step in re-analyzing published work is to use the data set provided by the original author. This can be a first check to see whether the results can be "duplicated" or "reproduced" (King 2003:98). Errors in the data set, faulty coding procedures, or other issues with the variable construction can be detected to test "reliability in research results" (King 2003:99).

⁴The class is the "Cambridge Replication Workshop," see <http://nicolejanz.de/teaching/replication.html>.

⁵While there is a very recent and important discussion on reproducibility in qualitative research (Elman and Kapiszewski 2014; Moravcsik 2014), this article concentrates on these terms as applied to statistical analyses.

⁶The information on data collection, sources, and transformations is sometimes called "metadata" (Carsey 2014:75).

⁷Ideally, the data would be stored permanently in a public repository where it can be easily found and identified by a DOI, such as at the Harvard Dataverse Network. If the author changes institutions, a complete transfer of all data files might not always be secured, and some authors do not have personal webpages or update their pages regularly.

Re-analyzing work based on the same data (assuming full data access is provided by the original author) is therefore important as an initial step but, as King (2003) states, to advance knowledge, the results must be *replicated* using newly collected data. Other scholars agree on this additional requirement for a good replication study (Herrnson 1995; Carsey 2014).

Similar criteria apply when considering how closely one should follow the original statistical models. To assess the robustness of an analysis, the researcher must add different statistical techniques, variables, or specifications (Herrnson 1995). King also states that in order to advance knowledge from existing research, one must “follow the precise path taken by a previous researcher, and then improve on the data or methodology” (King 1995:445). Carsey (2014) points out that leading journals should not publish replication studies based solely on the same data and methods as the original paper, which points to the fact that “more” is expected.

To sum up, a duplication study verifies previous research results by attempting to produce the exact same results based on the exact same data set with exactly the same methods. A replication study further tests the robustness of previous research results by employing newly collected data, and/or new variables, and/or new model specifications. An ideal “gold standard” replication study would perform most of these three extensions while ensuring that it is transparent and reproducible itself. Table 1 in Appendix S1 describes the difference between duplication and replication in more detail and provides a checklist of items that should be achieved by replicators.

A “Failed” Replication

A replication attempt can fail at different stages. If the results cannot be duplicated at the first stage, there is clearly little reason to trust the work, and if at the second stage, after using new data and improved methods, the results cannot be reproduced, one would have to describe exactly at which point the replication has failed. Different measurements of concepts that are hard to quantify, for example, human rights, can naturally yield different results (Meyer 1999). Therefore, different results do not necessarily mean that the original article was faulty, and so it is all the more important to make sure that the replicator fully understands the methods and variables of the original study. In fact, when the replication of an article is reported as “failed,” original authors often claim that the replication itself was flawed. For example, one original author criticized a replication of his published article as “less realistic,” “inconsistent with the substantive literature,” and “of limited utility” (Mansfield, Milner, and Rosendorff 2002:167),⁸ and others complained of a “fundamentally flawed” replication of their work (Peffley, Knigge, and Hurwitz 2001b:421),⁹ while a further author stated that a replication of his work contained “statistical, computational, and reporting errors that invalidate its conclusions” (Gerber and Green 2005:301).¹⁰ This means that students need to be very careful to provide clearly documented evidence before calling a replication “failed.” By being even more diligent and transparent, students can prevent the original author claiming—justifiably or not—that they simply lack the necessary skills (Ishiyama 2014).

⁸Mansfield, Milner, and Rosendorff (2000) was replicated by Dai (2002), and in turn, the original author commented on the replication study in Mansfield, Milner, and Rosendorff (2002).

⁹The original article was Peffley, Knigge, and Hurwitz (2001a), the replication was published by Miller et al. (2001), and the original author in turn commented on the replication in the same issue (Peffley, Knigge, and Hurwitz 2001b).

¹⁰The original study was Gerber and Green (2000), the replication study was Imai (2005), and the original author in turn reacted in Gerber and Green (2005).

Defining Replication Studies in Teaching

I suggest that the requirements for a replication study conducted by students should depend on the purpose of the course. A “minimum” duplication version might be more suitable for pure methods training, and an “extended” version for advanced courses and potential publication. In basic statistics courses, or courses with limited time, a replication study would be a one-stage process of analyzing the same data as the original author, following the same statistical procedures. If the tables, figures, and results can be reproduced, the re-analysis succeeds; if not, it fails. This has been called duplication or re-analysis by scholars in the field. For the sake of simplicity and practicality, I would find it permissible to call this a replication study if the teacher makes it clear when assigning the task that a second, more advanced stage can follow. For additional learning purposes, for example, in advanced and more time-intensive methods courses, one could extend the initial assignment by letting students re-collect the data using different measurements, criticizing the models or theory, doing robustness checks, etc.¹¹

Benefits of Replication Studies for Students

Students might ask why they should replicate published papers. Isn't it a waste of time, given that journals and universities expect original (doctoral) research? Here are my answers:

