SOCIETY FOR THE TEACHING OF PSYCHOLOGY

Teaching of Psychology
1-9

DOI: 10.1177/0098628319872605
journals.sagepub.com/home/top
SAGE

Statistical Software Output in the Classroom: A Comparison of R and SPSS

Jacob B. Rode[1] and Megan M. Ringel[1]

Abstract

The current study examines whether R or SPSS output induces greater initial anxiety in students and whether anxiety toward one or both changes after being taught one type of software output throughout the course. The authors each taught an introductory statistics course, with the first course ($n = 43$) teaching R output exclusively and the second course ($n = 39$) teaching SPSS output exclusively. Students in both courses were given surveys assessing their anxiety and confidence toward R and SPSS output on the first and last days of class. Students initially reported greater anxiety and lower confidence when viewing R compared to SPSS output. However, the initial difference between R- and SPSS-related anxiety and confidence level disappeared when students were taught R and decreased substantially when students were taught SPSS. The results suggest that although R output may seem more intimidating initially, students adapt to it as well as they do to SPSS.

Keywords

SPSS, R software, statistics anxiety, software output

Statistics anxiety is a major concern in undergraduate introductory statistics courses (Zeidner, 1991). For social science majors in particular, many students lack statistical or mathematical backgrounds and thus enter their first statistics course with anxiety and a low sense of self-efficacy (Onwuegbuzie & Wilson, 2003). In response to this, instructors have tried a wide variety of approaches to improve statistical education and decrease student anxiety. One approach—recommended by the Guidelines for Assessment and Instruction in Statistics Education (GAISE) college report—involves decreasing the emphasis on hand calculation of formulas in favor of learning to conduct and interpret data analysis more realistically using statistical software (GAISE College Report, 2016). Aligned with these recommendations, many instructors expose students either to the actual use of statistical software through lab-based assignments or to interpretation of statistical tables and graphs generated from popular software (Bartz & Sabolik, 2001; Sosa, Berger, Saw, & Mary, 2011).

Considerable research demonstrates that computer-based tools such as common statistical software can improve students' understanding of and attitudes toward statistics (e.g., Ciftci, Karadag, & Akdal, 2014; McCulloch, 2017; Sosa et al., 2011). Moreover, Goal 8 from the recent GAISE report highlighted the importance of having students learn from statistical software, stating, "Students should be able to interpret and draw conclusions from standard output from statistical software" (GAISE College Report, 2016, p. 11). However, instructors must also remain cognizant of their students' differing levels of prior knowledge and computer skills, as well as the anxiety that may arise when learning both statistics and a new software program. Considering that instructors are increasingly incorporating the use of statistical software and/or statistical output from programs ranging from Excel to SPSS, SAS, and R, it would be beneficial to better understand how incorporation of statistical software may increase or decrease student anxiety.

Given instructors' desire to minimize student anxiety, is one type of software superior to another in this regard? Little research has been conducted on this topic, and the use of R in particular is an understudied area. Some have argued that R should be taught to enhance statistics education (Baumer, Cetinkaya-Rundel, Bray, Loi, & Horton, 2014; Farrell & Carey, 2018; Ferreira, Kataoka, & Karrer, 2014; Franklin, 2018; Yu, 2017). Baumer et al. (2014) reported favorable results teaching undergraduate students in introductory statistics courses at Duke University and Smith College how to use R and R Markdown (Allaire et al., 2018), noting that students' lack of prior coding experience and initial frustration learning R was not an impediment to either their performance or their reported enjoyment of the class at the end of the semester. However, there were many features of the courses that may not be representative of typical introductory statistics courses for social science majors, including the use of a flipped

[1] Department of Psychological Science, University of California Irvine, CA, USA

Corresponding Author:
Jacob B. Rode, Department of Psychological Science, University of California Irvine, 4201 Social & Behavioral Sciences Gateway, Irvine, CA 92697, USA.
Email: jrode@uci.edu

classroom, team-based learning, and weekly R labs and lab-based assignments. Professors of courses without lab components, and/or courses in which students have high levels of statistics anxiety and diverse mathematical and computational backgrounds, may be left wondering whether it is worthwhile to introduce students to R over other software types. Indeed, ongoing debates in online education communities suggest that the use of R with undergraduates, and how the experience compares to teaching software such as SPSS, is very much an open question that many educators would like to see answered empirically (e.g., see https://www.researchgate.net/post/Is_it_easier_for_students_to_learn_statistics_using_SPSS_or_R). Statistics professors have likewise written blogs about the benefits and drawbacks of R and SPSS (e.g., Anglim, 2013; Franklin, 2018; Wall, 2014). These debates capture the concern that R is a highly useful program for students but comes with a steeper learning curve and fewer resources available for beginners compared to other software, leading instructors to question whether it is wise to emphasize R in an introductory course (especially those for non-statistics majors). To the best of our knowledge, no study has explicitly compared the teaching of R to statistical software more commonly used with undergraduates, such as SPSS. Moreover, there is little research on incorporating statistical output in the introductory classroom, much less whether one type of output is more advantageous than another.

