Research Ethics Training Materials:

Questionable Research Practices in the Quantitative Humanities

Margaret Wood, Northern Arizona University
Tove Larsson, Uppsala University / Northern Arizona University
Luke Plonsky, Northern Arizona University
Scott Sterling, Indiana State University
Merja Kytö, Uppsala University
Katherine Yaw, University of South Florida

Please cite as:
Research Ethics Training Materials: Questionable Research Practices in the Quantitative Humanities

Margaret Wood, Northern Arizona University
Tove Larsson, Uppsala University / Northern Arizona University
Luke Plonsky, Northern Arizona University
Scott Sterling, Indiana State University
Merja Kytö, Uppsala University
Katherine Yaw, University of South Florida

2023

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

This document is produced under a Creative Commons BY-SA NonCommercial license, which means that you can use, distribute, and/or remix the material so long as you do not charge money for the resulting work. You must provide appropriate credit for any material you use from this text, including the use of a creative commons license equal to the one included on any contributions that you distribute. More information about Creative Commons can be found at https://creativecommons.org/. We particular encourage translation of these stories into additional languages.
TABLE OF CONTENTS

1. Introduction .................................................................................................................................................4

2. QRP Taxonomy – Descriptions and Video Links .........................................................................................5
   2.1 Section I: Funding..................................................................................................................................5
   2.2 Section II: Design and Data Collection ...............................................................................................9
   2.3 Section III: Data Analysis and Interpretation ......................................................................................13
   2.4 Section IV: Write-up and Dissemination .............................................................................................18

3. Lesson Plans ...............................................................................................................................................27
   3.1 Directions .............................................................................................................................................27
   3.2 Sample Lesson Plans ............................................................................................................................27
   3.3 Activity Bank .......................................................................................................................................30
   3.4 Scenarios / Discussion Questions / Extra Readings .........................................................................31
   3.5 Suggested Readings .............................................................................................................................39

4. Authors’ Contact ........................................................................................................................................41
Introduction

Questionable Research Practices (QRPs) fall in the gray area between responsible conduct of research (RCR) and blatant misconduct (i.e., falsification, fabrication, plagiarism) (Isbell et al., 2022; Larsson, et. al., submitted; Sterling & Gass, 2017). While there is generally consensus that certain research practices, such as plagiarism, qualify as research misconduct, researchers do not always agree on the appropriateness of practices that fall into the ethical gray zone, for example, a researchers’ responsibility to share data and the proper handling of outliers (Larsson et al., submitted). Further complicating the matter is that these practices are sometimes justifiable and reasonable under certain circumstances, and debatable under others (Plonsky et. al., submitted). Because these practices are not particularly well understood or addressed in research ethics training, researchers may be making ethically questionable decisions without knowing they are doing so.

While there is growing interest in research ethics, and overall, more attention given to these ethical gray-zone practices (Yaw et al., forthcoming), it appears that a majority of the training materials available to the field (i.e., textbooks, research methods level courses) focus research ethics training on IRB-related matters, such as informed consent, confidentiality, and treatment of participants. When the gray area topics related to QRPs are breached in these training materials, they are rarely framed as decision-points that are matters of ethical concern (Wood et. al., submitted).

These teaching materials were created with the goal of providing both researchers and students with more robust and thorough research ethics training. This document contains the following materials:

1. A taxonomy of 58 QRPs (see Plonsky et al., submitted) arising at four stages of research. Each QRP is presented with a description and short informational video
2. Three sample lesson plans for lessons of different lengths
3. A list of suggested readings
4. An activity bank for tailoring lesson plans to particular classes
5. A list of scenarios and/or extra readings for select QRPs
QRP Taxonomy – Descriptions and Video Links

2.1 Section I: Funding

1. Cherry-picking samples/data/results that favor the funder
   
   **Description:** Cherry-picking is when a person purposefully picks evidence that supports their claim while omitting data that would potentially nullify their argument. Cherry-picked data will often seem to validate the claim made by an author (they would be bad cherry-pickers otherwise). When it comes to project funders, they might only provide limited access to a dataset or restrict the type of data that could be discussed. Scholars might also only point to positive data to keep their funders happy and willing to provide them with more money. Science and society are the losers here since low quality, unrepresentative data is being published skewing our understanding of how a phenomenon works.

   **Video:** [https://www.youtube.com/watch?v=KHqQV8sWvKQ](https://www.youtube.com/watch?v=KHqQV8sWvKQ)

2. Choosing a topic on the grounds that the funder might expect the study to portray them in a positive light

   **Description:** Imagine the publisher of a prominent test or of a textbook series put out a call for grant proposals that involve their products. This is a fairly common situation, actually. Upon seeing the call, a researcher might be tempted to submit a proposal that would likely cast the publisher/company in a positive light. At first glance, this might seem objectively wrong. However, there is a fair bit of gray area here, too. The findings of the study might also portray the company positively simply because of the nature of their products. The key here is to let theoretical and/or practical issues from the literature drive the questions that we ask rather than how we think the findings of our proposed studies might present the funder.

   **Video:** [https://www.youtube.com/watch?v=rw401UJYZxM](https://www.youtube.com/watch?v=rw401UJYZxM)

3. Not reporting a conflict of interest (financial or otherwise)

   **Description:** It's often the case that we have personal and/or professional links to a particular strand of research or even to a commercial product such as a test that generates revenue. These connections can introduce the potential for bias in our research and, in many cases, cannot always be avoided. The important thing is to simply declare any potential conflicts of interest in a grant proposal. That way the reviewers and grant funding agency is aware in advance of funding the project.

   **Video:** [https://www.youtube.com/watch?v=KagzmiSPmQs](https://www.youtube.com/watch?v=KagzmiSPmQs)
4. Misrepresenting researcher qualification/experience in the proposal

**Description:** Grant proposals almost always require a section that describes the qualifications of the team involved. This might include their areas of expertise, publication record, and experience with other funded projects. This information is then used, along with the proposal to evaluate the likely success of the project. In order to provide the grant-funding agency with an accurate view of the team and its potential, it is imperative to describe such qualifications honestly and without embellishment.

**Video:** [https://www.youtube.com/watch?v=83yDs27GXMc](https://www.youtube.com/watch?v=83yDs27GXMc)

5. Misrepresenting study importance in the proposal (e.g., exaggeration of impact and value of proposal to society)

**Description:** Many grant applications will require some kind of impact statement, which asks the author to describe what will be gained from the project in terms of knowledge, outputs, products, and so forth. Some researchers exaggerate in this section, which is not appropriate. In order to avoid stepping into QRP territory, it's critical that we take a reasoned, balanced view on the potential contribution of the studies we seek funding for.

**Video:** [https://www.youtube.com/watch?v=7yKKLx8GJJM](https://www.youtube.com/watch?v=7yKKLx8GJJM)

6. Over budgeting (i.e., knowingly asking for more money than is actually needed)

**Description:** This QRP should be self-explanatory. The general idea, though, is that we should only request funds that we actually expect to need and use. It can be tempting to pad the budget a bit in order to allow for funding for other projects or, for example, for equipment that isn't strictly needed for the project at hand. This is not appropriate. Furthermore, over-budgeting can also lead to problems related to getting approval to use unspent funds for purposes other than those identified in the proposal.

**Video:** [https://www.youtube.com/watch?v=lzIkiQ41vAY](https://www.youtube.com/watch?v=lzIkiQ41vAY)
7. Not producing the promised project outcomes due to project mismanagement (e.g., producing fewer articles than promised)

**Description:** We of course want to write a grant proposal that is as competitive as possible to the funding agency. An important part of writing a successful proposal is assuring the funder that the money will be put to good use, that is, that the study will yield results that are shared appropriately and widely. While not producing the promised project outcomes could be due to either over-promising or under-performing, it can be considered a 'breach of contract' with the funding agency, and thus a QRP.

**Video:** [https://www.youtube.com/watch?v=j9KziSzkoo](https://www.youtube.com/watch?v=j9KziSzkoo)

8. Using funds for a purpose other than what was stated in the proposal (e.g., for a research assistant instead of for participant-related expenses)

**Description:** Grant proposals almost always include a budget. This budget is itemized to show exactly how all the requested funds will be used. For a budget of $8,000, for example, you might allocate $5000 to paying an hourly research assistant, $2000 for compensating participants, and $1000 for purchasing new software you need for the project. It is generally not appropriate to re-allocate these funds for other purposes without approval from the funder. An example of this would be deciding to use some of the money from the $8000 budget to cover conference-related expenses, which were not part of the budget that the funder approved.