1. A better way to learn statistics: Replication is essential to a deeper understanding of statistical tests and modeling. The advantage over textbook exercises is that students use real-life data with all bugs and complications included. In addition, by going through the data and codes of the original study, students realize what kinds of decisions the author made, for example, about variable transformations, missing observations, or model specifications. As King stated, one can see “replication not as an end in itself but as a means for acquainting yourself with the methods used in a study, the original author's line of thinking, the complications he or she must have faced, and the solutions” (Price 2011).
2. Jumping to the research frontier: The replication of recently published results allows students to find out how to add knowledge to their field in the best way possible. Compiling a literature review is not always sufficient to appreciate the details of the data challenges and state-of-the-art methods that drive cutting-edge research. Victoria Stodden, who assigns replications in her courses, emphasizes: “The remarkable difficulties students have in replicating published articles teaches more about the state of the literature (...) than reading all the published literature.”¹²
3. Getting published early: Working at the research frontier based on replications also facilitates early publication. King highlights that “If (...) you begin a project from scratch without replication, you need to defend every coding decision, every hypothesis, every data source, every method - everything. In contrast, if you start with replication, you only need to defend the one area you are improving” (King 2006:119). A range of replication articles which began as class projects, such as Bell and Miller (2013), can now be found in political science journals.
4. Creating a reproducibility routine: Replication studies as part of methods training not only improves understanding of statistics. Replication almost

¹¹See a list of how to add value in King (2006), and in a handout provided in Appendix S1.

¹²Victoria Stodden, Columbia University, on her syllabus for “STAT 8325: Topics in Advanced Statistics: Fall 2012”. Syllabus kindly provided by V. Stodden.

always involves frustration because data are not accessible, software codings are unclear (or not available), and methods and variable transformations are not described in detail.¹³ This frustration is an effective, if painful, way to learn firsthand when published results are really reproducible and when they are not, and will ideally help students to improve transparency in their own work: while “the experience is in part disheartening, (...) it also seems to empower students who (correctly) conclude that they can do better” (King 2006:120). If students understand the value of keeping logs and providing their own data, they should develop a reproducibility routine which will hopefully feel automatic and natural to them (Carsey 2014).

5. Introducing fun into statistics teaching: Replication studies are not always frustrating: The kind of “detective work” involved in replicating cutting-edge work can be “exciting and fun” (Frank and Saxe 2012:600). For example, a human rights scholar remembers a replication class project: “There was a typo in one of the tables and the challenge for the students was to find the typo. That was a great exercise.”¹⁴ Student feedback has also shown that replication studies can be motivating.¹⁵
6. Developing professionalism: Finally, by engaging with a published study in depth, including its methods, coding decisions, and presentation of results, students learn firsthand about scientific norms and will better understand what kinds of decisions in all steps of an analysis are acceptable. Therefore, teaching based on replication helps to “professionalize students into the discipline” (King 2006:119).

Types of Courses

Introductory lectures in Political Science departments should not only emphasize research transparency, reproducibility standards, and data access, but also discuss practical steps such as keeping full logs of files from day one of the doctoral research. When ideas about reproducibility “are blended with discussions of developing research questions, formulating initial research plans, and developing research designs” (Carsey 2014:74), students can incorporate these principles in their own research—at least in theory.

To ensure that ideas about reproducibility are put into practice, the most common implementation in teaching is to assign replication studies in standard methods courses. The goal is to teach statistical techniques, but instead of being given problem sets, students must replicate (parts of) a published study employing the methods they learn in class. While not yet standard practice, this seems to be a “growing trend” (Carsey 2014).¹⁶ The most widely known course of this kind is “Government 2001” at Harvard University, taught by Gary King.¹⁷ According to the syllabus,¹⁸ the students team up in small groups and conduct a replication

¹³See the articles “Replication Frustration in Political Science,” “Nightmare after Nightmare: Students Trying to Replicate Work,” and “Replication Workshop: What Frustrated Students the Most, and Why They Still Liked the Course,” on the Political Science Replication blog, <http://politicalsciencereplication.wordpress.com> (accessed March 15, 2014).

¹⁴Todd Landman, University of Essex, on the Political Science Replication blog, <http://politicalsciencereplication.wordpress.com/2013/01/30/sharing-of-qualitative-data-is-possible-but-the-volume-of-information-is-gigantic-says-todd-landman/> (accessed March 12, 2014).

¹⁵See <http://politicalsciencereplication.wordpress.com/2014/02/25/replication-workshop-what-frustrated-students-and-why-they-still-liked-the-course/> (accessed March 15, 2014).

¹⁶Several syllabi of such courses can be found in Appendix S1.

¹⁷“Advanced Quantitative Political Methodology: Government 2001,” see class materials at <http://projects.iq.harvard.edu/gov2001/home> (accessed March 15, 2014).