The Present Research

The present research is a step toward addressing the potential benefits and drawbacks of using R by examining whether R or SPSS statistical output induces greater initial anxiety in students, and whether anxiety toward one or both change after being taught one type of software output throughout the course. The study was conducted on two undergraduate introductory statistics courses taught back-to-back in the same summer and university. The university's curriculum does not include a lab-based component in the introductory course, leaving it up to the instructor to decide if and how to incorporate statistical software within the lecture-based structure of the course. Instead, students are typically exposed to statistical output in the form of tables and graphs from software such as SPSS (or others depending on the instructor's preferred software). In our design, one course exposed students to R output, whereas the other course used SPSS output. In most other ways, however, the courses were nearly identical. Both instructors were doctoral students with the same amount of pedagogical training and experience. Both courses were of a similar size and composition, and students in both classes had equivalent levels of background experience in statistics and performance on a pretest of basic math and research design knowledge. Furthermore, the same topics and instructional techniques were used in both courses, and the instructors presented statistical output to students frequently and approximately the same number of times.

The instructor of the first course taught students to interpret R output only, whereas the instructor of the second course taught SPSS output only. Throughout the rest of the article, we will refer to these as the R and the SPSS Courses, respectively. Our outcomes were self-reported anxiety and self-confidence toward interpreting (or learning to interpret) statistical output. These were measured using surveys in which students were given examples of statistical output in R and SPSS and then responded to items probing for anxiety and confidence. Surveys were administered on the first and last days of each course.

With this design, we were able to test several hypotheses. First, at the beginning of each course, we hypothesized that students would report greater anxiety when viewing R output compared to identical content presented in SPSS output. Although there is little empirical research to speak directly to this hypothesis, both our anecdotal experiences as instructors and others' teaching experiences (e.g., Anglim, 2013) suggest that many aspects of R, including its output structure, are more difficult initially for novices to learn compared to other software. Unlike the formatted tables that SPSS produces, R output more closely resembles lines of code, with information more tightly clustered together and labels often misaligned with their corresponding values. Given these differences, it may be that R output has a less visually appealing, intuitive design to the novice learner. Output that resembles code may also be problematic for students with greater "programming anxiety," as research suggests that this is a common anxiety among undergraduates learning to code (Connolly, Murphy, & Moore, 2009; Nolan & Bergin, 2016; Owolabi, Olanipekun, & Iwerima, 2014). Thus, it was logical to predict that students would initially report greater anxiety and lower confidence toward interpreting R compared to SPSS output.

Our second hypothesis was that students would report lower anxiety and greater self-confidence at the end of the course (compared to the beginning of the course) toward the software which they were taught. Several studies have found that students improve their attitudes toward statistics and gain greater knowledge of statistics after having been taught SPSS software (e.g., McCullough, 2017; Šebjan & Tominc, 2015). Other studies report similar success in teaching students R software (e.g., Baumer et al., 2014; Farrell & Carey, 2018). Likewise, we expected that by teaching our students how to interpret output and prioritizing this ability throughout the course, students would become proficient in interpreting their course's output (whether R or SPSS) and thus would feel less anxious and more confident about that type of output by the end of the course.

Third, we predicted that learning how to interpret one type of software output would translate to less anxiety and higher self-confidence toward the other type of output. We expected this outcome because gaining experience reading one type of output would likely make a different type of output less anxiety provoking, given that it is simply the same statistical tests presented in a different format. In other words, we surmised that because students at the end of the course would be familiar with the types of statistical tests shown in the output, an unfamiliar output format would still be processed with greater ease compared to the beginning of the course, when students were likely unfamiliar with all of the content shown. Finally, we explored (but had no specific prediction) whether the initial difference between R- and SPSS-related anxiety (and/or confidence) would persist or disappear after having been taught R or SPSS throughout the course. In other words, although we did expect an improvement in anxiety and confidence over time for the type of output students were taught, we did not have enough prior evidence to specify if the initial gap between R- and SPSS-related anxiety (and/or confidence) would remain by the end of each course. For example, it may be that students in the SPSS Course feel less anxious toward R output at the end of the course but still report greater anxiety toward it than they do toward SPSS.

Method

Participants and Procedure

The current research took place at a large, public West Coast university during the summer in subsequent 5-week introductory statistics courses, from late June 2018 to early September 2018. The introductory course is mandatory for undergraduates in the school and requires research methods to be taken as a prerequisite. Both instructors taught the same course, with one during the first summer session and the other during the second summer session. Students were mostly juniors and seniors, with a few sophomores and no freshmen (see Table 1 for specifics). We did not collect gender or race/ethnicity of the students in our specific courses, but the demographics of students in those courses tend to be similar to those of the university overall. As of 2017, the gender and race/ethnicity breakdown of undergraduates at the university was 53% female, 35% Asian, 26% Hispanic, 14% White, and fewer than 5% identifying as another race/ethnicity.