**Video:** [https://www.youtube.com/watch?v=J_1JS9mUT2E](https://www.youtube.com/watch?v=J_1JS9mUT2E)

9. Misrepresenting literature in the proposal (e.g., over-emphasizing previous research that supports the proposal and/or ignoring conflicting evidence)

**Description:** It often happens that researchers find themselves looking for a reference to support a point they want to make, rather than (more appropriately) basing one's ideas on the theoretical or empirical evidence in the literature. Along these same lines, researchers often highlight the studies that support or point to the need for the research they're hoping to get funding for. This problem may be especially prevalent in the absence of meta-analytic evidence to show overall findings across a given body of work. A related but somewhat opposite problem is when a researcher finds out about a study that seems to do more or less what they're seeking funding to do. Might you feel drawn to either omit or at least to de-emphasize that paper from your literature review?

**Video:** [https://www.youtube.com/watch?v=IDCfJ5QPQXE](https://www.youtube.com/watch?v=IDCfJ5QPQXE)
10. Making changes in context after proposal submitted (e.g., proposing to carry out funded work in one context/with one population but then actually carrying it out elsewhere/with another population)

**Description:** Any and all major changes made the design of a funded study need to be reported to and cleared with a funder. For example, if you got funding to conduct a study of classroom-based learners of English as a second language, but you had to then change the sample to a smaller set of online learners of English as a foreign language, you would need to ensure that the funder approved this. Very minor changes are sometimes acceptable without approval. As a general rule, though, it's better to check than to assume.

**Video:** [https://www.youtube.com/watch?v=z_KLXFqLuu8](https://www.youtube.com/watch?v=z_KLXFqLuu8)

11. Not disclosing impacts that funder directly had on research decisions (e.g., using particular datasets, selection of published outcomes)

**Description:** Grant-fundings agencies can sometimes influence a study by encouraging the use of certain instruments or by requesting that the authors publish some of the results instead of others. Sometimes funders even write into grant contracts that they have the right to do so. Such influence is not necessarily inappropriate but it is necessary to declare when this happens in a written report.

**Video:** [https://www.youtube.com/watch?v=kohbD03cyLA](https://www.youtube.com/watch?v=kohbD03cyLA)
2.2 Section II: Design and Data Collection

1. Selecting variables out of convenience and/or familiarity when more theoretically grounded variables are available

**Description:** Some researchers seem to include variables in their designs not because of any strong theoretical or practical reason but, rather, because those variables are familiar or convenient to collect. I've been seeing this lately in SLA with working memory, for example. A more appropriate approach is to think through our variables more thoroughly and to ensure that we're including them in our design because of a hypothesized relation. It's also OK to include exploratory variables in a design but these need to be clearly labeled as such.

**Video:** [https://www.youtube.com/watch?v=eFBnUlwDMmw](https://www.youtube.com/watch?v=eFBnUlwDMmw)

2. Choosing a design and/or instrument type that provides comparatively easy or convenient access to data instead of one that has a strong validity argument behind it

**Description:** When we decide on a study design, including an instrument type, we want to make sure that we measure what we set out to measure, that is, that we have high validity. It is likely more time-consuming and otherwise costly to prioritize validity over convenience. That is, we're likely tempted to go with an instrument type that provides easy or convenient access to data - or we may even have to for different reasons. So as a field, we'd want to think about when (if ever) it would be more all right to sacrifice validity to some degree in the interest of completing a study.

**Video:** [https://www.youtube.com/watch?v=vOR9GSetjJU](https://www.youtube.com/watch?v=vOR9GSetjJU)

3. Defaulting to convention (e.g., choosing a design or instrument type because it is used in previous research, without making sure that it is the most appropriate design or instrument for the target relationships and/or constructs)

**Description:** As creatures (and researchers) of habit, we often choose research designs, instruments, analyses and so forth simply because they are familiar, and not because they necessarily provide the strongest evidence or because they have a strong validity argument behind them, for example. This is a pattern that can lead to inefficiencies, noisy data, and flaws in our results. When setting up a study, it's imperative to question whether we are choosing all aspects of our design and analysis due to the high validity evidence they provide as opposed to defaulting to convention or convenience.

**Video:** [https://www.youtube.com/watch?v=XizoOcdqacA](https://www.youtube.com/watch?v=XizoOcdqacA)
4. Employing instruments/measures without a strong validity argument

**Description:** We often have many choices about which measures we use for a particular variable. In deciding which measure to use, researchers fail to consider whether or not the different choices have a strong validity argument behind them, relying more on other factors such as convenience or simply whether others in the domain have used them. Here's a tip: If you find that measures in your area do not have strong validity arguments behind them, try doing some of the validation work before moving ahead with more studies on unvalidated measures.

**Video:** [https://www.youtube.com/watch?v=DGxzVyuF8-4](https://www.youtube.com/watch?v=DGxzVyuF8-4)

5. Not being transparent with regard to the decisions made in the data collection phase

**Description:** In a typical study, we make a great deal of decisions in the data collection phase. We may for example decide to exclude certain populations from our sampling frame or choose to make changes to our instrument and therefore start over -- or we may expand our sample after a while if we don't get enough data. Decisions made at this stage are important for readers of our study to know about, and failing to report these steps in a transparent manner could be an issue.

**Video:** [https://www.youtube.com/watch?v=kOM7bMTSAyw](https://www.youtube.com/watch?v=kOM7bMTSAyw)

6. Biasing the design/instrument so that outcomes are favorable to researcher beliefs (e.g., choosing a design/instrument that will likely lead to similar outcomes as previous research)

**Description:** Many times we, as researchers, go into running studies with a preference about what we find. We might hope, for example, to find an advantage for one particular treatment over another, or for one variable A to be a stronger predictor of Y than variable B. This is, to some extent, natural and unavoidable. What lands us in the realm of questionable, however, is when these preferences lead us to make design-related decisions that are more likely to yield our preferred outcomes. We need to avoid such choices whenever possible. When we're not sure about our objectivity, we might consider consulting with others who have a different point of view and/or disclosing our preferences or biases going into the study.

**Video:** [https://www.youtube.com/watch?v=P-f8XE71tvQ](https://www.youtube.com/watch?v=P-f8XE71tvQ)
7. Not reporting the effect of decisions about method, design, or instrumentation on study outcomes (e.g., operationalizing proficiency as grade level instead of using an accepted measure of language proficiency)

Description: The research process holds a substantial number of decision points that you will have to make along the way that will affect the outcome of your results. For example, you'll need to make the larger decisions of what type of methodology to select and what type of variables to include, but you'll also need to make smaller decisions that arise unexpectedly throughout the process - for example, how do you handle potential outliers? You may end up having to make these larger decisions based on outside factors that are out of your control. This includes things such as accessibility to the population that you are hoping to collect data from, and journal submission guidelines.

When it comes to reporting our decisions, it’s important to consider which will be the most meaningful and useful to the audience. For example, when in the semester you invited your participants will likely be less meaningful information than who you invited, and why. When making these decisions, it is helpful to consider factors such as audience and journal submission guidelines. Who is the target audience of this paper - practitioners or researchers? What is the word limit for the journal, and how much space will you have to report such decisions?

Importantly, as researchers, it is crucial to have good reasons for your decisions. Whether or not other researchers agree with the decisions that you make, this type of information can help inform future studies and are important for replication research. This also helps the research community have more confidence in what you did.

Video: https://www.youtube.com/watch?v=RfYHvEUDLx4

8. Leaving out known/likely moderator variables or covariates from the study design without explanation or acknowledgment

Description: For a variety of practical reasons, we certainly can't include every variable that matters in every single study. We have to make choices about which variables to include. However, there are sometimes well-known and well-studied covariates or moderating variables that really should be examined and which, when left out, can be considered questionable. An example of this might be a study that looks at reading comprehension only as a function of grammatical knowledge without considering vocabulary knowledge, when both of these variables are known to be highly correlated with reading ability. Including only one of these variables in the study will almost necessarily yield results showing that variable to have an outsized influence on the dependent variable, in this case, reading comprehension.

Video: https://www.youtube.com/watch?v=XlR6MMZrQHQ
9. Fishing for results by collecting information on unnecessary variables

**Description:** We often collect more data than we actually need to address our RQs or that we have a strong theoretical justification for. This is inefficient, but not necessarily unethical or even questionable. We venture into the realm of questionable when we start to analyze data from those variables. For example, it can be tempting to start running correlations for variables not because we hypothesized that they are relevant, but simply because we have the data. This can lead to fishing for results and HARKING (i.e., hypothesizing after results are known).

**Video:** [https://www.youtube.com/watch?v=RuTomZFevN4](https://www.youtube.com/watch?v=RuTomZFevN4)

10. Having an unnecessarily long/burdensome data collection for participants

**Description:** Data collection is time consuming and can even be expensive. We therefore always want to make sure that we get all the relevant information from our study participants. At times, in order for us to get all the relevant information, we would ideally have participants take part in very time consuming and/or mentally taxing activities. While finding a balance between our research needs and the participants' well-being is not always easy, we would want to make sure that we don't take up too much of our participants' time or energy.