¹⁸Download of the most recent syllabus as pdf: <http://projects.iq.harvard.edu/gov2001/book/syllabus> (accessed March 15, 2014).

study, aiming “to produce a publishable article, and, in fact, most students do publish their final paper in a scholarly journal.” In order to encourage students to follow a reproducible workflow, they must hand over all data to another student team, which will then replicate and assess their manuscript.¹⁹

Thomas M. Carsey, University of North Carolina at Chapel Hill, has been assigning replication studies to his students for the last decade (Carsey 2014). In his intermediate statistics course,²⁰ students write a replication paper modeled after articles in high-ranking journals. First, students must reproduce the findings by re-collecting the data from the original sources. Then, they extend the study by building on the analysis, which should be “derived from a clear theoretical proposition,” as the syllabus states.

Another example is Carlisle Rainey’s statistics course at the University at Buffalo.²¹ Students have to submit a high-quality replication paper which “should make a contribution to a political science literature,” including a replication data set and a conference-style presentation. Jeff Gill, Washington University in St Louis, requires students to “find a published work in your field of interest, obtain the data, and exactly replicate the author’s model results.”²² Similarly, Christopher Fariss at Penn State University asks his students to replicate a research paper published in the last five years.²³ Students must describe the initial article and “the ease with which the results replicate,” in addition to improving the research design.

These examples show that courses vary in their requirements of statistical knowledge, depth of analysis, and extension after re-analysis. Some teachers ask students to submit an individual paper; others require them to work in teams. It is not always clear from the syllabi (except for King’s course)²⁴ whether the assignments will remain unpublished or will be submitted to a public data repository. The advantage of such courses over a stand-alone interdisciplinary replication workshop is that they are often a mandatory part of methods training, so that a complete cohort of students is exposed to replication and reproducibility standards. Many of the courses are graded, which is an incentive to put up with the frustration involved. A disadvantage of this type of course might be that students have to spend additional time preparing readings for lectures and solving problem sets, which takes time away from conducting the replication study. In addition, if departments cannot fund such courses, or if lecturers hesitate to take up the extra workload, other formats might be more appropriate (see the next section).

An alternative to assigning replications in methods courses is a stand-alone replication workshop, which could be integrated into a summer school or run during term time. Here, students with advanced statistics skills learn about reproducibility and are guided through the process of conducting a replication study. I do not know how many courses like this, if any, are currently conducted in political science or the wider social sciences. This section of the article is based on my experience of running a stand-alone, interdisciplinary replication course for several

¹⁹See <http://projects.iq.harvard.edu/gov2001/book/replication-paper> (accessed March 15, 2014). See also King (1995, 2006).

²⁰“POLI 784: Intermediate Statistics, Spring 2014,” see syllabus at <http://carsey.web.unc.edu/teaching/> (accessed March 15, 2014).

²¹“PSC 531: Intermediate Statistics for the Social Sciences,” see syllabus at <http://www.carlislerainey.com/teaching/linear-models/> (accessed March 15, 2014).

²²“Political Science 582: Quantitative Analysis in Political Science II,” see syllabus at <http://artsci.wustl.edu/~jgill/PS582.2013.html> (accessed March 15, 2014).

²³“PLS 501: Methods of Political Analysis (Research Design),” see syllabus at <http://cfariss.com/> (accessed March 15, 2014).

²⁴Students have to upload their papers and data to the Harvard Dataverse, see <http://projects.iq.harvard.edu/gov2001/data> (accessed March 15, 2014).

years. In the Cambridge Replication Workshop, graduate students replicate a paper in their field over the course of eight weekly sessions. There is, in fact, more time to conduct the replication because of a two- to three-week break that allows for self-directed work before the results are presented. The course is offered by an integrated methods center²⁵ at the author's university, which provides methods training for graduate students at Masters and PhD level in all social science fields and is therefore interdisciplinary in nature. The students' statistical and software skills vary considerably. The main prerequisites for the course are (i) a good knowledge of basic statistics including multiple linear regression and data handling in R, (ii) a commitment to at least six hours' self-directed work per week, and (iii) the thesis committee/supervisor agreeing to the participation of their students.

In the last academic years, the course admitted about 15 students. The first four sessions focused on picking a suitable paper, downloading the data, and reproducing the results. During the second half of the course, students added value to the replication and drafted a paper or report, which was uploaded to the class data repository.²⁶ Each session consisted of a lecture introducing reproducibility standards and tools followed by a practical element to establish a reproducible filing and logging system and to help students with R coding, model specifications, and other problems during the replication of "their" paper. At some point during the course, students exchanged their code and data to provide and gain feedback. In order to ensure that they were all kept informed about the others' projects, they shared a drop box and gave weekly updates in class.

Students were confronted with the following challenges: (i) The data were nowhere to be found and the original sources were not clear, (ii) the original author did not respond to queries for data, (iii) the authors did not remember where they had stored their files, (iv) the steps in the analyses were not well described, (v) it was not clear how the variables were transformed before entering the analysis, and (vi) statistical models remained opaque. This irreproducibility across all social science fields led to frustration among students and demonstrated the consequences of lack of transparency. Even the experienced teaching assistants were surprised at the challenges students had to face.