Interpreting statistical output was listed as an objective for each course, with both instructors emphasizing this task. Students were exposed to R output in the R Course a total of 46 times: 21 through lecture and in-class activities and 25 on assignments and assessments. Students were exposed to SPSS output in the SPSS Course a total of 44 times: 30 through lecture and in-class activities and 14 on assignments and assessments. The courses used the same textbook, which did not include R or SPSS output (Caldwell, 2012). Although the courses were intentionally taught similarly with comparable content, there was no random assignment to course, making it a quasi-experimental study.

An introductory survey was given on the first day of each course before any material was taught, and a similar survey was given during the last class session, after all course content was finished. Both surveys contained the same output examples, and the order of presentation of SPSS and R output was varied to control for order effects. The output contained descriptive statistics (means, standard deviations) and an independent samples *t*-test. Students were not informed that the surveys were part of a study but were told the surveys were a way for their instructor to get to know them better and understand how they feel about statistics. Forty-three students were enrolled in the R Course, and 39 in the SPSS Course, but not every student took both the pre- and posttest survey (i.e., some students were absent on the first or last day of class). Forty-three students took the survey at the beginning of the R Course and 39 at the end of it. Thirty-nine students took the survey at the beginning of the SPSS Course and 38 students at the end of it. The university's institutional review board (IRB) approved the research.

Table 1. Background Information Comparison Between the Two Courses.

	R Course		SPSS Course	
Variable	*M* (*SD*)	*n* (%)	*M* (*SD*)	*n* (%)
Knowledge[a]				
Research methods	2.26 (1.05)	43	2.49 (0.91)	39
Math	3.35 (0.81)	43	3.41 (0.72)	39
Taken statistics		18 (41.86)		18 (46.15)
Major[b]				
Psychology		21 (48.84)		22 (56.41)
Criminology*		21 (48.84)		9 (23.08)
SE[c]		5 (11.63)		5 (12.82)
Other[d]		9 (20.93)		8 (20.51)
Year[e]				
Sophomore		2 (4.76)		4 (10.26)
Junior		23 (54.76)		16 (41.03)
Senior		17 (40.48)		19 (48.72)

Note. To test for background differences between the R and SPSS Courses, independent samples *t* tests were conducted for research methods and math. χ^2 tests were conducted for each categorical variable, except those with five or fewer students in a cell, in which case Fisher's exact test was conducted.
[a]Research methods and math were scored from 0 to 4, with 1 point for each correct answer. [b]Students could select more than one major, so the four listed are not mutually exclusive. The χ^2 test was conducted for each major separately. [c]SE = social ecology. Fisher's exact test was conducted. [d]Fisher's exact test was conducted. [e]There were no freshmen in either course. Fisher's exact test was conducted.
*$p < .05$.

Measures

Anxiety. Because existing scales did not sufficiently tap anxiety toward the output itself and were too cumbersome for the short survey, we measured anxiety with two face valid items on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*): "I do not feel any anxiety when I look at this statistical output" (reverse coded) and "I feel very anxious when I look at this statistical output." These items were identical on the pre- and posttests. We calculated standardized coefficient α for anxiety separately by output and time, resulting in four α calculations for anxiety: pretest R (standardized $\alpha = .84$, $N = 82$, $M = 4.94$, $SD = 1.59$), pretest SPSS (standardized $\alpha = .86$, $N = 82$, $M = 4.41$, $SD = 1.68$), posttest R (standardized $\alpha = .71$, $N = 77$, $M = 2.19$, $SD = 1.47$), and posttest SPSS (standardized $\alpha = .70$, $N = 77$, $M = 1.98$, $SD = 1.30$; for a discussion on the reliability of 2-item scales, see Eisinga, Grotenhuis,

& Pelzer, 2013). The R and SPSS Courses each had 38 students who completed the anxiety measures for both the pretest and posttest.

Confidence. Confidence was measured with four items on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*): "I'll never be able to understand statistical tables like this" (reverse coded), "This statistical output seems like it would be impossible to understand" (reverse coded), "This statistical output seems like it would be easy to understand," and "I feel confident that I'll learn enough statistics to be able to understand statistical output like this." The pre- and posttest items were identical except for the last item, which was changed to the present tense for the end-of-course survey ("...that I know enough statistics to be able..."). We calculated Cronbach's α for confidence separately by output and time, resulting in four α calculations for confidence: pretest R ($\alpha = .79$, $N = 82$, $M = 4.89$, $SD = 1.29$), pretest SPSS ($\alpha = .84$, $N = 81$, $M = 5.21$, $SD = 1.28$), posttest R ($\alpha = .64$, $N = 77$, $M = 6.23$, $SD = 0.87$), and posttest SPSS ($\alpha = .69$, $N = 77$, $M = 6.20$, $SD = 0.86$). The R Course had 37 students and the SPSS Course had 38 students who completed the confidence measures for both the pretest and posttest. See Table A1 in Appendix A for correlations between the measures.

Background knowledge. To understand prior knowledge of students on the pretest survey, we included four multiple-choice questions on research methods and four multiple-choice questions on basic math calculations. The four research methods questions probed concepts like hypotheses and control groups, while the four math questions involved working with fractions, percentages, and algebra. The methods and math questions can be found in Appendix B. We assessed statistics exposure by asking students if they had taken a statistics class before in either high school or college.

