Teaching Replication to Graduate Students

Stojmenovska, D.; Bol, T.; Leopold, T.

DOI
10.1177/0092055X19867996

Publication date
2019

Document Version
Final published version

Published in
Teaching Sociology

Original Article

AMERICAN SOCIOLOGICAL ASSOCIATION

Teaching Sociology
2019, Vol. 47(4) 303–313

DOI: 10.1177/0092055X19867996
ts.sagepub.com
SAGE

Teaching Replication to Graduate Students

Dragana Stojmenovska[1], Thijs Bol[1], and Thomas Leopold[1]

Abstract

Replicating published studies promotes active learning of quantitative research skills. Drawing on experiences from a replication course, we provide practical tips and reflections for teachers who consider incorporating replication in their courses. We discuss teaching practices and challenges we encountered at three stages of a replication course: student recruitment, course structure and proceedings, and learning outcomes. We highlight that by engaging in replication, students learn from established scholarly work in a collaborative and reflective manner. Students not only improve their quantitative literacy but also learn more generally about the scientific method and the production of research.

Keywords

replication, teaching replication, quantitative literacy, quantitative methods, sociology

In this article we advocate for engaging graduate students in replicating published studies. We highlight three main benefits of replication. First, students who replicate can learn from prominent scholars as they follow them step-by-step in designing and conducting a quantitative study. Second, replication gives students a more realistic view of the research process in the social sciences, involving numerous practical problems that are solved "behind the scenes" and cannot be observed in a published paper. Third, students can form a vital link in the replication chain, contributing to the accumulation of knowledge in a discipline by testing the robustness of important empirical findings. In this way replication has the potential to improve students' quantitative literacy beyond what can be achieved in standard courses on quantitative methods.

While there is no consensus on a single definition of quantitative literacy (Sharma 2017), two main properties can be discerned: (a) statistical skills that enable students to understand, conduct, and interpret analyses and (b) knowledge that enables them to critically evaluate the use of statistical methods. In other words, quantitative literacy entails abilities that go beyond merely reproducing a mathematical function. Social science research, with its hands-on application of quantitative skills to substantive questions, is particularly suitable for promoting students' quantitative literacy (Sweet and Strand 2006).

The literature on the challenges of teaching quantitative literacy points to the reliance on textbooks and traditional lecture methods as not sufficiently effective for developing students' ability to critically evaluate the processes that led to a particular finding (Howery and Rodriguez 2006). Empirical evidence on strategies for improving quantitative literacy suggests that teaching practices that engage students in hands-on activities are more effective. These include asking substantive questions relevant to students' lives (Atkinson, Czaja, and Brewster 2006; Burdette and McLoughlin 2010; Lindner 2012; Stroup et al. 2004), having students collaborate in groups

[1]University of Amsterdam, Amsterdam, The Netherlands

Corresponding Author:
Dragana Stojmenovska, University of Amsterdam, Nieuwe Achtergracht 166, Amsterdam, 1018 WV, The Netherlands.
Email: d.stojmenovska@uva.nl

(Caulfield and Caroline 2006), write reflective learning journals (Denton 2018) or collect their own data (Strangfeld 2013), and using computers for working with data (Wilder 2009).

Replicating existing studies is another way in which active learning can take place. In replicating, students closely study established scholarly work, experience the complete inquiry cycle and learn through collaboration and reflection. These benefits improve not only students' research skills but also their understanding of the production of science more generally. While the value of replication for the scientific community has received substantial attention, these pedagogical aspects of replication are far less acknowledged. The learning benefits of replication have received some attention in psychology (Chopik et al. 2018; Frank and Saxe 2012; Grahe et al. 2012), economics (Ball and Medeiros 2012), and political science (Janz 2016; King 2006) but not in sociology.

The goal of this article is to encourage replication in teaching sociology, drawing on experiences from teaching courses on replication in the Research Master Social Sciences program at the University of Amsterdam. In the remainder of the article we discuss three different stages of a course on replication: (1) recruiting students to join the course (starting the course), (2) course structure and proceedings (during the course), and (3) learning outcomes (finishing the course). The article concludes with a discussion of how active learning can take place through replication and how replication contributes to quantitative literacy and to the discipline of sociology more broadly.

COURSE DESCRIPTION

The course Replication, taught by two of the authors, is a stand-alone course offered within the Research Master Social Sciences program at the University of Amsterdam. The Research Master Social Sciences is a two-year research master program admitting about 35 students each year. The focus is on advanced research methodology and theory to prepare students for a career in academia or related fields. About 50 percent of the students who graduate from this program go on to pursue a PhD degree. The course is offered in the second year of the program. This means that students who enter the course already have advanced knowledge of quantitative methods in the social sciences. The duration of the course is seven weeks, with sessions of four hours each week. The total workload of 168 hours—including face-to-face meetings and expected number of self-study hours—is similar to other methodology and theory courses offered in the program.