**Video:** [https://www.youtube.com/watch?v=xYgoGMNp-NY&t=10s](https://www.youtube.com/watch?v=xYgoGMNp-NY&t=10s)

11. Recruiting participants to join a study in a way that makes refusal difficult or uncomfortable

**Description:** We of course want to try our best to up our chances of getting as many participants as we need for a study. If we fear that we may not reach that number, we may be tempted to try to nudge participants to sign up by making refusal difficult or uncomfortable. We might for example talk about what the consequences might be if we fail to get enough participants (e.g., how students may not be able to graduate). This may well be true and may up our numbers, but where do we draw the line for what recruitment strategies are considered acceptable?

**Video:** [https://www.youtube.com/watch?v=YF11DwNVDVs&t=6s](https://www.youtube.com/watch?v=YF11DwNVDVs&t=6s)
2.3 Section III: Data Analysis and Interpretation

1. Removing whole items/cases knowingly/purposefully to obtain favorable results

**Description:** There are cases when removing items or cases might be justified. For example, items may have proven too difficult to understand for participants and will thus not yield results with high validity, or we may wish to remove cases if we believe that the participant did not take the test/survey seriously. However, we should avoid removing anything from the analysis just to get more favorable results, or results that align with our hypotheses.

**Video:** [https://www.youtube.com/watch?v=oI_IDP1Yelw](https://www.youtube.com/watch?v=oI_IDP1Yelw)

2. Using unjustified methods of handling outliers

**Description:** We have many options for how to handle outlying data points in a given set of data. The same applies to even defining and identifying such data points. Consequently, researchers are often tempted to choose a definition and approach to handling those data points that produces a more favorable outcome. For example, by removing (or not removing) some data points that the researcher considers to be outliers, the analysis may then cross the threshold for statistical significance. Thus, it is imperative to decide how outliers will be identified and dealt with prior to collecting and analyzing a given data set.

**Video:** [https://www.youtube.com/watch?v=ZtjHg_e0too](https://www.youtube.com/watch?v=ZtjHg_e0too)

3. Not being transparent with regard to the reporting on what steps were taken for data cleaning (e.g., removing cases/items without a stated criterion or justification for doing so)

**Description:** Quantitative linguists make a lot of use of language corpora. One’s output files can be enormous, and one often needs to clean the masses of data by applying various screening strategies. It is important to report on these strategies in detail in one’s presentations and publications. One should explain carefully on what grounds examples have been included in the data, and on what grounds they have been excluded from the analyses. Failing to do so would make it difficult for another party to use the results for comparisons. Not being transparent about steps taken in data cleaning can be considered a questionable research practice and something we should not do.

**Video:** [https://www.youtube.com/watch?v=qZ0QZUNP9pE](https://www.youtube.com/watch?v=qZ0QZUNP9pE)
4. Being ambiguous about whether an exploratory vs. confirmatory analysis was employed

**Description:** We use exploratory designs to learn more about a topic that hasn't been researched much. Such studies are hypothesis generating. By contrast, confirmatory studies are hypothesis-testing; that is, we have hypotheses based on theory and previous research that we wish to put to the test on new data, for a new population, or in a new context more broadly. Most confirmatory studies involve some elements of exploration, but we always want to be transparent about what parts of the analysis were and weren't to help future studies.

**Video:** [https://www.youtube.com/watch?v=QpCDBVZKu4E](https://www.youtube.com/watch?v=QpCDBVZKu4E)

5. HARKing (i.e., hypothesizing after results are known)

**Description:** HARKing is hypothesizing after you have already analyzed your data and know your results. You want to avoid this because it can lead to issues where you are not really confirming the data that you found, because you're using the data to make up hypotheses. A stronger way to do research is to make hypotheses, and then do tests to see if those hypotheses are true.

**Video:** [https://www.youtube.com/watch?v=IakIYORcBaQ](https://www.youtube.com/watch?v=IakIYORcBaQ)

6. Cherry-picking data to analyze

**Description:** Cherry picking data to analyze can be more or less problematic in different contexts. For example, in many study designs, we are unable to present all the findings - we have to be selective, and present the most relevant findings. In corpus linguistics studies, we tend to get more results than could ever fit in a research article; in more qualitatively oriented research designs, we similarly have to choose what to focus on. Contexts where it could be more problematic include when we choose to disregard some findings to make the results better line up with our argument or hypotheses.

**Video:** [https://www.youtube.com/watch?v=1V4_54pHS0s](https://www.youtube.com/watch?v=1V4_54pHS0s)

7. Choosing a method of analysis that will likely lead to favorable outcome (e.g., in favor of the researcher’s hypothesis)

**Description:** The same data set can often be analyzed in a variety of different ways to address the same research question(s). Sometimes different analyses will yield essentially the same outcome, but this is not always the case. This situation, which highlights the combined art and science of data analysis, can lead researchers to choose an approach that yields a more favorable outcome, which is problematic and certainly questionable as well.

**Video:** [https://www.youtube.com/watch?v=ZewOiiOothc](https://www.youtube.com/watch?v=ZewOiiOothc)
8. p-hacking (i.e., running analyses in a manner that produces statistical significance)

**Description:** P-hacking occurs when a researcher continues analyzing their data in different ways until they observe statistically significant findings. This could occur by means of adding new background variables to the model, for example, as new covariates, or by breaking up the sample using different grouping variables. This practice often occurs as a result of researchers' (mistaken) belief that their findings are only important or meaningful if they obtain statistical significance.

**Video:** [https://www.youtube.com/watch?v=ZJnG8twSTvY](https://www.youtube.com/watch?v=ZJnG8twSTvY)

9. Ignoring alternate explanations of data

**Description:** If we only look at one possible explanation of our data, then we can usually validate what we have found in a number of different ways. If we take a second closer look, however, we might find a different explanation. It is good practice to test your hypothesis to make sure that that explanation has scientific validity. When you are doing your data collection, you want to think of as many possibilities as you can that could take down your hypothesis. If you find that you can't eliminate your hypothesis by any other alternate explanation, and re-tests show that it is the most valid one, you can have increasing confidence that it is the best possible explanation of your data.

**Video:** [https://www.youtube.com/watch?v=TzjEXNJENH8](https://www.youtube.com/watch?v=TzjEXNJENH8)

10. Using unjustified methods of handling missing data (e.g., imputing / inserting values for missing data that are not justified and/or that are more likely to yield desired outcomes)

**Description:** Missing data is a reality that we can face in quantitative research for a variety of reasons, such as incomplete survey responses, participant attrition, or technology recording errors. Missing data can impact the power of the statistical model used to analyze the data, so this becomes more critical in research domains where large sample sizes are not the norm and researchers feel that they cannot afford to remove entire cases if one data point is missing. One way of preserving all cases in the data set is to impute or insert values for any missing data points. Yet this can be done using a variety of techniques which may in turn impact the results of our study. Similar to the way we manage outliers, our handling of missing data should be done in a principled and transparent manner. As we encounter missing data, we may want to ask ourselves questions like: What justification do we have for selecting a particular data handling method? How does this method impact our overall results compared to other possible methods for handling missing data? Are we being transparent in reporting the decisions we have made around handling missing data, including their potential impact on our study findings?

**Video:** [https://www.youtube.com/watch?v=0-q6nfGKF0A](https://www.youtube.com/watch?v=0-q6nfGKF0A)
11. Categorizing continuous variables without sufficient justification

**Description:** Statistical techniques such as ANOVA can require categorical variables rather than continuous variables. Therefore, if you have a continuous variable and you really want to use ANOVA, you'd presumably turning this variable into a categorical one, by grouping the data points into bands (e.g., 0-4, 5-10, and so on). This is problematic in that we are losing a lot of information and nuance that way - just to be able to use a statistical test. Ideally, we'd want to collect continuous data when possible and then use them as such in our models.