Background information. We asked students about their year in school and their major. Year in school was a multiple-choice question with four options: freshman, sophomore, junior, or senior. Major was an open-ended question, so that students could indicate multiple majors. We broke major down into four categories: the three main majors within the school where the courses were taught—psychology, criminology, and social ecology—and an Other category that represented a major other than these three. We separated the four major categories into their own yes/no dichotomous variables, allowing students to be coded as "yes" for multiple majors. This helped us to assess the differences between the two courses by each major.

Statistical Analyses

Analyses were conducted using the R statistical software (R Core Team, 2018). We used anxiety and confidence as our two outcome variables. We conducted separate linear mixed models for both of these outcome variables, with time (pre, post) and output (R, SPSS) as within-participants factors and course (R Course, SPSS Course) as a between-participants factor. All three of these factors were dummy coded with pretest, R, and R Course as the reference groups, respectively. We used restricted maximum likelihood estimation and included a random intercept by-participants and random slopes for time by-participants.[1] Analyses were conducted using the *lmerTest* package in R (Kuznetsova, Brockhoff, & Christensen, 2017), and p-values were calculated using Satterthwaite's degrees of freedom approximation (Luke, 2017; Satterthwaite, 1941).

Results

Course Comparison

The background information for the two courses is shown in Table 1. Independent samples t-tests, χ^2 tests, and Fisher's exact tests were conducted to detect any differences between the two courses, with only criminology major resulting in a significant difference between the courses, $\chi^2(1) = 5.85$, $p = .02$. The two courses appear to be composed of students with similar backgrounds.

Anxiety

To assess students' anxiety toward the different software outputs, we conducted a 2 (time: pre vs. post) $\times$ 2 (output: R vs. SPSS) $\times$ 2 (course: R Course vs. SPSS Course) linear mixed model predicting anxiety from the interaction of our three categorical predictor variables. There was not a significant effect of course ($b = -0.27$, 95% CI_{boot} [-0.96, 0.45], $p = .45$), nor a significant course $\times$ time interaction ($b = 0.66$, 95% CI_{boot} [-0.19, 1.52], $p = .14$), nor a significant course $\times$ output interaction ($b = 0.16$, 95% CI_{boot} [-0.34, 0.63], $p = .54$). But there were significant effects of time ($b = -3.07$, 95% CI_{boot} [-3.66, -2.49], $p < .001$) and output ($b = -0.60$, 95% CI_{boot} [-0.93, -0.25], $p < .001$), such that students felt more anxiety at the beginning of the course (compared to the end) and toward R (compared to SPSS).[2] However, these effects are qualified by a significant time $\times$ output interaction ($b = 0.81$, 95% CI_{boot} [0.31, 1.30], $p = .002$) and a significant three-way interaction between time, output, and course ($b = -0.99$, 95% CI_{boot} [-1.72, -0.29], $p = .01$). As seen in the R Course panel of Figure 1, anxiety decreased significantly more for R than for SPSS (time $\times$ output interaction). The same interaction was not significant when using the SPSS Course as the reference group ($b = -0.18$, 95% CI_{boot} [-0.73, 0.34], $p = .49$), where anxiety about R and SPSS decreased by roughly the same amount from pre to post.

To further probe this three-way interaction, we conducted pairwise comparisons with Tukey-corrected p-values (within session, for two families of six comparisons each).[3] In the R Course, students initially had more anxiety about R ($M = 5.06$) than SPSS, $M = 4.46$, $t(155.00) = 3.43$, $p = .004$. By the end of the R Course, however, anxiety levels were nonsignificantly different between the two programs, $M_R = 1.99$, $M_{SPSS} = 2.19$, $t(155.00) = -1.11$, $p = .68$. In the SPSS Course, similarly, students began with marginally more anxiety about R ($M =$

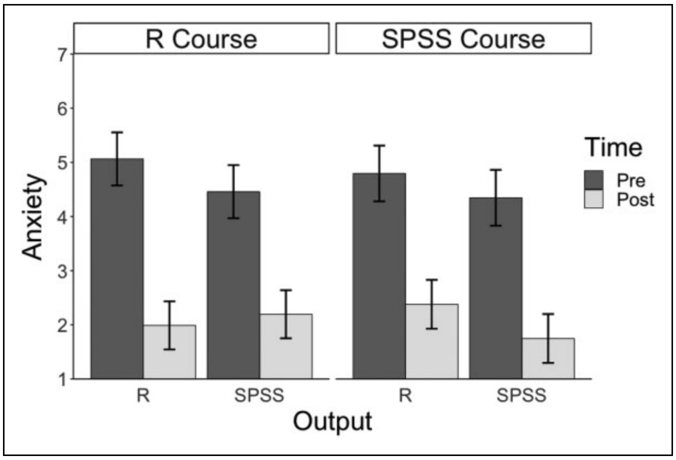

Figure 1. Students' anxiety toward learning statistical output reported at the beginning and end of the R and SPSS Courses. Error bars are 95% confidence intervals.