The course is open to students in the quantitative or mixed methods track of the Research Master Social Sciences (about 20 students) as well as all PhD students from the Amsterdam Institute for Social Science Research, which hosts the departments of sociology, geography, planning and development studies, political science, and anthropology (about 300 students). While students do not necessarily have to be trained in sociology to participate in the course, most participants come from sociology and political science.

The formal requirements for participating in the course are familiarity with ordinary least squares (OLS) and logistic regression. Often students already have additional methodological skills, notably on multilevel and longitudinal methods of analysis. The instruction language is English, and the statistical analysis software used in the teaching materials is Stata, although using other software, such as R, is allowed. Given the importance of close and continued contact between the lecturers and the students, attendance of all class meetings is mandatory, and students are expected to come prepared and participate actively in all sessions.

RECRUITING STUDENTS

The main selling point of the course is learning about replication as a defining feature of the scientific method and an essential tool to verify and extend published findings. We emphasize students' freedom to shape the course according to their research interest by replicating a published paper of their choice. A further motivating factor is that the outcome of the course, a replication article, may be submitted to a professional journal.

As of 2019, we have taught the course three times, recruiting five students for the first iteration, eight for the second, and seven for the third. While the small number of participants may suggest low interest to engage in replication, it is important to note that the course was not obligatory, and the primary student population (20 students enrolled in the quantitative or mixed-methods track of the Research Master Social Sciences) is small. At the same time, the small group size is essential for the success of the course. The activities of the course necessitate a small scale, close student–teacher contact, and individual supervision. Given the demands of replicating a research article in only a few weeks, we recommend a student-to-teacher

ratio of 5 or less to 1. To ensure close supervision of every student, two teachers were present during all sessions.

Whereas all interested students were permitted to enroll in the course, participants self-selected positively on quantitative literacy. This means that their quantitative methods skills and interests were more advanced than those of the average student already before engaging in the course. Almost all participants planned to pursue an academic career, had heard about replication before enrolling in the course, or had even engaged in replication in previous courses or workshops. One could argue that if anyone needs to be engaged in replication, it is exactly this group of skilled and motivated students with academic aspirations. At the same time, there are benefits to addressing a larger and more diverse student population. In future iterations of the course, we plan to attract more students by placing a stronger emphasis on the learning outcomes of engaging in replication, beyond framing replication as an important feature of the scientific process.

COURSE PROCEEDINGS

Introducing Students to Replication

The course starts by introducing students to the concept and practice of replication and the criteria for choosing a published article to replicate. We start with a narrow concept of replication as "duplication," whereby students try to get the exact same results using the same data and methods as described in the published article. In articles that present a large set of analyses, students can limit their exact replication to the article's core results.

For the purposes of teaching the course, exact replication has several advantages compared to conceptual replications that use different data and/or different methods to answer the same research questions. First, students follow the work of professional scholars step-by-step. The article is used as a "blueprint" that provides important guidance as students start their projects. Second, exact replication requires students to reconstruct not only key analytical decisions described in detail in the article but also the many small decisions that are often insufficiently or not described. In this way, students get firsthand experience with the difficulties and importance of documenting a quantitative study in a fully transparent way.

Third, exact replication streamlines the course proceedings because all students have to solve a similar set of problems—recovering the analytic sample, reconstructing all variables, and reconstructing the models used in the analysis. Fourth, the exact replication forms a base from which students can add value. Toward the end of the course, students are asked to perform at least one extension that adds to the study that they replicate (see Adding Value section).

In the first two sessions of the course, students are also asked to read about and reflect on debates surrounding replication in sociology. For example, we discuss potential problems with replication considering the methodological variety in the quantitative social sciences, the risk of replicating at the start of a career, and concerns about code as intellectual property. We review various examples of replication studies in the social sciences and discuss arguments in favor of (Freese 2007) or against (Abbott 2007) stricter replication standards in sociology.

Choosing the Article

The key assignment for students in the first week of the course is to choose a published article for their replication projects. Students are presented with three criteria for picking an article: (1) interest, (2) impact, and (3) feasibility. Choosing a paper that falls within their research interests increases students' motivation to work on the replication. Knowing the field also helps to evaluate the article and come up with an extension to the original analysis. After surveying their research interests, students are instructed to look for either an article that was published in a flagship journal of the discipline or a subfield or one that has been cited frequently (in the case of older papers), as indicated by Web of Science or Google Scholar.