The advantage of a stand-alone replication course is that it can be offered on a voluntary basis in addition to mandatory statistics classes. It is therefore possible to build up a strong reproducibility routine and to further stress the value of replication studies in new courses without changing standard modules. The voluntary nature of such a course also implies prior motivation and interest among students who sign up. Considerable time and effort was concentrated on the replication instead of teaching statistics or software, so that much hands-on help could be provided. The interdisciplinary approach allowed students from different fields to interact and exchange ideas, which fostered an understanding of different approaches to social science puzzles. In student feedback, many reported that they learned more about statistical methods than on standard statistics courses.

There are some disadvantages of a stand-alone research workshop. If the course is offered to all social science students, the interdisciplinary setup helps to permeate the sometimes artificial boundaries between disciplines; but it can also hinder in-depth, discipline-oriented discussions. While the methods were similar, students sometimes had difficulty understanding the details of their peers' projects. In addition, students underestimated the workload and found it difficult to submit the weekly assignments as steps of the replication process. Each year, some students dropped out of the class due to time issues. Finally, skills levels varied considerably. While the prerequisites were intended to filter out students who

²⁵The Social Sciences Research Methods Centre (SSRMC), see <http://www.ssrmc.group.cam.ac.uk/>.

²⁶See <http://thedata.harvard.edu/dvn/dv/CambridgeReplication> (accessed March 15, 2014).

needed more (basic) statistics training, some were still overwhelmed by methods in the papers they chose. The teaching assistants provided *ad hoc* tutorials, but the schedule did not plan for time to teach new methods in depth. The same difficulties arose regarding software skills. While some students found the necessary R packages and functions easily online, others struggled with simple data management. Teachers of a stand-alone course might have to develop pre-assessment mechanisms to identify those students who genuinely meet the requirements. Finally, a stand-alone replication course might involve very intense tutoring by teaching assistants and the instructor. We found that a ratio of one assistant to no more than four students was effective.

Navigating the Replication Process

When trying to integrate replication into teaching in different course setups, some steps will probably be the same. These include (i) selecting a paper to replicate, including data access, (ii) reproducing the models, (iii) adding value, and potentially also (iv) a cross-check between students, (v) uploading the class assignment to a repository, and (vi) a conference or journal submission. Many of these steps have been described by King (2006). Additional thoughts from the experiences of my replication course and other courses, and some possible solutions to challenges and pitfalls, are included here to encourage teachers to consider the adoption of replication in different kinds of teaching formats.

Selecting a Paper to Replicate, Including Data Access

How to select papers is an important learning process for students and should be part of a replication study assignment. The best tips on how to do this are provided by King (2006), and they can be easily adapted. In my stand-alone replication course, students are asked to pick a paper published in recent years (in a top journal), where the data set is available from the original author. They also should find a paper using methods they know already or have learned during the course. In my course, I insist that students locate the data set for the paper they want to replicate. While courses such as Carsey's "POLI 784"²⁷ require students to re-collect all data from the original sources, I found this too challenging for my students, at least to start with. In the first run of the replication workshop, one of the students tried to download all data from public sources because they were not available as a replication data set. It took the student several weeks to download, rearrange, clean, and subset the data, and he subsequently dropped out of the course because by then other students had already finished re-analyzing the models. In the second year, a student tried to obtain replication data from five different original authors who had not uploaded the files. None of them obliged, and by week three of the course, the student had dropped out. In a longer course, re-collecting the data instead of using the original files might still work, but I find it preferable to have access to the data used by the original author.

Students must also assess whether the methods employed are manageable. Students often underestimate this. A method might sound "easy," but when it comes to coding the specifications (especially if the software code in STATA or R is not available), suddenly more questions arise than are answered. In a methods course that teaches advanced statistics over a longer period of time (as most do), this may not be a major problem, but in a stand-alone course with varying student statistics levels, it may become so. In the second run of my replication course, a student had difficulties with ordered logit specifications that deviated from the

²⁷Syllabus at <http://carsey.web.unc.edu/teaching/> (accessed March 15, 2014).

standard versions. Since the original author only provided a generic STATA command in a footnote, the student had to invest a considerable amount of time in studying STATA manuals for the model specifications and then translating that into R code. In order to combat the problem of “too advanced models,” I have tried to recommend to my students recent papers using simple OLS. This solution has not been suitable so far, since it is difficult to find papers using OLS in good journals and students prefer to find their own papers which match their interests better (rather than their methods skills).

Another challenge in the stand-alone replication course is that students have to pick a paper by the end of the second week at the latest, while a methods course such as those described earlier might give them longer. Students in my course had little time to do this, so in the second year of running the course, I asked them to bring one to three papers to the very first session, after I had provided guidance by email beforehand. This helped to speed up the process, but some students found it overwhelming to pick a paper themselves without discussing the criteria and practical implications in class. A solution might be to arrange meetings with teaching assistants before the course starts or to develop stand-alone replication courses of more than eight weeks.