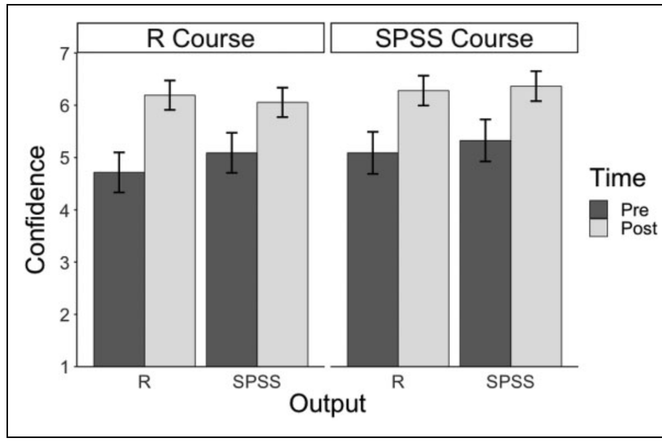

Figure 2. Students' confidence in learning statistical output reported at the beginning and end of the R and SPSS Courses. Error bars are 95% confidence intervals.

4.79) than SPSS, $M = 4.35$, $t(155.00) = 2.43$, $p = .08$. But at the end of the SPSS Course, students still had significantly more anxiety about R ($M = 2.38$) than they did about SPSS, $M = 1.75$, $t(155.00) = 3.37$, $p = .01$.

Confidence

To examine how confidence differed between time, output, and course, we conducted the same linear mixed model as with anxiety but used confidence as the outcome variable. There was not a significant effect of course ($b = 0.37$, 95% $\mathrm{CI}_{\mathrm{boot}}$ [–0.16, 0.90], $p = .18$), nor a significant course × time interaction ($b = –0.29$, 95% $\mathrm{CI}_{\mathrm{boot}}$ [–0.85, 0.25], $p = .33$), nor a significant course × output interaction ($b = –0.14$, 95% $\mathrm{CI}_{\mathrm{boot}}$ [–0.47, 0.21], $p = .40$), nor a significant three-way interaction ($b = 0.36$, 95% $\mathrm{CI}_{\mathrm{boot}}$ [–0.06, 0.82], $p = .12$).

However, as seen in Figure 2, there was a significant effect of time ($b = 1.48$, 95% $\mathrm{CI}_{\mathrm{boot}}$ [1.09, 1.85], $p < .001$) such that students felt more confidence at the end of the course than at the beginning. There was also a significant effect of output, where students felt more confidence overall toward SPSS than R ($b = 0.37$, 95% $\mathrm{CI}_{\mathrm{boot}}$ [0.15, 0.60], $p < .001$). Furthermore, these effects are qualified by a significant interaction between time and output ($b = –0.51$, 95% $\mathrm{CI}_{\mathrm{boot}}$ [−0.82, –0.22], $p = .002$). As seen in Figure 2, although students' confidence in both R and SPSS increased over time, there was a larger increase in confidence for R than SPSS.

Due to the nonsignificant three-way interaction, we collapsed across session to probe the time × output interaction using pairwise comparisons, adjusting for multiple comparisons using Tukey-corrected *p*-values (family of six comparisons after collapsing across session). At the beginning of the course, students had significantly more confidence in interpreting output from SPSS ($M = 5.21$) than R, $M = 4.90$, $t(154.46) = −3.81$, $p = .001$. By the end of the course, however, these differences were largely eliminated and students felt just as much confidence in interpreting R ($M = 6.24$) as they did in SPSS, $M = 6.21$, $t(154.17) = 0.32$, $p = .99$.

Discussion

As hypothesized, students in both courses initially reported greater anxiety and lower confidence when viewing R output compared to SPSS output. However, after exposure to either R or SPSS throughout the course, students indicated significantly lower anxiety and higher self-confidence toward interpreting output from the software they were taught as well as toward the other type of output. In the R Course, anxiety toward R and SPSS decreased substantially and were nearly identical at the end of the course. In the SPSS Course, anxiety toward R and SPSS likewise decreased substantially, though R anxiety remained statistically significantly higher than SPSS anxiety. In both courses, students reported feeling less confident about learning R output but ended the course equally confident in their ability to interpret R and SPSS output. In sum, the initial difference between R- and SPSS-related anxiety and confidence level disappeared when students were taught R and decreased (though remained significant for anxiety) when students were taught SPSS. These results suggest that learning how to interpret one type of software output decreases anxiety and elevates confidence toward it, while also improving attitudes toward a new type of output.

Anecdotally, many of our students in both courses expressed positive attitudes toward learning how to read output and told us this was one of the most interesting parts of the class to them. This was also evident on the end-of-course surveys, in which 20.5% of students in the R Course and 21% of students in the SPSS Course spontaneously mentioned that learning to read statistical output was their favorite part of the course. Both instructors emphasized that they were foregoing an emphasis on hand calculations in favor of teaching output, in order to expose students to how researchers use statistics in the real world. Both instructors also generated output from real data and presented them in the context of different research scenarios to emphasize the connection between research design and corresponding statistical tests. The instructors' positive attitudes toward output and connection of it to real-world applications may have helped students gain more confidence and feel less anxiety toward it. Although we do not have data to speak to this, we would encourage future instructors to use real data and communicate to their students the purpose and benefits of learning to interpret statistical output. Past research likewise supports the use of real-world data and statistical software to improve a wide range of student outcomes (Ciftci et al., 2014; McCulloch, 2017; Neumann, Hood, & Neumann, 2013).