This impact criterion assumes that articles published in flagship journals and/or highly cited articles are of higher quality and have made important contributions to the field, thereby ensuring that students are "learning from the best." Students can also choose to engage in a recent scholarly debate by choosing to replicate a recent study published in a flagship journal, even if it has few or no citations. While one can be critical of this assumption—many important articles are not published in high-impact-factor journals (Amin and Mabe 2003)—focusing on flagship journals provides a convenient filter for students to find their way around the large number of published articles in a (sub)field.

Finally, replicating the chosen article needs to be feasible in terms of data access and complexity of analysis. Students should be able to get access to the data fast and easily, preferably by downloading

the data directly or getting access via the authors of the original articles. Since the course takes place in a relatively short time frame (seven weeks), obtaining access to restricted data is often not possible. Given the time constraints, the analyses in the article should not be too complex in the sense that the students and the teachers are at least broadly familiar with the methods that were used. Feasibility is of general importance but especially in courses of relatively short duration, such as ours. For the second session of the course, we ask students to prepare short presentations of at least two potential papers to replicate.

Preparing to Start the Project

The second week marks the start of the student projects. Most time is spent on discussing which article each student should choose to replicate. We carefully discuss the opportunities and challenges of the studies that students have selected as potential replication papers. Each student presents two candidate papers for the replication project. The goal of this presentation, and the class discussion that follows each presentation, is to help the student make a good choice of an article to replicate. In their presentations, students describe (1) the premise of the paper, (2) the data and empirical design, (3) the key results, (4) why they think the paper would be a good candidate for the replication course, and (5) potential problems that might arise in the replication.

As instructors, our main input concerns the criteria of feasibility and impact. Feasibility is our primary concern, given that students need to complete their replication within the remaining six weeks of the course. Our experience with the presentations indicates that students often underestimate the time demands of data preparation. We advise against articles if data access takes more than a few days and/or if data preparation requires large investments of time. The latter assessment is based on the teachers' experience and often not more than an informed guess, as it is difficult to assess the complexity of data preparation solely on the basis of a research article. As problems with data access and/or preparation put a student's replication project at a risk of failure within the time boundaries of the course, our advice at this stage is conservative: our primary concern is that the replication should be doable within a limited amount of time.

Impact and quality are other concerns that usually require teacher feedback. Even skilled and motivated students have difficulties in assessing which articles are of high impact or quality. In the first iteration of the course, we allowed a student to choose an article that was very close to the student's research interest but poorly conducted and lacking any influence on the field. Although the criteria of student interest and feasibility were met, the student had to learn from a bad example instead of from a good one. Moreover, that the article received very little attention in the field undermined the motivation to replicate the study. Based on this experience, we advised more strongly against low-impact studies in later iterations of the course. Moreover, if a study appeared to be poorly conducted at first glance, we advise against replicating it unless it had a high impact.

In some cases, we advised against all studies proposed by a student in the second week and asked the student to start over and find new papers. Although this process can be frustrating, intervening at an early stage is important to prevent failure of projects at a later stage. Table 1 lists the journals students chose to replicate articles from, the data used in the articles, and the way of accessing the data.

Although we spend most of the second week deciding on which papers students should replicate, we also discuss the bigger picture in which their replication projects are embedded. To this end, we ask each student to prepare a brief presentation on an existing replication case of his or her own choice. Each student presents and discusses one classic or recent example of replication studies in the social sciences. In their presentations, students are asked to describe their example and explain what we can learn from it when it comes to the benefits and problems associated with replication in the social sciences.

Starting the Project: Creating an Effective Workflow

In week 3, we focus on how to organize a project and create an effective workflow of data analysis. In our teaching material, we illustrate this using Stata, but students can adapt the principles to other software. Our goal is that students start their projects based on a clear and effective structure of organizing their data and documentation. We follow textbook recommendations (e.g., Kohler and Kreuter 2012; Long 2008) that can be adapted to individual needs and preferences. Students are encouraged to use a template that includes folders for data files, code, log files, tables, and figures as well as configuration files and ado-files that

Table 1. Journals, Data, and Data Access of Replicated Articles.