These experiences show that even the first step of picking a suitable paper can be difficult and might deter students from conducting replication studies. It is therefore all the more important to provide guidelines and tailored advice on this. Once a course has been running for several years, the teacher might want to set up a database of which kinds of papers “worked” for students.²⁸

Reproducing Models and Results Tables

The second step after picking a paper and becoming familiar with the data²⁹ is to re-analyze the models used in the original piece. Papers report results in various ways, for example, as tables, figures, or text. Not all papers describe the model specifications clearly or provide STATA commands or R functions used. In order to create a full list of models to re-analyze, I use an assignment which requires students to copy and paste screenshots of all tables and figures from the original paper into a document, and to quote word-for-word the phrases describing models in the text. This gives students a step-by-step guideline on what they should reproduce. Without this step, students were at times confused and overwhelmed as to which of all the reported results they should concentrate on.

After the analysis, students must report back in class and clarify “the extent to which you were able to replicate the author’s results” (King 2006:120). During this stage, I encouraged my students to discuss in class to what extent the original authors were really “wrong,” or if the students themselves might have misunderstood the analysis. Students also discussed how the original author should have presented the results (more clearly) and how the author should have given access to data, code, variable codebooks, etc. This demonstrated to students what a good transparent workflow looks like. During the stage of duplicating the results, I also asked students to work transparently themselves, keeping their data and analysis files in a shared drop box that is separated into separate folders for data, analysis, figures, etc. (Gandrud 2013:62).

²⁸I recommended my students to look at the previous year’s uploaded replication studies in our class repository (<http://thedata.harvard.edu/dvn/dv/CambridgeReplication>) or to read those by Gary King’s students at <http://projects.iq.harvard.edu/gov2001/data> (accessed March 15, 2014).

²⁹A first assignment in my replication workshop is to create a file called “descriptive.R” in which the students provide summary statistics on all data. This often reveals when it is less than clear which variables were used by the original author or whether they were transformed, for example, by logs.

Adding Value to the Re-analysis

A pure re-analysis, as mentioned above, is a good learning exercise. However, for more advanced statistics students, and to increase the prospect of publishing the paper, value must be added. Carsey's syllabus for "POLI 784" recommends using a different coding of a variable, adding new variables, considering different model specifications, or adding new data. All these extensions must be "derived from a clear theoretical proposition and/or a clear methodological critique."³⁰ King advises starting with "the smallest number of improvements possible to produce new results," including the handling of "missing data, selection bias, omitted variable bias, the model specification, differential item functioning, the functional form, etc., adding control variables or better measures, extending the time series and conducting out-of-sample tests, applying a better statistical model" (King 2006:120). In my replication course, I ask students to explore how replication studies published in journals in their field are structured and how these authors extended the initial re-analysis.³¹ This way, students learn which kinds of improvements are necessary in their field to turn a re-analysis into a publishable paper.

Cross-checking Between Students

In some of the existing courses using replication in class, students are required to cross-check each others' work. They exchange their draft papers, software codes, full variable codebooks, and data. Ideally, they note specifically which results they could replicate or not, why they think that is, and how they added value. There are several benefits to this exercise. First, exchanging drafts for feedback is a form of professionalism in scientific work (King 2006). Second, other students in class may be able to help solve problems with models, coding, or writing. This not only improves the paper, but also potentially explains why a replication did not succeed. Third, exchanging papers and codes can demonstrate whether the students work reproducibly themselves.

Publishing the Replication Study in a Repository

In King's class "Government 2001,"³² students are required to upload their final paper and data to the Harvard Dataverse after being checked by the instructor. The upload will get a DOI, permanent URL, and suggested form of citation of the study. This is another step toward establishing transparency and data sharing among graduate students and also makes the results of the replication study available to the community. King advises that a copy of the paper should first be sent to the original author, who can respond to the critique and comment on possible failed replications.

So far, a look at course syllabi shows that few require students to upload the results of the replication study. A recent survey³³ among teachers assigning replications, and students doing replications, shows similar results. More than 70% of the respondents said that the results were never (or rarely) shared outside of the

³⁰Syllabus at <http://carsey.web.unc.edu/teaching/> (accessed March 15, 2014).

³¹These papers include Bell and Miller (2013). Each student is assigned a published replication paper and presents the outline to the class, saying how they added value. We then discuss which of the improvements are theoretically grounded or just "playing around," which are easy and quick, and which could be suitable for their own paper. A handout of a list of possible improvements compiled from these discussions in class can be found in Appendix S1 under "How to add value."

³²See class materials at: <http://gking.harvard.edu/classes/advanced-quantitative-political-methodology-government-2001-government-1002-and-e-2001> (accessed March 15, 2014).

³³The survey is collecting responses on an ongoing basis now: <http://tinyurl.com/onqy34b> (accessed March 15, 2014).

classroom; only 13% noted that the replication studies were afterward published on the course Web site or other data archives (Janz, Werfel, and Wykstra 2014). Reluctance to publish student replication studies in repositories is not surprising, as results would have to be polished and quality-checked by the instructor. Therefore, many replication studies by students remain an unused resource and are not discussed in the community, although they might contain important corrections to published work.