**Video:** [https://www.youtube.com/watch?v=kfmlGg1CquE](https://www.youtube.com/watch?v=kfmlGg1CquE)

12. Using too many statistics tests without correction (e.g., Bonferroni)

**Description:** This is actually one of those concepts in statistics that took me quite a while to fully appreciate in different ways and I’ll set up an example of why it took me a while to understand this. Let's imagine a scenario in which I have done 20 different independent samples t-tests. The first 19 of them had results that had a higher p-value than .05, which is what we set as our alpha level. And so, we say those tests were non-significant. I then run a 20th test that actually has a significant p-value, which for us is below .05. And just for sake of argument, let's imagine it has a medium effect size. I found something here, right? The data is showing that there is some sort of a score or relationship that we want to investigate. But the problem here is that I've run 19 other tests and I've added a 20th, for 20 total tests. And so you might be thinking, or at least I was always thinking, “why would it matter?” If I only ran one test right here, or that 20th test by itself, the p-value wouldn't change. The effect size wouldn't change. The confidence intervals wouldn't change. Nothing should change as long as we don't manipulate the data. And so why would it matter if we run it 1 out of 1 times or 1 out of 20? And the problem comes down to confidence, right? If we are confident in our results, that we decrease our number of research degrees of freedom and we say “no, we think this is going to be the correct answer.” The chances that there would be higher errors are much more minimal. The fact that we've run 20 tests, what it basically is doing is increasing our signal to noise ratio. When you're at 20 tests, you might be thinking to yourself. “Yeah, that's not too many”, but imagine you did this at a much larger scale. Let's say that you ran 400 tests, or 1000 tests, and you got one out of 1000 tests to be significant. Again, is it because there's a true and natural reason for why that one test was significant, or was it just sort of a fluke? The numbers might have worked out that way. Maybe we have the right sample size, we have the right test conditions, all that kind of stuff. Because again we wouldn’t expect it to be the case that one out of 100 or one out of 1000 tests is significant. And so, when you're running multiple tests you want to use some sort of correction. That is going to make it much harder to get over, and typically our corrections are having a decreased alpha level, so maybe instead of setting a .05 we set at .001 alpha level that we feel more confident in. Or again, we do fewer tests because that again increases our confidence in the tests that we did run. And so, for me this is always a concept that didn't make a lot of sense until I started thinking about it in that fashion, right? That we're doing too much. And inside of that too much, there's just a better chance we'll get a hit and again hit being a positive result than if we did a much smaller size. So, I hope that helps explain this a little bit better.

**Video:** [https://www.youtube.com/watch?v=7i2IPKxavoU](https://www.youtube.com/watch?v=7i2IPKxavoU)
13. Using incorrect statistical methods (e.g., tests that are not appropriate for the type of data being analyzed)

**Description:** Quantitative research often involves application of statistical techniques. As very few of us are trained statisticians, knowing everything we need to know about a technique and how it should (not) be applied is a big ask. However, if we are to be able to trust the results we get from statistical techniques, we need to use them correctly. Questions of whose responsibility it is that the analyses are correct are tricky: are we responsible for checking a co-author's statistical analyses? Is it OK to claim ignorance (i.e., that we did not know that we weren't using the statistical techniques correctly)?

**Video:** [https://www.youtube.com/watch?v=1x70n0rNhhg](https://www.youtube.com/watch?v=1x70n0rNhhg)

14. Interpreting statistical results inappropriately (e.g., claiming equivalence between groups based on a non-statistically significant difference; undue extrapolation)

**Description:** When we use statistical tests, we not only want to make sure they're appropriate for our data (by checking assumptions, etc.), but we also want to make sure we interpret them correctly. For example, we cannot claim equivalence between groups just because we don't find a statistically significant difference between them; all we can say is that we do not have enough evidence to reject the null hypothesis. It is our responsibility as researchers to learn enough about a method such that we are able to interpret the results correctly.

**Video:** [https://www.youtube.com/watch?v=MosWCcA617o](https://www.youtube.com/watch?v=MosWCcA617o)
2.4 Section IV: Write-up and Dissemination

1. Failing to refer to relevant work by other authors

**Description:** Failing to refer to relevant work by other authors is bad practice, and a potential QRP. Few people would disagree with the statement that we should do our best to find and read relevant studies, and then make sure to refer to them, where relevant. Failing to do so for competitive reasons, or because you don't agree with the author's perspective may give readers the impression that that author's work is not relevant, which may reflect poorly on both you and the other author.

**Video:** [https://www.youtube.com/watch?v=CtK4Dnk2dE](https://www.youtube.com/watch?v=CtK4Dnk2dE)

2. Not providing sufficient description of the data analyses or other procedures

**Description:** Is failing to describe one’s data analyses in sufficient detail a questionable research practice? Well, superficial descriptions of data analyses leave many questions open to a reader or an audience and usually make it difficult to assess the results and the value of the study. One can think of many reasons why data analyses are not reported on accurately. Some things may seem so self-evident to the researcher that one doesn’t come to think of having to spell them out explicitly. Or then there may be a length limit to the paper or a conference talk making it necessary to cut down the details. Be that as it may, we should not be happy with insufficient descriptions of data analyses. They are an example of questionable research practices, and something we should not promote.

**Video:** [https://www.youtube.com/watch?v=zIE7X2GDKsg](https://www.youtube.com/watch?v=zIE7X2GDKsg)

3. Not providing sufficient description of the data and the results (e.g., exact p-values, SD)

**Description:** Quantitative studies tend to produce substantial amounts of results in the form of tables, graphs and other numerical information. It is customary to give basic numerical information in one’s publications. Failing to describe the data used for the analyses in sufficient detail, and not to report basic numerical information makes it difficult for readers to grasp the dimensions of a study. For instance, only reporting one’s results in terms of percentages, without giving the corresponding raw figures, usually leaves the reader in the blue about the scope of the study and the value of the results obtained. Indeed, not providing sufficient description of the data and the results is a questionable research practice and something we should not do.

**Video:** [https://www.youtube.com/watch?v=22J-euVi2tE](https://www.youtube.com/watch?v=22J-euVi2tE)
4. Not reporting or publishing results because they are not statistically significant (i.e., the ‘file drawer’ issue)

**Description:** Our field places a lot of emphasis on statistical significance, which results in the tendency to either not report results that are statistically non-significant, or to fail to publish these studies all together. This is known as the file-drawer problem. One reason that this happens is that journals are far less likely to accept studies with statistically non-significant results. Because publishing is important for career advancement, there’s little incentive to attempt to publish statistically non-significant results when the researcher can expect for the study to be rejected.

It’s important to remember that statistically non-significant results that come from a methodologically sound study are still valid and help to build our cumulative knowledge. If you fail to report results that are statistically non-significant, you may be withholding important scientific findings that will help to move the field forward.

**Video:** [https://www.youtube.com/watch?v=FgV_pQEjwkM](https://www.youtube.com/watch?v=FgV_pQEjwkM)

5. Employing selective reporting of results/instruments

**Description:** Wouldn’t you wish to have your study confirm the hypotheses you have presented when launching the analyses? It may be tempting to resort to selective reporting of results to achieve this goal. Or then one might opt for analytical instruments that would enhance a certain outcome of the study to the disadvantage of others. However, reporting on the results only selectively so as to be able to draw conclusions to one desired direction is something we need to consider a questionable research practice.

**Video:** [https://youtu.be/-qjpokm4RWQ](https://youtu.be/-qjpokm4RWQ)

6. Not sharing data when allowable

**Description:** When we conduct a study, it’s important for us to share our data with the research community if we’re able to. Sometimes it isn’t possible to share your data if confidentiality or copyright are at play, but if not, we should be putting our data out there for others to access. There are a few reasons people may be reluctant to share their data: perhaps you put lots of hard work and time into the collection of the data and you feel some level of ownership over it. Maybe you are concerned that another researcher may use the data you collected to publish a study that you had intended to conduct yourself. Regardless of your concerns, there are several reasons that data sharing is important. When you make your data accessible, you may be saving someone else a lot of time that they would have spent collecting a dataset that already exists. Sharing data also helps with accountability. You want people to trust you are a researcher, and the results that you publish, and making data accessible is a good way to do that. Finally, data sharing is important for the sake of collaboration and good-will in the field.

**Video:** [https://www.youtube.com/watch?v=c3qRiOdnY2I](https://www.youtube.com/watch?v=c3qRiOdnY2I)
7. Not sharing scripts used to analyze the results

**Description:** Scripts take time to write and may be messier than we would like, so some may be a bit hesitant to share them. However, making sure to share our scripts increases transparency (shows researchers what we actually did in our study) and helps future researchers carry out replication studies and build on our work.

**Video:** [https://www.youtube.com/watch?v=28eZybxHME&t=2s](https://www.youtube.com/watch?v=28eZybxHME&t=2s)

8. Not sharing instruments/coding schemes

**Description:** There are several reasons why we would want to share instruments and coding schemes. For example, in sharing what we did in a study, we are transparent about all the decisions made, thus enabling others to scrutinize our results and learn from what we did and didn't do. Also, sharing our instruments and coding schemes enables other researchers to use them, thus enabling replication studies and cumulative knowledge building in the field, which makes for results that are more easily comparable. So, failing to do so could be problematic for the field in terms of transparency and community-building of research.