Journal	Data	Data Access
American Journal of Political Science	Author's data	From website
American Journal of Political Science	Multiple data sources	From website
American Journal of Sociology	National Longitudinal Survey of Youth 1979	After registration
American Sociological Review	National Organizations Survey 1996–1997	After registration
American Sociological Review	World Values Survey	After registration
Archives of General Psychiatry	National Longitudinal Study of Adolescent to Adult Health (Add Health)	Only subsample of data available
European Journal of Political Research	European Values Survey	After registration
European Political Science Review	European Social Survey, Comparative Manifesto Project	After registration
European Sociological Review	European Social Survey, World Bank indicators	After registration
European Sociological Review	European Social Survey	After registration
European Union Politics	Eurobarometer	After registration
International Organization	European Social Survey	After registration
Journal of Child and Family Studies	Add Health	Only subsample of data available
Migration Studies	European Social Survey	After registration
Journal of Family Violence	World Values Survey	After registration
Party Politics	Eurobarometer	After registration
Perspectives on Politics	Author's data	From website
Poetics	Author's data	From website
Political Studies	International Social Survey Programme	After registration
The Economic Journal	Socio-Economic Panel (Germany)	After registration

Note: This list is based on the 20 articles replicated by the students we have taught in the three iterations of our course.

facilitate data analysis and documentation. These folders constitute a so-called replication package that can be used by others to reproduce all tables and figures resulting from the analyses conducted by the student. A replication package typically contains a README file that gives a structured overview of the different types of content stored in separate folders and explains which parts of the code should be run to reproduce the analyses. By setting the path of the replication files in the do-file (where code in Stata is documented), students can directly save modified data, tables, and figures in the respective folders of the replication package.

We also require students to use extensive annotation during their data preparation and analysis. The overall structure of data preparation and analysis as well as all decisions regarding sample selection and coding of variables need to be explained in the do-files. The raw data remain unchanged, and any amendment is documented in the do-file and stored in a working directory. Establishing a reproducible filing and logging system at the beginning of the project means that the students have shareable replication packages ready at the end of the project. Each project is usually divided into two separate steps: data preparation and data analysis. We ask students to carry out all changes to the data, such as merging, sample cuts, and coding variables, in a preparation do-file. The analysis do-file should produce all empirical material included in the replication article. Figure 1 gives an example of an annotated Stata do-file used for running analyses.

To track the progress of each project, we ask students to prepare weekly reports in which they document each step that they have taken as they proceed with the replication, the problems that they are facing, and questions for the lecturers. Students

```
*=================================================================*
*                        Description of dofile                    *
*=================================================================*

This dofile generates Table 1 and Figure 1 that are presented in [Name Article].
The structure of this dofile is as follows:
  1. Setting paths and opening data
  2. Analysis (Table 1)
  3. Robustness checks (Figure 1)

*=================================================================*
*                  1. Setting paths and opening data              *
*=================================================================*

global dir "SET YOUR PATH"
global pathd "$dir/data"
global pathp "$dir/posted"
global patht "$dir/tables"
global pathf "$dir/figures"
global pathl "$dir/logfiles"
set more off
cap log close
log using "$pathl/analyses.log", replace
use "$pathp/workingdata.dta", clear

*=================================================================*
*                        2. Analysis                              *
*=================================================================*

/*Table 1. [Table Title as in Article]*/
xtreg depvar var1 var2 var3 var4, fe i(serno)
eststo m1
esttab m1 using "$patht\Table 1.rtf", [options]

*=================================================================*
*                     3. Robustness checks                        *
*=================================================================*

/*Figure 1. [Figure Title as in Article]*/
margins i.var1, at(var2=(0(10)100))
marginsplot, [options]
graph export "$pathf/Figure 1.pdf", replace
**end do-file
log close
```

Figure 1. Annotated Stata do-file.

are also asked to include annotated code in their progress report. From this point onward in the course, students generally follow their own route. Some students spend little time on data preparation, while others struggle with getting the data in shape. This means that the course becomes increasingly focused on the individual, and teacher–student discussions are often one-to-one.

Continuing the Project: Avoiding and Detecting Errors

In week 4, all student projects are running and some may have already completed their data preparation. The teachers' general input in this week is avoiding and detecting errors. Several practical tips are given. These include making a plan that describes

and puts into a logical sequence all the steps needed to get from the raw data to the prepared data that can be analyzed. Other advice focuses on efficient programming techniques that minimize code and thus the potential for error. Data preparation usually contains repetitive elements, and loops can decrease the amount of code but also the clarity of it. Moreover, we caution students against trusting their eyes. Students are instructed to not rely on browsing through a subset of their data to check whether a command has worked as expected. Instead, error-detecting routines should rely on logical checks that can uncover any inconsistencies in the entire data.

Another important piece of advice is to design a step-by-step procedure for selecting the analytic sample, starting from a clearly defined starting sample (usually the raw data) and specifying cases and observations lost due to each sample cut, as per the standards of the American Psychological Association. Our experience with replication projects in this course is that sample cuts are often poorly documented in research articles, and almost all students encounter problems in recovering the analytic sample of the paper that they replicate.