Conference or Journal Submission

Submission to a journal is a final step and ultimately the most rewarding for students. The initial class project would have to be rewritten following the standard guidelines of journals. The results, and the criticism of the original paper, must be presented in a nuanced, neutral, and professional way (King 2006). In my course, students examine published replication studies to learn how to write one themselves. Many of the published replication articles are presented as original research while mentioning that they build on the work and data of a previous article.³⁴ If no journal submission is (yet) intended, some of my students turn their replication into a PhD chapter, or they present the replication paper at a conference or aim to publish it on their laboratory Web site. For any course assigning replication studies, it is important to find similarly rewarding ways to utilize the output.

Criticism of Replication in the Classroom

Not everyone agrees that students should replicate published work during their graduate studies. Some criticism of the practice aims to protect students, and some questions the motives and professionalism of young researchers who replicate existing work.³⁵

Criticism 1: Letting students believe they can later publish their replication study could encourage destructive “error hunting.” There might indeed be publication bias toward replications that failed.³⁶ However, students do not have the time to work on several projects until they find one that does not replicate. I have experienced that students felt successful when they could replicate tables and figures, and frustrated when they could not. No student was eager to find an error; on the contrary, when students could not replicate a table, they spent weeks re-doing their own coding, assuming they (not the original author) had made mistakes. In addition, the problem that failed replications might be more publishable is a serious issue which needs to be addressed by journals and in the peer review process. It should not deter teachers from assigning replications.

Criticism 2: If young scholars start their career by correcting “rogue scientists,” it provides an unhealthy socialization in the discipline because it creates a distorted picture of what science is about. I would argue that replicating existing work is actually an excellent way to introduce them to the discipline. The painful process of re-analyzing data and adding to an existing study helps to understand that science is about reproducibility. Learning firsthand what it means to work transparently is the best socialization graduate students can have, and they even make their own contribution when they add knowledge to the re-analysis.

³⁴A handout with a list of what to do with the manuscript after class, including publication, is presented in Appendix S1.

³⁵Much of the criticism described here was discussed at the ISA Annual Convention, Toronto, 2014. I thank the panel members, and particularly the discussant Nils B. Weidmann, for raising these points.

³⁶There is evidence from experimental psychology that a publication bias for replication studies exists (Francis 2012).

Criticism 3: There could be reputational repercussions for young scholars if their first appearance in the “journal arena” is a paper that aims to denigrate “big names.” Such criticism seems patronizing. Introducing replication in the classroom ensures that students learn to conduct replication studies professionally, using adequate methods and language. I am not sure that the community really punishes replicators in the job market; if it does, then it must change. The answer cannot be to stop students from checking existing work.

Criticism 4: Students might not have the resources and expertise of the original authors; they might wrongfully label a study as “failed” and damage the original author’s reputation. A biologist recently wrote in *NATURE* that a failed replication could “jeopardize the original scientists’ chances of obtaining funding” (Bissell 2013).³⁷ An author in the field of social psychology, whose paper failed to replicate, wrote of the “defamation” of her work. She was asked about the failed replication of her research in a grant interview, and a peer reviewer of another of her articles questioned the validity of her overall work (Schnall 2014).³⁸ The potential reputational damage when published articles are not reproducible should not be ignored. Therefore, it is all the more important that replicators work in a professional way. Students need to learn how to draft their replication papers with care and make sure that they call a replication “failed” only after extensive analysis. Some responsibility also rests with journals, which could invite comments from original authors when they publish a replication of their work.

Criticism 5: If students only replicate those studies that provide their data and code openly, this could create a bias toward checking work of “good” researchers who work reproducibly. This criticism can only be dealt with if students do not stop at duplication based on provided data but turn to replication that involves collecting new data. Even if students only conduct duplications or re-analyze based on provided materials, I would assume that those researchers who do work transparently have little to fear (and nothing to hide). Embedding replication in teaching will encourage new waves of “good” researchers who work transparently so that the “bad” ones stand out, not vice versa.

Criticism 6: The discipline should not relegate the important task of cross-checking published articles to unpaid graduate students. This is a crucial and valid point. Unfortunately, some senior researchers might not wish to do replication studies because they have usually completed their methods training, and journals prefer “original” work. However, we should not forget that the students to whom the community might give the task of cross-checking are future researchers who will hopefully go on to perform valuable replication studies when they are more senior. By letting students replicate, we do not outsource replication, but we integrate it into the field for the future. The fact that the publishing process does not always reward replications should lead us to question journal practices and not prevent students from replicating work in their methods training.