**Video:** [https://www.youtube.com/watch?v=dCozHRUkLrk](https://www.youtube.com/watch?v=dCozHRUkLrk)

9. Not attempting to publish results in a timely manner

**Description:** This QRP may seem less problematic at first glance as it seems that it would only affect ourselves. If we collect data and do not get around to publishing the study for a year or two, it could negatively affect our chances for tenure promotion, our qualification for a job, or our ability to present in conferences. While it may seem that others will not be negatively affected by this QRP, there are in fact other effects that we don’t always think about. First, if you are working with others on the project, this QRP has the potential to negatively affect your co-authors in the same way that it would affect you. If you are not completing the parts of the paper that you are responsible for, and in doing so, holding up the project, they too may not be able to apply for tenure or have a good shot at job promotion. You also have your own participants to think about. If your study stands to benefit your participants and you do not publish the paper for others to read, you may be impeding the ability for your participants to benefit from the research that they participated in. This is important to consider when you decide to take on new projects and plan your research agenda. It’s a good idea to consider which projects you will have time to complete so that these projects do not sit and collect dust.

**Video:** [https://www.youtube.com/watch?v=hFX1PqzBSpo](https://www.youtube.com/watch?v=hFX1PqzBSpo)
10. Presenting misleading figures of data (e.g., displaying a truncated entire y-axis)

**Description:** Data visualization is a key part of how we share our research findings. As we decide how to visualize our data, we need to be mindful of how readers might interpret our data figures. Let’s take a fairly standard design for researching the effects of instruction on second language learning. Imagine a study where one group of learners received an experimental type of teaching and the other group received a traditional type of teaching. The effects of these two teaching approaches were measured by examining learners’ changes in test performance before and after the intervention. Upon analysis, the results showed that the experimental group’s average scores increased from 74% on the pre-test to 77% on the post-test, while the comparison group’s average scores went from 74% to 75%. One way we could present these findings is using a line graph, with time on the x-axis (i.e., pre and post) and average score percentage on the y-axis. For percentage data, we might expect the y-axis to range from 0 to 100. However, this might make it harder to see the changes between the pre- and post-test scores, or we may not feel that this highlights the differences we want our audience to note between the two groups. So, we might be tempted to present a graph with the y-axis ranging only from 70% to 80%, which makes the difference between the pre- and post-test scores for the two groups really noticeable.

By zooming in like this, however, we are removing important context, namely that the two groups only improved by 3% and 1% out of 100% and not out of 10%. While our y-axis may be labeled accurately, a casual reader might glance at a graph of only 10% of the possible data range and get the impression that our findings were more substantial than they actually were. So, as we prepare our figures, it is helpful to ask ourselves how a reader might reasonably interpret the figure and whether misinterpretation is likely.

**Video:** [https://www.youtube.com/watch?v=67w3lHMiYUU](https://www.youtube.com/watch?v=67w3lHMiYUU)

11. Salami publication (e.g., dividing up the results of a single study into multiple manuscripts in order to publish more)

**Description:** We tend to want to have a smaller number of research questions so that we can say more about each one. The problem with salami publications is that you are trying to cut too many pieces out of your dataset, so your questions start to become less meaningful. It's good to use as much information from your dataset as you can and still write a good paper that is comprehensible to your reader. Having a paper that you cut into smaller bits is fine if you can't talk about your whole dataset, but you don't want to cut it down so far that now you are talking about way too detailed of a project. This becomes a questionable practice when it is done for the sake of increasing the number of publications on a CV without actually conducting additional research.

**Video:** [https://www.youtube.com/watch?v=nBYbm4WYP6w&t=35s](https://www.youtube.com/watch?v=nBYbm4WYP6w&t=35s)
12. Not managing time well for one’s own conference presentations, resulting in less time for other presenters, limiting discussion, and impacting others at the conference

**Description:** At academic conferences, we typically view all presentations as having the same level of impact on the field, and thus, there is the expectation that everyone will have an equal amount of time to give their presentation. For this reason, it is important to prepare for a conference presentation in such a way that you will be able to deliver it and take questions within your allotted time. Going over your allotted time will negatively impact the presenter that follows you, potentially causing them to rush through their slides, cut out important parts of their presentation, and harm their overall experience at the conference. This may also negatively impact the audience by causing them to be late for the next presentation they wish to attend. Occasionally, we may go a minute or two over our presentation time, but if we do, we should be prepared to end our Q&A session at the proper time and cede the floor to the next presenter.

**Video:** [https://www.youtube.com/watch?v=1yoInSRz2_U](https://www.youtube.com/watch?v=1yoInSRz2_U)

13. Presenting same presentation at multiple conferences

**Description:** Sometimes there are good justifications to give the same presentation at multiple conferences. It's important to consider who your audience is. If the conference is research-focused, you can expect the attendees to be primarily interested in that aspect of the project, meaning that those interested in pedagogy are less likely to attend. This means that if your study carries important pedagogical implications, those who would be most interested in this information, and most likely to use it, may not ever hear the presentation. This works in reverse - if you are presenting your study at a practitioner-based conference, those who are primarily interested in the research aspect of the project (i.e., methodology, data analysis) are unlikely to attend and will not have access to the information that you are sharing.

The location of the conference often has a similar effect on the audience. Local or regional conferences are generally easier to attend than international conferences as they often mean less travel time and lower cost. Conferences that are farther away and more costly to attend will likely attract fewer of the scholars that you intend to present to. In this case, it may be appropriate to give the presentation at a regional conference even if you've already given it at an international one, and vice versa. If you can make an argument that the audience you will be presenting to is vastly different, this practice may not be so questionable. However, if you expect the audiences at the conferences to be highly similar, this looks more like CV padding. In this case, the practice is more questionable.

**Video:** [https://www.youtube.com/watch?v=TYG-MU8sOyM](https://www.youtube.com/watch?v=TYG-MU8sOyM)
### 14. Employing excessive self-citation

**Description:** What is considered ‘excessive’ can be hard to pin down. Generally, self-citation that appears to be excessive occurs when a researcher is one of a highly limited number, or potentially the only researcher, who conducts a particular type of niche research that is relevant to the paper. If this is truly the case, then this QRP may not be so questionable as it is also not appropriate to purposefully limit self-citation. This becomes problematic when the author does so in order to pad a CV or increase name recognition and readership. In addition to this, if you are failing to cite relevant studies by other authors, this may also be unethical as you are indirectly impeding another researcher’s advancement in the field. Even if you believe that you are one of the few people who conducts your type of research, be certain to do some digging in the literature in order to be certain that you are not failing to cite other authors. Be certain to consider whether a citation is truly necessary based on the language you used, and whether the citation is absolutely relevant to the study.

**Video:** [https://www.youtube.com/watch?v=gWwqc-VWsyQ](https://www.youtube.com/watch?v=gWwqc-VWsyQ)

---

### 15. Intentionally omitting relevant work because it does not align with one’s theoretical or methodological approach

**Description:** We all tend to work within some theoretical or methodological framework. It is therefore not inconceivable that we will disagree with statements or approaches from a competing framework. It gets a little tricky, however, when it comes to the background section of an article - to what extent is it OK to omit relevant work because it doesn't align with your own approach? One could maybe argue that those papers aren't as relevant as ones produced in your framework, but we of course do not want to end up in an echo chamber where we only cite people we agree with.

**Video:** [https://www.youtube.com/watch?v=CIcc8xTOA4](https://www.youtube.com/watch?v=CIcc8xTOA4)

---

### 16. Inappropriately including or excluding authors

**Description:** Including authors on a paper that have not adequately contributed is known as ‘gifted authorship’. This is particularly problematic when the individual is included in order to reward the individual in some way, return a favor, or pad their CV. This sometimes also happens when lesser-known researchers include a well-regarded, senior scholar on a paper in order to increase readership, citations, or recognition of their own names. This is problematic not just because it gives the impression that the individual has contributed to the field in ways that they haven’t; it also means that the individual then becomes responsible for the paper, regardless of the quality of the study. Inappropriately excluding authors is also problematic, simply because researchers are not getting credit for the work that they have done. It’s important to remember to have these conversations early in the research process, and often.

**Video:** [https://www.youtube.com/watch?v=jf6gPww0Paw](https://www.youtube.com/watch?v=jf6gPww0Paw)
17. Inappropriately attributing author roles when listed in publications

Description: Clearly delineating authorship is an important aspect of the publication process, especially in fields where collaborative research and publication are more common. Part of this process is accurately attributing the roles that each author played in the project. A growing number of journals now provide opportunities for authors to report not only that an individual was involved in a project but also in what ways the author was involved. This is a matter of transparency, and it allows readers who may have questions about certain aspects of a study to know which author(s) would best be able to answer these questions. For example, if only one author from a team of three conducted the data analysis, this role should not be attributed to all team members as they may not be equally prepared to explain the nuances of the statistical tests used. This level of transparency can offer a more realistic picture of what collaboration looks like in our research while also serving as a check that everyone listed as an author contributed to the publication in a meaningful way. Publications serve as a record of our scholarly work, so appropriate attribution of author roles helps us ensure that that record is accurate.