Limitations and Future Directions

The results are encouraging for any educator wishing to incorporate R in an introductory statistics course, but the study also had multiple limitations to consider. First, the study was limited to one public university and conducted during the summer sessions rather than during the regular academic year. Additionally, the class sizes were relatively small. Given these limitations, it would be helpful to replicate the study in other institutions and with more courses (or larger class sizes) before drawing firm conclusions. Larger samples would also allow for more fine-grained analyses for which the present research was too underpowered to attempt, such as exploring whether statistics anxiety or demographic characteristics moderate the effects of output exposure.

Another limitation of the study is a lack of a validated scale of anxiety. We created two face valid measures for the purposes of this study, but they have not been validated in other samples. Other researchers have created and validated scales of anxiety toward statistics (e.g., Cruise, Cash, & Bolton, 1985; Vigil-Colet, Lorenzo-Seva, & Condon, 2008). However, these scales measure anxiety toward statistics in general and do not tap into anxiety about specific software output. Because we wanted to examine reactions to different software output, we needed the items to be particular to software output itself rather than a general anxiety toward statistics. We hoped to capture initial student reaction to the specific software output and wanted items to point directly to that output. On a practical level, we also wanted to keep the survey short in order to resemble a more natural introductory survey and prevent suspicion from students. Despite our reasons for choosing to create our own measures, the lack of psychometric testing results in uncertainty about the robustness of the items. The items could be measuring a slightly different construct than what we intended, may not generalize to other samples, or may not correlate with other measures of anxiety. The items have not undergone thorough testing by other researchers nor been revised after rigorous assessment. Future research should investigate the validity and reliability of measures specifically focused on statistical software. The field would also benefit from creating new measures that capture attitudes toward different aspects of software use and different types of software programs, including R.

There was also no random assignment of students to the R or SPSS Course. As the study was not a true experiment, we cannot rule out the possibility that additional, unmeasured student characteristics influenced the outcome. Moreover, each course was taught by a different instructor rather than having the same instructor for both courses, which could have introduced differences that may have affected the results in a particular direction. Although we attempted to make each course comparable in design, content, instructor style, and frequency of exposure to output, unmeasured differences between the students and/or the instructors' unconscious biases could have influenced students' responding on the course surveys. For example, a more rigorous experiment that includes random assignment of students to study condition as well as instructors blind to students' condition would help prevent demand characteristics from influencing students' survey responses.

Perhaps the biggest limitation is that the students learned only to interpret output rather than how to actually use the software. The courses in the present study were lecture-based and thus could not teach students how to analyze data in R and SPSS. The present results may not generalize to situations in which the students are expected to use the software along with interpreting its output. Additional research is needed before instructors can assume that student outcomes will be equivalent in courses that teach students to actually use R software compared to SPSS.

Given the aforementioned limitations, more work is needed to better understand how students' attitudes toward statistics, as well as understanding and academic performance, fare when learning R compared to a more commonly used program such as SPSS. However, the promising results we obtained comport with previous research suggesting that students in an introductory course benefit from learning R (Baumer et al., 2014). Future research should be conducted at a variety of institutions (community colleges, public and private institutions), explore any potential moderating characteristics of the students themselves (e.g., prior anxiety levels, mathematical and computer programming background), and compare outcomes from classes taught with different statistical programs (e.g., R, SPSS, and SAS). Experimental work in particular would be most valuable to give instructors precise information on the benefits and drawbacks of teaching students how to conduct data analysis in R compared to other software types. There is currently a dearth of research on incorporating R in psychology statistics courses compared to research on SPSS. Instructors would no doubt be interested in better understanding whether R presents greater barriers initially compared to other software types, and what the consequences of those barriers may be in terms of student retention and achievement. Studies should also investigate what types of support (instructional, textbook, etc.) students need to succeed with R, and whether student attitudes and understanding of statistics are enhanced by the utilization of R. Future research would also benefit from examining longer term benefits of learning different types of software, such as what software experience is more prevalent or sought after in the types of industry careers that undergraduate psychology majors often enter and what software skills are desired in psychology graduate programs.

Conclusion

The present results suggest that in courses that incorporate the goal of teaching students to interpret statistical output, students' anxiety and confidence levels can improve just as much when learning R compared to SPSS, even if R appears more intimidating at the beginning. Not only did we observe no drawbacks to teaching R but also it appears that teaching any type of output may decrease student anxiety and increase confidence toward learning a new type of output. Future studies should investigate whether these results hold in other contexts, and more importantly, how psychology undergraduates fare in courses that teach students to use R compared to other types of software.

Appendix A

Table A1. Correlations Between Confidence and Anxiety Scales Separated by Time and Output.