Finally, we encourage students to extensively use descriptive statistics to find errors in their data before moving on to multivariate analyses, during which it is much more difficult to detect data preparation errors. We instruct students to carefully inspect statistics, such as averages, in order to detect anomalies and inconsistencies. In addition to carefully looking at all descriptive statistics, students sometimes need a substantive understanding—knowledge of the relevant literature—to assess what is unusual. For example, knowledge about the age demographics of union dissolution is required to assess whether the risk of divorce found in a certain sample is unusually high or low (for an example, see Karraker and Latham 2015). Often, however, common sense suffices. This can be as simple as identifying a coding error by finding that the maximum value on the variable age in years is 999 (for an example of this type of error, see Stojmenovska, Bol, and Leopold 2017).

To summarize, our four practical tips for avoiding and detecting errors are (1) plan well; (2) less code, fewer errors; (3) don't trust your eyes; and (4) do descriptives.

Project Nearing Completion: Adding Value to the Original Study and Cross-Checking Code

In general, students are free to choose what they consider to be the best ways of adding value to the original article. After spending several weeks on replicating the study, students often have various ideas for extensions to the original analyses. This means that teacher advice focuses primarily on prioritizing and evaluating the options in terms of feasibility.

A crucial message to convey is that any extension needs to be substantiated by a theoretical and/or a methodological concern. That is to say, it should not be done just for the sake of extending the study in some way. We included the extension element in the course because students often come up with ideas on how to improve or change the analysis during their replication of the original findings. This part of the course provides the space to do so and gives students the opportunity to adjust aspects of the analyses that they would have done differently or check if the results hold when using other data or methods.

In week 5, we also pair up students and ask them to cross-check each other's code. We ask them to evaluate the code of their classmate with the following questions in mind: Is the code clearly structured and well annotated? Is the code written efficiently and is it free of errors? Is there other room for improvement? Students provide feedback to one another, and in week 6 we have a class discussion about important issues that came up in the process of cross-checking code. Our experience is that checking the code of someone else is beneficial not just for the author of the code but also for the evaluator. There are always different approaches to similar problems that will provide useful insights for one's own research practices.

Contacting the Authors

In some cases, students have to contact the authors of the original articles for clarification. Writing the e-mail with the right tone is important as some authors are not welcoming of replication, among other reasons because they feel that their work is under investigation. Our advice to students in case of discrepancies that cannot be reconciled is to always assume that the error is with the replicator, not the author of the original study. This assumption, which often turns out to be warranted, sets the tone of the e-mail when contacting authors. We ask students to send us a draft of the e-mail for review before they contact the authors.

Completing the Project: The Replication Paper

The final two weeks of the course are dedicated to finishing the replication paper, to which

the following guidelines apply. The paper should consist of four main parts: (1) introduction, (2) replication, (3) extension, and (4) conclusion. The introduction should be brief and start with the problem that the original paper has addressed and its relevance. The latter is used to motivate the replication and the extension. If applicable, the introduction should also define the scope of the replication (if full or partial, and why).

In the replication section, students describe the exact replication in a brief and focused way, presenting their empirical material closely aligned with the material presented in the original study. In writing whether the results replicate, we caution against a binary criterion of yes/no (for an example of multiple standards used for evaluating replication success, see Open Science Collaboration 2015). Instead, we ask the students to make a theoretically and empirically informed statement of whether they think the replication was successful or not. This means that students should focus on what is most relevant for replicating the original paper, that is, the key findings, and evaluate, in case of discrepancies, whether these are theoretically meaningful. Students should generally describe similarities or deviations in terms of point estimates and significance levels for all key findings. In the extension section, we ask students to show how the extension adds value to the original analyses and what the results of the extension mean for the substantive conclusions of the article. The conclusion of the article should be brief, summarizing the results from the replication and the extension, adding interpretation, and discussing implications.

LEARNING OUTCOMES

The course has three main learning objectives. Students who complete the course should be able to (1) conduct a replication study according to pertinent standards of reproducibility and add value to a published study, (2) understand what decisions and difficulties authors of published work have typically faced, and (3) critically evaluate the publication process and the role of replication in the social sciences.

The first objective was met by all students, as at the end of all iterations of the course all students submitted an article in which they presented results from their replications and extensions, accompanied by replication packages. By the end of the course, students had finished their replication projects in accordance with the completion criteria. In some cases, where potential for sharing the project results outside of the classroom was present, students continued working on their projects after the official end of the course. Judging from the quality of their replication packages and attention to detail, students had learned how to build a structured workflow that guides their work, at the same time making it easier for the instructors (and other potential parties) to evaluate their work. Students' extensively annotated code and replication packages followed high standards of reproducibility. These include the American Psychological Association standards of documenting the selection of analytic samples and the *American Journal of Political Science* standard of full correspondence between the replication package and all empirical material shown in the paper. In the course evaluations of the 2018 iteration, a student stated that the replication course "will help [them] write [their] own study."