Video: https://www.youtube.com/watch?v=F1Akx2R42M0

18. Inappropriate ordering of authors

Description: Not everyone will agree on which contributions to a paper are most valuable and should be rewarded by a higher position in author order. We may believe that author order should reflect the amount of time or writing that an individual puts into the paper. Others might feel that the variety of contributions at different stages of the process matter most, or that the individual who emerges as the ‘leader’, scheduling meetings, setting deadlines, etc. deserve a higher position in the author order. We may also be tempted to put well-known or respected scholars, academics working towards tenure, or early-career graduate students who would benefit greatly from being a first author on a paper higher in author order. In the latter two cases, this begins to look like gifted authorship. It is best if this conversation can happen before the research process begins. Have conversations about how each author can best contribute to the study, which pieces of the research they believe they are best equipped to take the lead on, and importantly, what this means for author order.

Video: https://www.youtube.com/watch?v=B5O17Xw4XDo
19. Not giving research assistants due credit in publications

**Description:** This is a difficult QRP because there are many different opinions about the role of a research assistant. Research assistantships often come with benefits such as mentorship and financial compensation. In these situations, some may feel that RAs are already compensated for their work, and that authorship credit should only be awarded to them if they go above and beyond their expected role, for example, participating in the study design conversations or data analysis. Other academics fall into the opposite camp on this issue, feeling that if RAs contribute to the project, they are naturally a part of the research team. Those who take this position feel that RAs should be held to the same standards for authorship as the other researchers on the project are, regardless of their status as a research assistant. This decision, however, is highly dependent upon various other contextual factors, such as the level of the student (e.g., undergraduate students versus Ph.D. students), the nature of the project, the mentor or PI, and the roles that are expected of the research assistant. The best way to handle this is to get ahead of this issue. It’s useful for the mentor or PI to have an open discussion with the research assistant before the project starts in order to determine what will qualify the RA for authorship credit moving forward.

**Video:** [https://www.youtube.com/watch?v=JpzlxIshyKQ](https://www.youtube.com/watch?v=JpzlxIshyKQ)

20. Exaggerating the implications and/or importance of findings in order to increase likelihood of publication

**Description:** When we have worked hard on a project for a long time, we are likely to be (a) excited about the topic and the findings and (b) eager to see it in print. It might therefore be tempting to exaggerate the importance of the findings. This may or may not be an issue in all contexts, in that reviewers and editors tend to be able to see through overly enthusiastic accounts of something, but we should of course always try to be as objective as possible when presenting our findings.

**Video:** [https://www.youtube.com/watch?v=sErMAjica4Q](https://www.youtube.com/watch?v=sErMAjica4Q)

21. Lifting short phrases from others without quoting directly

**Description:** Especially when we start out in the field, we may struggle to know how to avoid having overly cautious citation practices while still not plagiarizing ideas and text from other researchers. To make matters worse, we may need to know the field relatively well to know which concepts and ideas are considered 'common knowledge' and which always require citation. As a rule of thumb, it's of course better to air on the side of caution, and 'overcite' rather than 'undercite'.

**Video:** [https://www.youtube.com/watch?v=fTdl6mZ2NPU](https://www.youtube.com/watch?v=fTdl6mZ2NPU)
22. Irresponsibly co-authoring (e.g., not being involved enough to be able to verify accuracy of analysis)

**Description:** In academics, getting credit for publications is the way we get paid (other than the actual money they give us for working). As scholars we use this credit to help us join graduate programs, get jobs, gain tenure, boost our CVs, secure funding, and more. The more academic bling you earn the higher your status in the field tends to be and the more opportunities open to you. So it is important to make sure that all authors listed on a publication deserve to be there. If a person is “gifted” authorship, they unfairly benefit from gaining academic credit without having done the work, meaning they might have a better chance of getting a job or a prestigious grant over someone else who has done the work. From an ethical standpoint, gifting authorship is also problematic. If I conduct research and do a lot of shady things and then list you as an author, you become responsible for the things that I did. You agree to what was discussed in the paper, even if it is questionable work or includes offensive language. Because you are now an author who helped “write” the paper. What counts as authorship can be hard to pin down. The APA says: An author is considered anyone involved with initial research design, data collection and analysis, manuscript drafting, or final approval. But not everything counts. Things like mentorship or being paid in a different way (such as a research assistant) are usually not enough to earn authorship credit. Saying “I wonder if X impacts Y” is not the same thing as sitting down and defining X and Y and then figuring out how to test the impacts. The best way to avoid issues of authorship is with open communication. Talk with people connected to the project and see if they want to be part of a research team and contribute meaningfully to the work. Write up contracts explaining what everyone’s roles and expectations are from the start. Avoid headaches before they start.

**Video:** [https://www.youtube.com/watch?v=u5gIUuAEYK8](https://www.youtube.com/watch?v=u5gIUuAEYK8)
Lesson Plans

3.1 Directions

Sample lesson plans have been provided for courses of the following lengths: 30 – 50 min, 1 – 1.5 hours, and 2 – 2.5 hours (Section 3.2). The activity bank (Section 3.3) includes a mixture of teacher-lead and student-lead activities that can be substituted into the sample lesson plans based on the participants’ familiarity with QRP. A set of scenarios and reading excerpts are provided for select QRP in Section 3.4. A list of suggested readings is provided in Section 3.5.

3.2 Sample Lesson Plans

If you will only have the opportunity to lead a single session related to QRP, we suggest that you select QRP on that come from each of the four stages of research (Funding, Design/Data Collection, Data Analysis/Interpretation, Write-up/Dissemination) (Plonsky et al., submitted). Stress that these ethical decision points occur throughout the research process.

If you will have the opportunity to integrate multiple sessions focused on QRP into your curriculum, we suggest that you select QRP for each lesson as they become relevant (work through one stage of research at a time). Mix in more advanced activities as the students learn more about QRP.

Please draw from the activity bank provided in Section 3.3 to tailor these lesson plans to the needs of your students.

30-40 min lesson

(Before class)
- QRP readings (e.g., Sterling et al., 2023)
- 3-4 scenarios to read and think about (one from each stage of research or 4 from the same stage if the QRP lessons are recurring throughout the semester)

Pre-task (5-10 min):
- Opening comprehension check from reading – what are QRP?
  - Discuss in pairs, check in as a class

Main activity (20 min):
- In small groups, discuss the scenarios provided before class
  - Discuss responses as a class

Post-task (5-10 min)
- Discussion: Why are QRP a problem – what is at risk? What can we do to avoid making questionable decisions in these circumstances?
1 - 1.5-hour lesson

(Before class)
- QRP readings (e.g., Yaw et. al., 2023)

Pre-task (10 min):
- Rate a list of QRPs individually for opinion on severity

Main activity (30-40 min) (select one activity or shortened versions of both depending on time):
- Activity 1: Provide students with survey data from Larsson et al. (submitted). Compare personal opinions (from pre-task ratings) to survey data. How do your opinions compare to the consensus of the field?
- Activity 2: Mini-debates - put students in small groups, assign each group a QRP, pair up the groups to have mini-debates about severity (students make arguments based on contexts, risks, etc.)

Post-task (15 min):
- Re-rate same QRP list for severity
  - Class discussion
    - Did your opinions change?
    - What were changes in opinion based on?
    - How do we recognize “it depends” circumstances?

2 - 2.5-hour lesson

(Before class)
- QRP readings (e.g., Larsson et al., submitted)

Pre-task: (20 – 30 minutes)
- Provide students with a list of QRPs (sample from all four areas, or, do all QRPs from one area) – let students rate QRPs individually for their opinion on severity
- Talk in small groups about responses: Why did students respond the way that they did?

Main activity (1 - 1.5 hour)
- Put students in small groups (2-3), assign (or let students select) a QRP to focus on
- Watch QRP vid (from YouTube), give access to short reading excerpts about the specific QRP (if accessible)
- Students prepare to give a small presentation to the class about the QRP
  - Explain the QRP
    - Explain why it can be problematic and in what contexts: what is at risk?
    - Discuss what forces may influence the decision: what may lead us to make a questionable decision and what makes it questionable?
    - Discuss contexts when the QRP may be problematic (if students believe it is), and contexts in which it may not be a problem
  - Discuss ways that we can minimize risk for that specific QRP
Post-activity (20 min):

- Full class discussion or personal reflection
  - Why are QRPs a problem – what is at risk?
  - Have you ever found yourself in a situation where you had to make one of these decisions? What did you do? Would you do something different now? What could you have done to avoid making a questionable decision?
3.3 Activity Bank

Substitute activities from the activity bank into your lesson plans in order to tailor the session to the needs of your students.