³⁷Bissell claimed that in her field of biology: “People trying to repeat others’ research often do not have the time, funding, or resources to gain the same expertise with the experimental protocol as the original authors, who were perhaps operating under a multiyear federal grant and aiming for a high-profile publication. If a researcher spends six months, say, trying to replicate such work and reports that it is irreproducible, that can deter other scientists from pursuing a promising line of research, jeopardize the original scientists’ chances of obtaining funding to continue it themselves, and potentially damage their reputations.” See <http://www.nature.com/news/reproducibility-the-risks-of-the-replication-drive-1.14184> (accessed July 18, 2014).

³⁸Simone Schnall wrote in a blog: “... careers and funding decisions are based on reputations. The implicit accusations that currently come with failure to replicate an existing finding can do tremendous damage to somebody’s reputation, especially if accompanied by mocking and bullying on social media. So the burden of proof needs to be high before claims about replication evidence can be made.” See <http://www.spsblog.org/simone-schnall-on-her-experience-with-a-registered-replication-project/> (accessed July 18, 2014).

Conclusion: Where to Take It from Here?

This article has argued that reproducibility and replication as the gold standard for scientific research is inadequately implemented in the field of political science. One way to improve adherence to such standards is to embed them in teaching practice for graduate students. Universities should encourage instructors of different types of courses to assign replications to establish a culture of replication and reproducibility during early career stages.

In order to show how the gap between ideal and implementation can be reduced, this article first clarified the difference between reproducibility, duplication, and replication studies. Secondly, it presented a range of benefits of doing replication studies. The main contribution of the article was a thorough discussion of practical implications. It documented which scholars currently assign replications in the classroom and how they did so. The article also aimed to encourage other teachers to start assigning replications by laying out the replication process in detail, from selecting a paper to final manuscripts. Drawing on the author's own teaching experience as well as that of others, particular emphasis was placed on the pitfalls and challenges of letting students replicate work. Finally, the paper listed criticisms of replication work and offered responses to replication skeptics.

For the future, it is important to establish networks among teachers who assign replications. At present, course instructors do not always know who else is doing so. A platform or Web site that collects syllabi and encourages exchange of experiences and solutions to problems in class could be useful, and this article is intended to stimulate discussion in the community on how to connect teachers and store class materials in a more systematic way. One way of bringing together instructors who assign replication is the online platform Political Science Replication Initiative,³⁹ which invites students and their course instructors to upload replication studies conducted in class. Greater awareness of the state of replication studies in teaching is crucial to "document the impact of promoting data access and research transparency principles" in universities (Carsey 2014:73).

Secondly, it will be beneficial if more teachers discuss new software developments that make reproducible research much easier and which can be taught to students as part of their methods and software training (Carsey 2014). For example, personal data repositories like GitHub,⁴⁰ or the use of Sweave and Knitr (Gandrud 2013), which integrate analysis code with text and figures, could establish an even more transparent way of working reproducibly. In many other fields, these developments have already been embraced, and in political science and international relations, more emphasis on data management in PhD programs as part of the skills set must be included as well.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. (1) a checklist of items necessary for a duplication versus a replication, (2) excerpts from syllabi assigning replication in methods courses, (3) an example syllabus for a stand-alone replication course, (4) a table on how to add value to a replication in the class room, and (5) what to do with the manuscript after class to get it published.

³⁹Political Science Replication Initiative (PSRI), see <http://projects.iq.harvard.edu/psreplication> (accessed November 16, 2014).

⁴⁰<https://github.com/> (accessed March 15, 2014).