The replication paper also served as basis for evaluating how students added value to the original studies. Students' extensions of (parts of) the original analyses ranged from running alternative models to using other or more recent data. In some cases, the student found errors in the original study, in which case both the replication and extension focused on correcting these errors. One example is the replication of a study by Herring (2009). The student encountered two problems with the original article. The first problem emerged while the student performed the replication. Correcting an error in the coding of the dependent variables (missing values were treated as substantive values) led to a considerably smaller sample size in the replication study than in the original study. The second problem emerged in the adding-value stage of the project. The student found that the distribution of a key control variable used in the original analyses was heavily skewed, and the results of the study were not robust to a correction of this skew. This replication paper resulted in a published comment (Stojmenovska et al., 2017).

It is important to note that most replication projects did not find errors that invalidated the original study's conclusions. However, in the early stages of a replication project, students often encountered problems such as recovering the analytic sample or reproducing all variables. These problems more often reflect insufficient documentation of important analytic decisions than errors made by the authors of a published study. As the replication projects progressed, students' results often converged with the results shown in the original study. It is therefore important to alert students that

extensive checks are needed before concluding that an original study contains an error. If extensive checks cannot reconcile the replication with the original study, we advise students to contact the authors for clarification. Students should always work from the assumption that the discrepancies they (initially) found are not due to errors made by the original authors, an assumption that in our experience is often warranted. If students claim to have found an error, this conclusion must be supported by systematic, extensive, and fully documented analysis and take into account the response of the original authors, if available.

The second objective, understanding what decisions scholars face in their research, forms a central part of the weekly reports (one page) that students wrote but also the replication paper. The instructors used the weekly reports not only as a reference for individual supervision but also to highlight overarching issues for discussion in the group. Important decisions with respect to coding, sampling, or modeling often form the backbone of the replication papers. What would have happened if the researcher had taken a different path? This means that even if the findings replicate, students learn about decisions that mostly take place behind the curtains. In the student presentations given toward the end of the course, students focused on presenting the results of taking different paths, thereby reflecting on the type of decisions often faced in research. One student, for example, reflected on the different results obtained from using pooled regression models with interactions involving the predictor of interest, educational level, as opposed to using single regression models for the different educational levels. Another student discussed how using more recent data and a larger number of measures for the outcome variable yields different results from those presented in the original article.

Finally, the weekly progress reports and written (anonymous) evaluations show that students also critically reflected on the publication process and the role of replication in research. The progress reports reveal that for virtually all students, it became clear quite quickly that most published articles do not contain sufficient information to reproduce all relevant findings. This in itself is an important learning experience for students who aim for a career in academia. A student reflected on the publication process in the following evaluation of the Replication course in 2015:

> It was amazing to actually work statistically on published articles. It was a great eye-opener in that publishing does not equal quality.

At a more general level, the course aims to professionalize students into their discipline and to develop a transparency routine for their future careers. In the course evaluations of the 2017 reiteration, a student touched upon on the broader role of replication in research:

> I learned a lot in terms of ethical research, working more diligently [on] the methods I applied for my replication.

This statement addresses ethical principles of research, the teaching of which has been promoted by others (e.g., Lowney 2014). Replication provides the type of experiential learning of research ethics found to be achievable through other experiential techniques (for a review, see Teixeira-Poit, Cameron, and Schulman 2011).

The course has consistently received very high student evaluation scores, indicating high popularity. However, it should be noted that the course was small, it was voluntary, and students were positively selected on motivation and academic ability. Moreover, student evaluation scores are not indicative of student learning, as shown in a recent meta-analysis (Uttl, White, and Gonzalez 2017).

ACTIVE LEARNING THROUGH REPLICATION

In this article we have argued that by engaging with a study in depth, from data preparation and methods to presentation of results, students gain a detailed understanding of all steps involved in a quantitative empirical analysis. This includes all the challenges and complications that data analysis typically brings as well as strategies of how to resolve them.

Unlike learning from methods textbooks—which are often stylized and focus only on certain segments of the research process, most commonly, the analysis and interpretation part—replication takes students through the complete cycle of inquiry. As a result, students gain an in-depth understanding of the steps that are taken from raw data to published analyses, exploring the highest

level of detail up to the smallest decisions that need to be taken in order to prepare and analyze quantitative data. The learning outcome is a professional level of quantitative literacy that, in our view, is hard to achieve by other means.

In our course, keeping track of students' learning progress was facilitated through continuous progress reports, close student–teacher contact, class discussions, and cross-checking of code. Documenting each step the student takes in the form of progress reports resembles the use of reflective journals in class, which has shown positive effects on student learning (Denton 2018). Students learned not only from the teachers but also from each other by discussing the challenges they have encountered in a small-class setting. In doing so, students become the source of knowledge rather than passive recipients.