Scenarios for scenario activities are provided in the following section (Section 3.4).

1. Provide scenarios, discuss in small groups, report back to class
2. Students rate a list of QRPs for severity on their own (or come to a consensus as a group), then compare student opinions to survey data from (Larsson et al.). Use survey ratings and qualitative comments. Compare personal opinions to the opinions of the larger field
3. Students rate a list of QRPs for severity on their own, discuss in small groups, make notes of differences in opinion, report back to class (shorter classes)
4. Mini-debates - select a set of QRPs, put students into small groups to have debates about the order of perceived severity
5. Students use materials (short readings, YouTube videos) to learn about a QRP in small groups, teach the class through short group presentations (long class only)
6. Provide students with a list of QRPs, have students rate perceived severity on their own (pre-task), students re-rate the same list of QRPs at the end of the class (post-task/reflection)
7. Learn about a QRP in groups (video, short paragraph-page reading, etc.), write scenarios, pass to another group. Groups discuss and report responses back to the class (longer classes, classes that have had some scaffolding).
   8. Whole class or small group discussion
      a. Why are QRPs a problem – what is at risk?
      b. Have you ever found yourself in a situation where you had to make one of these decisions? What did you do?
      c. What can we do to avoid making a questionable decision in these circumstances?
9. Have students discuss situations in which particular QRP items would be deemed acceptable.
10. Create a short scenario of your own in which you talk about a time that you made a questionable research decision. Explain the circumstances and what you would do differently now.
3.4 Scenarios / Discussion Questions / Extra Readings

Citations for all supplemental readings can be found in the Suggested Reading list (Section 3.5)

Section I: Funding

QRP #5: Misrepresenting study importance in the proposal

- **Scenario:** John has been told by his Chair that in order to have a stronger tenure portfolio to be able to get promoted, he needs to apply for external funding and ideally get a grant. John has very limited experience writing up grant proposals. He sends a first draft to two senior colleagues to get some feedback, and they both advise him to emphasize the importance of the implications of the project. John isn’t clear on how to do so, as the project, while interesting to him, doesn’t seem to have clear implications for either teaching or research. As he really values the advice from his senior colleagues, and as he feels some pressure to write a strong one to help his tenure case, he adds some language that he feels exaggerates the impact and value of his proposal to society. He sends it back to his senior colleagues who both give it a thumbs up. He is hesitant to submit this revised version, as it does not feel like he is accurately representing the project. Should he submit it?

QRP #7: Not producing the promised project outcomes due to project mismanagement

- **Scenario:** Nina is working on a grant application with a team of researchers at different stages of their careers. As she is the primary investigator (PI), she’s in charge of drafting a budget. She reaches out to two of the team members who have a high success rate at getting grant funding to ask for advice on how much to apply for. Her draft budget is $110,000, as she is thinking that $100,000 would cover the salary costs, given everyone’s salaries and months devoted to the project, and that $10,000 would be enough to cover travel costs for conferences for the duration of the project. The team members tell her to add at least 20% to that, as, according to them, grant agencies often cut budgets. Nina is highly uncomfortable doing so, as she feels that the higher predicted costs would be unwarranted, but she doesn’t want to go against the advice of her team members. Should she take her colleagues’ advice?

Section II: Design/Data Collection

QRP #3: Defaulting to convention and QRP #4: Employing instruments/measures without a strong validity argument

- **Extra readings:**
  - Wallis, 2021, p.27 (excerpt)
  - Mackey, 2020, pp. 202 - 203 (pertaining to appropriate methods of data collection)
QRP #10: Having an unnecessarily long/burdensome data collection for participants

- **Scenario:** Lin is conducting a study on intelligibility and comprehensibility of authentic audio materials in the listening classroom (Study 1). While designing the study, Lin realized that the target participant population overlapped with another research study on learner beliefs that they had wanted to conduct but had not yet started (Study 2). Knowing that participant recruitment can be challenging and time-consuming, Lin decided to include instruments from Study 2 as part of the data collection for Study 1 so that they could have additional data to analyze separately. When Lin piloted this data collection, they found that it took participants about an hour to complete the Study 1 procedures with an additional 20 minutes to complete the Study 2 instruments. Some pilot participants noted that the study felt long, while others found it interesting and did not have a problem with the amount of time it took. Should Lin go ahead with this plan for an expanded data collection?

QRP #11: Recruiting participants to join a study in a way that makes refusal difficult or uncomfortable

- **Scenario:** Fatma is taking a research methods course and is required to conduct a complete research project within a single semester. Fatma is ready to gather data for her project and has asked people in her department if it would be possible to visit their classes and recruit participants. Fatma has visited three classes already and things went mostly fine. She went to the final few minutes of class and read her IRB approved recruitment script. She then hung up a sheet with a QR code that students could use to sign up for the study as they left the classroom. Today, Fatma visited Agustín’s classroom and followed the same routine as before. However, near the end of Fatma’s pitch, Agustín added a few words to help motivate the students. He mentioned how important research was, how students at this campus need to support each other, and that Fatma might not pass her class without the help of the students. In the first three classes that Fatma visited, around 1 in 4 students agreed to take part in the study. In Agustín’s class, nearly 3 out of 4 agreed to do so.

Section III: Data Analysis/Interpretation

QRP #1: Removing whole items/cases knowingly/purposefully to obtain favorable results

- **Scenario:** Kai is conducting a study looking at changes in cultural understandings and attitudes held by first year language students. As part of this semester long project, Kai is conducting observations in four classrooms along with a repeated survey (4 times) that
attempts to track changes in attitudes among the students. While analyzing the survey data, Kai noticed that four of the 63 participants appeared to not take the survey task seriously. These participants selected 1 or 6 (the anchor points) for all questions in each of the four surveys. In looking back at their observation notes, Kai realized that these four students were absent from the classes at higher rates than other students and did not engage with class discussion unless directly called upon. Prior to running any statistical tests, Kai decided to remove the data of these four students since they felt that the data was unreliable. Kai later re-ran the tests including the removed participants and found that including the participants changed the outcome of the statistical tests in ways that did not align with Kai’s hypotheses.

QRP #2: Using unjustified methods of handling outliers

- **Extra reading:** Mackey, 2020, pp.192-193 (‘Outliers in Interaction, Feedback, and Task Data’)

QRP #4: Being ambiguous about whether an exploratory vs. confirmatory analysis was employed

- **Scenario:** Ren is interested in what makes students successful language learners. His previous research had shown that students whose background included having good language teachers in high school, living in an area where the second language was often used, having a desire to consume media in the L2, and seeing a future usage of the language resulted in participants rating their motivation to learn language higher than students who lacked more than one of those components. In trying to design a follow up study, Ren decided to send a survey to students in 400 level language courses and asked them about their past and future selves in terms of learning and using their language. The survey included items such as “Knowing how I will use my language in the future is important” or “Watching movies in my L2 is a waste of time”. The results of the survey showed almost exactly what Ren thought it would. Most of the students mentioned having positive language experiences prior to going to college and they all had plans for how they wanted to use the language in the future. Since the data was so robust, Ren put forward an argument about the need to strengthen language opportunities in high school and to include career readiness components in all levels of language courses to help motivate students.

QRP #5: HARKing (i.e., hypothesizing after results are known)

- **Scenario:** Finally, it was done. Noel had spent several months brainstorming a research idea, fought for weeks with the IRB, tried four different methods to recruit enough participants, and sent out hundreds of email reminders. And now, Noel was looking at his glorious spreadsheet full of information. Noel had originally wanted to investigate if the text color used for subtitles would have an effect on the amount of material
comprehended by participants. He used five different font colors (the only ones available on his media player) and found that light hues (gold and orange) had significantly higher levels of comprehension than did dark hues (green, black, and purple). In his write up, Noel hypothesized that lighter hues were more likely to draw attention without interfering with the onscreen images. His null hypothesis was that there would be no difference between the two hues, which was rejected by the t-test he ran.