	Pre/R		Pre/SPSS		Post/R		Post/SPSS	
Variable	Confidence	Anxiety	Confidence	Anxiety	Confidence	Anxiety	Confidence	Anxiety
Pre/R								
Confidence	1							
Anxiety	−.57***	1						
Pre/SPSS								
Confidence	.79***	−.49***	1					
Anxiety	−.42***	.71***	−.59***	1				
Post/R								
Confidence	.31**	−.14	.37**	−.11	1			
Anxiety	−.21	.16	−.23*	.17	−.56***	1		
Post/SPSS								
Confidence	.34**	−.17	.33**	−.10	.78***	−.44***	1	
Anxiety	−.17	.13	−.22	.18	−.46***	.68***	−.54***	1

*$p < .05$. **$p < .01$. ***$p < .001$.

Appendix B

Math and research methods prior knowledge questions (correct answer is indicated with *):

1) A midterm exam is worth 50 points. What percentage of your grade is the exam if the class is out of 300 points?
   a) 15%
   b) 16.67%*
   c) 18%
   d) 21.67%

2) A company reports that two fifths of its employees are women. If there are 90 employees, how many are women?
   a) 36*
   b) 40
   c) 44
   d) 50

3) Convert 12/40 to a decimal:
   a) .37
   b) .33
   c) 3.33
   d) .30*

4) $3X + 5 = -4$. What does X equal?
   a) 3
   b) −3*
   c) .33
   d) −.33

5) What is a hypothesis?
   a) an observation about phenomena
   b) a question about the relationships between variables
   c) a statement about the relationship between two (or more) variables*
   d) a theoretical construct that involves two (or more) variables

6) In a within-subjects design . . .
   a) only the experimental group gets the treatment
   b) individuals are observed but no variables are manipulated
   c) the sample is a single group that participates in every treatment condition*
   d) the sample is divided into multiple groups that only get one type of treatment

7) The independent variable is the variable that is ___________ by the researcher.
   a) manipulated*
   b) measured
   c) observed
   d) invented

8) What is the purpose of a control condition in an experiment?
   a) to control for the effects of the experiment
   b) to provide a baseline for comparison with the experimental condition*
   c) to have a backup group in case something goes wrong

Authors' Note

JBR generated the idea, conducted analyses, and wrote the Methods and Results sections. MMR wrote the Abstract, Introduction, and Discussion sections.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Because our design had only one observation per participant per cell (i.e., each student had one outcome variable measurement for R pretest, one R posttest, one SPSS pretest, and one SPSS posttest), we could not estimate the maximal random effects structure, which would have included the interaction between output and time as a random effect (Barr, Levy, Scheepers, & Tily, 2013). In situations like this, Brauer and Curtin (2018) recommend still running the maximal model by suppressing the error given by the software program. However, Barr (2013) and Singmann and Kellen (in press) state that researchers can simply drop the highest order interaction term as a random effect and treat the resulting model as the maximal one (also see Singmann, 2017). Using a data-driven approach to compare models (as recommended by Barr et al., 2013), we compared five models with different random effects structures: (1) the maximal model with software errors suppressed; (2) a model with a random intercept by-participants, random slopes for time by-participants, and random slopes for output by-participants; (3) a model with a random intercept by-participants and random slopes for time by-participants; (4) a model with a random intercept by-participants and random slopes for output by-participants; and (5) a model with only a random intercept by-participants. The simplest model that was not significantly different than a more complex model was the third model, one with a random intercept by-participants and random slopes for time by-participants.
2. Since we used dummy coding, the non-interaction effects reported in this article are simple effects: The effect of time is for the R Course with R output, the effect of output is for the R Course at the pretest, and the effect of course is at the pretest with R output.
3. We conducted pairwise comparisons using the *emmeans* package in R (Lenth, 2018). The means reported are the estimated marginal means from the mixed model, and we continued to use Satterthwaite's approximation for the degrees of freedom. All R code available upon request.

References

Allaire, Y. X., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., . . . Chang, W. (2018). *rmarkdown: Dynamic documents for R*. R package version 1.10. Retrieved from https://CRAN.R-project.org/package=rmarkdown

Anglim, J. (2013, July 17). *Evaluating the potential incorporation of R into research methods education in psychology* [Blog post]. Retrieved from http://jeromyanglim.blogspot.com/2013/07/evaluating-potential-incorporation-of-r.html

Bartz, A. E., & Sabolik, M. A. (2001). Computer and software use in teaching the beginning statistics course. *Teaching of Psychology*, *28*, 147–149. doi:10.1207/S15328023TOP2802_16

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, Article 328. doi:10.3389/fpsyg.2013.00328

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. doi:10.1016/j.jml.2012.11.001

Baumer, B., Cetinkaya-Rundel, M., Bray, A., Loi, L., & Horton, N. J. (2014). R Markdown: Integrating a reproducible analysis tool into introductory statistics. *Technology Innovations in Statistics Education*, *8*. doi:10.5811/westjem.2011.5.6700

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*, 389–411. doi:10.1037/met0000159

Caldwell, S. (2012). *Statistics unplugged* (4th ed.). Belmont, CA: Cengage Learning.