Our course takes place in a relatively short time frame of seven weeks. For courses of longer duration, we recommend that flexibility remain the guiding criterion for choosing articles to replicate. Even with more time at hand, students should be able to get access to the data quickly and easily, and the level of complexity of the analyses should not exceed the capabilities of the students and the teachers. If given more time, we would pay more attention to two elements of the course: the extension and the rounding off of the projects. Given that getting the data in shape and reconstructing the samples from the original articles took a substantial amount of time, students often did not have enough time to execute the extension the way they had ideally envisioned. This meant that the replication itself had more weight in most students' projects. An additional interesting activity would be to open up course discussions to a broader audience of students who did not participate in the course. This can be achieved when rounding off the projects at the end of the course, for example, by inviting other students from the graduate program to watch the course participants present their projects. That way, a larger number of students can be involved in thinking about replication and research practices more broadly.

Next to offering a stand-alone course on replication, teachers could consider incorporating replication into their existing quantitative methods courses. This can be as simple as having students read an article or book chapter about replication and transparent social science research (e.g., Christensen, Freese, and Miguel 2019) and discuss the arguments for and against replication. A more hands-on exercise could be asking students to recover the sample size or reproduce a core table in a published paper chosen by the teacher. We incorporated this activity in a statistics course for second-year undergraduate students. To streamline the exercise, we provided students with the data used in the paper and asked all students to reproduce the same table. We recommend choosing a paper with relatively simple analyses and well-documented data handling.

What does teaching replication to graduate students mean more broadly for the discipline of sociology? In our view, engaging students in replication contributes to the improvement of practices of transparency and research ethics. Recently, the editors of *American Sociological Review* have argued that "many submissions fail to provide enough detail on… sample size, missing data, decisions about who is included or excluded from the sample, what types of models are estimated and why… and so forth" (Mustillo, Lizardo, and McVeigh 2018:3). Considering their importance, these are major issues for which better practices are required. The Code of Ethics of the American Sociological Association also touches upon issues of reporting, in reading that "[i]n presenting their work, sociologists [should] report their findings fully and do not omit relevant data" and "take particular care to state all relevant qualifications" (American Sociological Association 2018:15). By asking students to clarify and annotate each step of data preparation and analysis, we socialize them into establishing transparent and replicable standards for future studies of their own. This, we believe, makes for a better sociology of the future.

EDITOR'S NOTE

Reviewers for this manuscript were, in alphabetical order, Matthew May and Kristjane Nordmeyer.

REFERENCES

Abbott, Andrew. 2007. "Notes on Replication." *Sociological Methods & Research* 36(2):210–19.

American Sociological Association. 2018. "Code of Ethics." Retrieved July 5, 2019 (https://www.asanet.org/sites/default/files/asa_code_of_ethics-june2018.pdf).

Amin, M., and M. A. Mabe. 2003. "Impact Factors: Use and Abuse." *Medicina (Buenos Aires)* 63(4):347–54.

Atkinson, Maxine P., Ronald F. Czaja, and Zachary B. Brewster. 2006. "Integrating Sociological Research into Large Introductory Courses: Learning Content and Increasing Quantitative Literacy." *Teaching Sociology* 34(1):54–64.

Ball, Richard, and Norm Medeiros. 2012. "Teaching Integrity in Empirical Research: A Protocol for

Documenting Data Management and Analysis." *Journal of Economic Education* 43(2):182–89.

Burdette, Amy M., and Kerry McLoughlin. 2010. "Using Census Data in the Classroom to Increase Quantitative Literacy and Promote Critical Sociological Thinking." *Teaching Sociology* 38(3):247–57.

Caulfield, Susan L., and Hodges Persell Caroline. 2006. "Teaching Social Science Reasoning and Quantitative Literacy: The Role of Collaborative Groups." *Teaching Sociology* 34(1):39–53.

Chopik, William J., Ryan H. Bremner, Andrew M. Defever, and Victor N. Keller. 2018. "How (and Whether) to Teach Undergraduates about the Replication Crisis in Psychological Science." *Teaching of Psychology* 45(2):158–63.

Christensen, Garret, Jeremy Freese, and Edward Miguel. 2019. *Transparent and Reproducible Social Science Research: How to Do Open Science*. Berkeley: University of California Press.

Denton, Ashley Waggoner. 2018. "The Use of a Reflective Learning Journal in an Introductory Statistics Course." *Psychology Learning & Teaching* 17(1):84–93.

Frank, Michael C., and Rebecca Saxe. 2012. "Teaching Replication." *Perspectives on Psychological Science* 7(6):600–604.