References

- BELL, MARK S., AND NICHOLAS L. MILLER. (2013) Questioning the Effect of Nuclear Weapons on Conflict. *Journal of Conflict Resolution* 59 (1): 74–92.
- BISSELL, MINA. (2013) Reproducibility: The Risks of the Replication Drive. *Nature*. Available at <http://www.nsf.gov/eng/general/dmp.jsp> (Accessed March 15, 2014).
- CARSEY, THOMAS M. (2014) Making DA-RT a Reality. *PS, Political Science & Politics* 47 (1): 72–77.
- DAFOE, ALLAN. (2014) Science Deserves Better: The Imperative to Share Complete Replication Files. *PS, Political Science & Politics* 47 (1): 60–66.
- DAI, XINYUAN. (2002) Political Regimes and International Trade: The Democratic Difference Revisited. *American Political Science Review* 96 (1): 159–165.
- ELMAN, COLIN, AND DIANA KAPISZEWSKI. (2014) Data Access and Research Transparency in the Qualitative Tradition. *PS, Political Science & Politics* 47 (1): 43–47.
- FRANCIS, GREGORY. (2012) Publication Bias and the Failure of Replication in Experimental Psychology. *Psychonomic Bulletin & Review* 19 (6): 975–991.
- FRANK, MICHAEL C., AND REBECCA SAXE. (2012) Teaching Replication. *Perspectives on Psychological Science* 7 (6): 600–604.
- GANDRUD, CHRISTOPHER. (2013) *Reproducible Research With R and RStudio*. Boca Raton, FL: Chapman & Hall/CRC.
- GERBER, ALAN S., AND DONALD P. GREEN. (2000) The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment. *American Political Science Review* 94 (3): 653–663.
- GERBER, ALAN S., AND DONALD P. GREEN. (2005) Correction to Gerber and Green (2000), Replication of Disputed Findings, and Reply to Imai (2005). *American Political Science Review* 99 (2): 301–313.
- GHERGHINA, SERGIU, AND ALEXIA KATSANIDOU. (2013) Data Availability in Political Science Journals. *European Political Science* 12 (3): 333–349.
- GIBSON, JAMES L. (1995) Cautious Reflections on a Data-Archiving Policy for Political Science. *PS, Political Science & Politics* 28 (3): 473–476.
- HERRNSON, PAUL S. (1995) Replication, Verification, Secondary Analysis, and Data Collection in Political Science. *PS, Political Science & Politics* 28 (3): 452–455.
- HOEFFLER, JAN H. (2013) Teaching Replication in Quantitative Empirical Economics. World Economics Association (WEA), Conference on the Economics Curriculum: Towards a Radical Reformation: May 3–31.
- IMAI, KOSUKE. (2005) Do Get-Out-the-Vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments. *American Political Science Review* 99 (2): 283–300.
- ISHIYAMA, JOHN. (2014) Replication, Research Transparency, and Journal Publications: Individualism, Community Models, and the Future of Replication Studies. *PS, Political Science & Politics* 47 (1): 78–83.
- JANZ, NICOLE, SETH WERFEL, AND STEPHANIE WYKSTRA. (2014) Replication in Political Science Graduate Courses: An Untapped Resource? *The Monkey Cage Blog at the Washington Post*. Available at <http://www.washingtonpost.com/blogs/monkey-cage/wp/2014/02/12/replication-in-political-science-graduate-courses-an-untapped-resource/> (Accessed March 15, 2014).
- KING, GARY. (1995) Replication. *PS, Political Science & Politics* 28 (3): 443–499.
- KING, GARY. (2003) The Future of Replication. *International Studies Perspectives* 4 (1): 100–105.
- KING, GARY. (2006) Publication. *PS, Political Science & Politics* 39 (1): 119–125.
- KOOLE, SANDER L., AND DANIEL LAKENS. (2012) Rewarding Replications: A Sure and Simple Way to Improve Psychological Science. *Perspectives on Psychological Science* 7 (6): 608–614.
- LUPIA, ARTHUR, AND GEORGE ALTER. (2014) Data Access and Research Transparency in the Quantitative Tradition. *PS, Political Science & Politics* 47 (1): 54–59.
- LUPIA, ARTHUR, AND COLIN ELMAN. (2014) Openness in Political Science: Data Access and Research Transparency. *PS, Political Science & Politics* 47 (1): 19–42.
- MANSFIELD, EDWARD D., HELEN V. MILNER, AND PETER B. ROSENDORFF. (2000) Free to Trade: Democracies, Autocracies, and International Trade. *American Political Science Review* 94 (2): 305–321.
- MANSFIELD, EDWARD D., HELEN V. MILNER, AND PETER B. ROSENDORFF. (2002) Replication, Realism, and Robustness: Analyzing Political Regimes and International Trade. *American Political Science Review* 96 (1): 167–169.
- MEYER, WILLIAM H. (1999) Confirming, Infirmiting, and “Falsifying” Theories of Human Rights: Reflections on Smith, Bolyard, and Ippolito Through the Lens of Lakatos. *Human Rights Quarterly* 21 (1): 220–228.
- MILLER, ARTHUR, TOR WYNN, PHIL ULLRICH, AND MOLLIE MARTI. (2001) Concept and Measurement Artifact in Multiple Values and Value Conflict Models. *Political Research Quarterly* 54 (2): 407–419.

- MORAVCSIK, ANDREW. (2014) Transparency: The Revolution in Qualitative Research. *PS, Political Science & Politics* 47 (1): 48–53.
- NATIONAL SCIENCE FOUNDATION. (2011) NSF Data Management Plan Requirements. Available at <http://www.nsf.gov/eng/general/dmp.jsp> (Accessed March 15, 2014).
- PEFFLEY, MARK, PIA KNIGGE, AND JON HURWITZ. (2001a) A Multiple Values Model of Political Tolerance. *Political Research Quarterly* 54 (2): 379–406.
- PEFFLEY, MARK, PIA KNIGGE, AND JON HURWITZ. (2001b) A Reply to Miller et al.: Replication Made Simple. *Political Research Quarterly* 54 (2): 421–429.
- PRICE, MICHAEL. (2011) To Replicate or Not to Replicate? *Science Career Magazine*. December 2. Available at <http://tinyurl.com/b2q7g3a> (Accessed March 15, 2014).
- SCHNALL, SIMONE. (2014) Simone Schnall on Her Experience with a Registered Replication Project. *Society for Personality and Social Psychology Website*, May 23. Available at <http://www.spsblog.org/simone-schnall-on-her-experience-with-a-registered-replication-project/> (Accessed March 15, 2014).

Copyright of International Studies Perspectives is the property of Oxford University Press / USA and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.