Ciftci, S. K., Karadag, E., & Akdal, P. (2014). Instruction of statistics via computer-based tools: Effects on statistics' anxiety, attitude, and achievement. *Journal of Educational Computing Research*, *50*, 119–133. doi:10.2190/EC.50.1.f

Connolly, C., Murphy, E., & Moore, S. (2009). Programming anxiety amongst computing students—A key in the retention debate? *IEEE Transactions on Education*, *52*, 52–56.

Cruise, R. J., Cash, R. W., & Bolton, D. L. (1985, August). *Development and validation of an instrument to measure statistical anxiety*. Paper presented at the Annual Meeting of the Statistical Education Section, Chicago, IL.

Eisinga, R., Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*, 637–642. doi:10.1007/s00038-012-0416-3

Farrell, K. J., & Carey, C. C. (2018). Power, pitfalls, and potential for integrating computational literacy into undergraduate ecology courses. *International Journal of Business Innovation and Research*, *17*, 7744–7751. doi:10.1002/ece3.4363

Ferreira, R. D.S., Kataoka, V. Y., & Karrer, M. (2014). Teaching probability with the support of the R statistical software. *Statistics Education Research Journal*, *13*, 132–147.

Franklin, R. (2018, June 21). *R—The future of psychology statistics is open source* [Blog post]. Retrieved from https://nobaproject.com/blog/2018-06-21-r-the-future-of-psychology-statistics-is-open-source

GAISE College Report ASA Revision Committee. (2016). *Guidelines for assessment and instruction in statistics education college report 2016*. Retrieved from http://www.amstat.org/education/gaise

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. doi:10.18637/jss.v082.i13

Lenth, R. (2018). *emmeans: Estimated marginal means, aka least-squares means*. R package version 1.2.2. Retrieved from https://CRAN.R-project.org/package=emmeans

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*, 1494–1502. doi:10.3758/s13428-016-0809-y

McCulloch, R. S. (2017). Learning outcomes in a laboratory environment vs. classroom for statistics instruction: An alternative approach using statistical software. *International Journal of Higher Education*, *6*, 131–142. doi:10.5430/ijhe.v6n5p131

Neumann, D. L., Hood, M., & Neumann, M. M. (2013). Using real-life data when teaching statistics: Student perceptions of this strategy in an introductory statistics course. *Statistics Education Research Journal*, *12*, 59–70.

Nolan, K., & Bergin, S. (2016). The role of anxiety when learning to program. In *Proceedings of the 16th Koli Calling International Conference on Computing Education Research* (pp. 61–70). ACM. doi:10.1145/2999541.2999557

Onwuegbuzie, A. J., & Wilson, V. A. (2003). Statistics anxiety: Nature, etiology, antecedents, effects, and treatments—A comprehensive review of the literature. *Teaching in Higher Education*, *8*, 195–209. doi:10.1080/1356251032000052447

Owolabi, J., Olanipekun, P., & Iwerima, J. (2014). Mathematics ability and anxiety, computer and programming anxieties, age and gender as determinants of achievement in basic programming. *GSTF Journal on Computing (JoC)*, *3*, 109–114. doi:10.5176/2251-3043_3.4.296

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika*, *6*, 309–316. doi:10.1007/BF02288586

Šebjan, U., & Tominc, P. (2015). Impact of support of teacher and compatibility with needs of study on usefulness of SPSS by students. *Computers in Human Behavior*, *53*, 354–365. doi:10.1016/j.chb.2015.07.022

Singmann, H. (2017, May 29). *Mixed models for ANOVA designs with one observation per unit of observation and cell of the design* [Blog post]. Retrieved from http://singmann.org/mixed-models-for-anova-designs-with-one-observation-per-unit-of-observation-and-cell-of-the-design/

Singmann, H., & Kellen, D. (in press). An introduction to mixed models for experimental psychology. In D. H. Spieler & E. Schumacher (Eds.), *New methods in neuroscience and cognitive psychology*. Psychology Press. Retrieved from https://cran.r-project.org/web/packages/afex/vignettes/introduction-mixed-models.pdf

Sosa, G. W., Berger, D. E., Saw, A. T., & Mary, J. C. (2011). Effectiveness of computer-assisted instruction in statistics: A meta-analysis. *Review of Educational Research*, *81*, 97–128. doi:10.3102/0034654310378174

Vigil-Colet, A., Lorenzo-Seva, U., & Condon, L. (2008). Development and validation of the statistical anxiety scale. *Psicothema*, *20*, 174–180.

Wall, M. (2014, November 27). *JASP might finally be the SPSS replacement we've been waiting for* [Blog post]. Retrieved from https://computingforpsychologists.wordpress.com/tag/spss/

Yu, F. (2017). Discussion on the application of R software in statistics teaching. *2017 5th International Conference on Modern Education (ICOME 2017) (Icome)* (pp. 47–53). Retrieved from http://www.dpi-proceedings.com/index.php/dtssehs/article/viewFile/19347/18838

Zeidner, M. (1991). Statistics and mathematics anxiety in social science students: Some interesting parallels. *The British Journal of Educational Psychology*, *61*, 319–328. doi:10.1111/j.2044-8279.1991.tb00989.x