Freese, Jeremy. 2007. "Replication Standards for Quantitative Social Science: Why Not Sociology?" *Sociological Methods & Research* 36(2):153–72.

Grahe, Jon E., Alan Reifman, Anthony D. Hermann, Marie Walker, Kathryn C. Oleson, Michelle Nario-Redmond, and Richard P. Wiebe. 2012. "Harnessing the Undiscovered Resource of Student Research Projects." *Perspectives on Psychological Science* 7(6):605–607.

Herring, Cedric. 2009. "Does Diversity Pay? Race, Gender, and the Business Case for Diversity." *American Sociological Review* 74(2):208–24.

Howery, Carla B., and Havidan Rodriguez. 2006. "Integrating Data Analysis (IDA): Working with Sociology Departments to Address the Quantitative Literacy Gap." *Teaching Sociology* 34(1):23–38.

Janz, Nicole. 2016. "Bringing the Gold Standard into the Classroom: Replication in University Teaching." *International Studies Perspectives* 17(4):392–407.

Karraker, Amelia, and Kenzie Latham. 2015. "Authors' Explanation of the Retraction." *Journal of Health and Social Behavior* 56(3):417–19.

King, Gary. 2006. "Publication, Publication." *PS: Political Science and Politics* 39:119–25.

Kohler, Ulrich, and Frauke Kreuter. 2012. *Data Analysis Using Stata*, 3rd ed. College Station, TX: Stata Press.

Lindner, Andrew M. 2012. "Teaching Quantitative Literacy through a Regression Analysis of Exam Performance." *Teaching Sociology* 40(1):50–59.

Long, J. Scott. 2008. *The Workflow of Data Analysis Using Stata*. College Station, TX: Stata Press.

Lowney, Kathleen S. 2014. "Ask *Teaching Sociology*: What Should Faculty Consider before Having Students Conduct Research in a Class?" *Teaching Sociology* 42(3):240–44.

Mustillo, Sarah A., Omar A. Lizardo, and Rory M. McVeigh. 2018. "Editors' Comment: A Few Guidelines for Quantitative Submissions." *American Sociological Review* 83(6):1281–83.

Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349(6251):aac4716.

Sharma, Sashi. 2017. "Definitions and Models of Statistical Literacy: A Literature Review." *Open Review of Educational Research* 4(1):118–33.

Stojmenovska, Dragana, Thijs Bol, and Thomas Leopold. 2017. "Does Diversity Pay? A Replication of Herring (2009)." *American Sociological Review* 82(4):857–67.

Strangfeld, Jennifer A. 2013. "Promoting Active Learning: Student-Led Data Gathering in Undergraduate Statistics." *Teaching Sociology* 41(2):199–206.

Stroup, Donna F., Richard A. Goodman, Ralph Cordell, and Richard Scheaffer. 2004. "Teaching Statistical Principles Using Epidemiology." *American Statistician* 58(1):77–84.

Sweet, Stephen, and Kerry Strand. 2006. "Cultivating Quantitative Literacy: The Role of Sociology." *Teaching Sociology* 34(1):1–4.

Teixeira-Poit, Stephanie M., Abigail E. Cameron, and Michael D. Schulman. 2011. "Experiential Learning and Research Ethics: Enhancing Knowledge through Action." *Teaching Sociology* 39(3):244–58.

Uttl, Bob, Carmela A. White, and Daniela Wong Gonzalez. 2017. "Meta-analysis of Faculty's Teaching Effectiveness: Student Evaluation of Teaching Ratings and Student Learning Are Not Related." *Studies in Educational Evaluation* 54:22–42.

Wilder, Esther Isabelle. 2009. "Responding to the Quantitative Literacy Gap among Students in Sociology Courses." *Teaching Sociology* 37(2):151–70.

AUTHOR BIOGRAPHIES

Dragana Stojmenovska is a PhD candidate in the Department of Sociology at the University of Amsterdam. Her research affinities lie with gender and gendered social structures. Recent publications include "Management Gender Composition and the Gender Pay Gap: Evidence from British Panel Data" (*Gender, Work & Organization*, 2019).

Thijs Bol is an associate professor in the Department of Sociology at the University of Amsterdam. His research interests are in social stratification and inequality in education, the labor market, and science. Recent publications include "School-to-Work Linkages, Educational Mismatches, and Labor Market Outcomes" (*American Sociological Review*, 2019).

Thomas Leopold is an associate professor in the Department of Sociology at the University of Amsterdam. His research interests are in family relations, demography, and social stratification. Recent publications include "Gender Differences in the Consequences of Divorce: A Study of Multiple Outcomes" (*Demography*, 2018).