QRP #6: Cherry-picking data to analyze
- Extra reading: Wallis, 2021, pp. 198-200 (excerpt)

QRP #8: P-hacking
- Extra readings:
  - Stefanowitsch, 2020, pp. 173-174 (excerpt)
  - Wallis, 2021, p.274 (‘Re-sampling and p-hacking’)

QRP #9: Ignoring alternate explanations of data
- Extra reading: Wallis, 2021, pp. 198-200

QRP #10: Using unjustified methods of handling missing data
- Extra reading: Wallis (2021) - Ch.16: Conducting Research with Imperfect Data

QRP #11: Categorizing continuous variables without sufficient justification
- Extra reading: Phakiti et al., 2018, p. 517 (‘When ANOVA is Inappropriate: Forcing Continuous Data to be Nominal or Ordinal’)

QRP #13: Using incorrect statistical methods
- Extra readings:
  - Wallis (2021) - Ch. 13: Choosing the Right Test
  - Egbert, Larsson, Biber (2020) - Ch. 6: The Role of Statistical Analysis in Linguistic Description [corpus research]

QRP #14: Interpreting statistical results inappropriately
- Scenario (Note: can also be used for Write-up/Dissemination #9): Julio is interested in the relationship between the correlation between two individual difference variables hypothesized to be linked to second language proficiency. He, therefore, administers an online survey to measure those variables along with a background questionnaire to gather
demographic data on his sample of Chinese learners of English. Thanks to connections with scholars in China who kindly forward his invitation to participate, Julio collects a fairly large sample (N = 650). The correlations between his two variables of interest turn out to be very small ($r = .03$). Due to the large sample, however, the correlation is statistically significant. Julio highlights the statistical significance of this finding in the abstract and throughout the results and discussion, only mentioning once the observed correlation of .03.

### Section IV: Write-up Dissemination

**QRP #6: Not sharing data when allowable**

- **Extra reading:** Paquot & Gries, Eds., 2021, p.16 (excerpt - sharing corpora)

**QRP #7: Not sharing scripts used to analyze the results**

- **Scenario:** Jean recently published a study that involved the important methodological component of creating a collocation/concordancing program in Python for corpus linguistic research. This program was capable of: (1) transforming an enormous corpus into a readable dataset that the program could process in a matter of seconds, (2) processing user-input for collocate searches, and (3) organizing output into an Excel document for the user. At the completion of her project, she was asked to share this Python script on a public platform so that other researchers could use it. As only an intermediate-level programmer, Samantha was hesitant to share the Python script because, in her view, the code was written rather crudely. She questioned whether publicly sharing the script would hurt her reputation in the larger research community as a programmer, and whether the script would actually be useful to other researchers in the field who were potentially more adept at programming. As she knew of a handful of researchers with programming skills, she reasoned that there were other researchers who could likely write a much more efficient, user-friendly program with the same capabilities. For these reasons, she opted not to share her Python script. Was this a reasonable decision to make?

**QRP #8: Not sharing instruments/coding schemes**

- **Extra reading:** McKinley & Rose, Eds., 2019, pp.100-101 (‘Open Science’)

**QRP #9: Not attempting to publish results in a timely manner**

- **Scenario (Note: also can be used for Data Analysis/Interpretation, QRP #14):** Julio is interested in the relationship between the correlation between two individual difference variables hypothesized to be linked to second language proficiency. He, therefore,
administers an online survey to measure those variables along with a background questionnaire to gather demographic data on his sample of Chinese learners of English. Thanks to connections with scholars in China who kindly forward his invitation to participate, Julio collects a fairly large sample (N = 650). The correlations between his two variables of interest turns out to be very small (r = .03). Due to the large sample, however, the correlation is statistically significant. Julio highlights the statistical significance of this finding in the abstract and throughout the results and discussion, only mentioning once the observed correlation of .03.

QRP #11: Salami publication

- **Scenario:** Cynthia and Sam are a pair of graduate students who have spent the better part of a year collecting data for a study investigating the effect that the gender and age of the interlocutors in the TOEFL iBT spoken exam has on test scores. It’s time to start the write-up, and Cynthia and Sam must decide what content to include in their paper. They have two options: (1) they can write one paper that presents the results and analysis of the gender/scores relationship and the age/scores relationship, or (2) they can write two shorter papers, one focused on the relationship between age and test scores, and one focused on gender and test scores. If they choose option 1, they will have a difficult decision to make as to which student will be the first author for the project. If they choose option 2, each of the graduate students will have a first authored paper. While they feel that they could easily fit the results and analysis into a single paper, they have put equal work into the project over the course of the past year, and would both benefit greatly from a first authored paper as they enter the academic job market. Which option should they choose? (adapted from scenario provided by the Center for Taiwan Academic Research Ethics Education: [https://ethics.moe.edu.tw/resource/case/detail/32/](https://ethics.moe.edu.tw/resource/case/detail/32/))

  - **Extra reading:** Mackey, 2020, pp. 45-46 (short excerpt)

QRP #13: Presenting the same presentation at multiple conferences

- **Discussion questions:**
  - Have you ever presented the same presentation at more than one conference? If so, why?
  - What circumstances can you think of when presenting the exact same presentation may be just fine? May be questionable?
  - What do you feel must change in a presentation for it to be considered ‘not the same’? The slides themselves? The findings focused on in the presentation? The data used?

QRP #14: Employing excessive self-citation

- **Scenario:** How many self-citations are too many? Imagine your co-author sends you the most recent version of a paper you are both working on. Your co-author has a fair bit of experience in the topic of the paper. However, it seems that he has inserted citations of
his own work in places where it doesn’t seem especially relevant and/or where reference to another work might be more appropriate. Would you bring this up with him? How so?

- Extra reading:
  

QRP #17: Inappropriately attributing author roles when listed in publications

- Scenario: Misha is a first-year assistant professor. She’s leading a research project where she’s working with two other researchers: an MA student and a senior scholar. The original plan was for all three to be equally involved in the project, but as time went on, it became clear that most of the work has been done by the MA student and by Misha. The senior scholar, Dr. Marke, has offered to have her name taken off the paper to reflect this, but Misha is hesitant to do so, as she worries that it will be more difficult to publish papers from the project without the name recognition of Dr. Marke. Also, Dr. Marke has been helpful in an advisory role. When it comes time to submit the first paper from the project, their target journal asks the team to detail the contributions of each of the team members. None of the categories to choose from cover the kind of support Dr. Marke has provided, so unless Misha adds something like ‘taking an active role in the write-up’ (the scholar did rewrite one of the paragraphs, she reasons), Dr. Marke’s box will be empty.

  Should Misha bend the rules a little to keep Dr. Marke listed as an author?

QRP #18: Inappropriate ordering of authors

- Scenario: Harriet and Gloria recently completed a study for a graduate seminar investigating language variation of users posting in political activist groups on different social media platforms. They put in a year of hard work completing the project, which required the extensive process of building a large stratified, representative corpus of 5,000 social media posts, writing various concordancing programs to analyze their data, interpreting the findings, and writing up the manuscript. As they began to look for a journal that was a good fit for their study, Gloria’s mentor, a well-respected researcher in the field, contacted Gloria and Harriet to inform them that he’d been asked to submit a study to a high impact journal for a special edition on a related topic. The professor asked if the three of them could co-author the manuscript for submission. Based on this seemingly great opportunity for the students to publish a prestigious journal, and based somewhat on the pressure felt by Gloria not to turn down the offer from her mentor, Harriet and Gloria agreed. However, following this conversation, the professor, without consulting the students, made minor edits to the methods section of the paper, placed his name as the first author, and submitted the manuscript to the journal. Harriet and Gloria both privately felt that, due to their independent conceptualization of the study, compilation of the corpus, analysis of the data, and write-up of the manuscript, they should have been first and second author on the paper. Due to the power dynamic in this
relationship, and the fear of losing the opportunity to publish, they are hesitant to approach Gloria’s mentor. What would you advise them to do?

QRP #22: Irresponsibly co-authoring

- **Scenario:** Maisy was a research assistant for Dr. Glass and was tasked with transcribing about 50 hours of focus group data. The project had something to do with how L2 French learners talked about trees and outdoor spaces. Honestly, Maisy wasn’t all that sure what the goal of the project was but she did hear a lot about trees. She spent weeks transcribing the data in the format that Dr. Glass had taught her. They had weekly meetings to talk about the transcriptions and Dr. Glass would often ask Maisy if she had noticed anything funny or interesting in the data. Since Maisy struggled to figure out who was talking and when, she had started to notice various cues the participants used to signal that they wanted a chance to speak. Maisy also mentioned the instances when the participants used incorrect French terms as every instance required Maisy to relisten to the same clip numerous times. Maisy kept working on creating the transcript throughout the semester and was surprised several months later when Dr. Glass mentioned that she had just submitted their manuscript. She was proud that Maisy was able to not only transcribe data but also helped with the data analysis.
3.5 Suggested Readings


Authors’ Contact

With questions or comments, please contact one of the following authors:

Margaret Wood, Ph.D.
Department of English, Northern Arizona University
mkw57@nau.edu

Tove Larsson, Ph.D.
Department of English, Northern Arizona University
Department of English, Uppsala University
Tove.Larsson@nau.edu

Luke Plonsky, Ph.D.
Department of English, Northern Arizona University
luke.plonsky@nau.edu

Scott Sterling, Ph.D.
Department of Languages, Literatures, and Linguistics, Indiana State University
Scott.Sterling@indstate.edu

Merja Kytö, Ph.D.
Department of English, Uppsala University
merja.kyto@engelska.uu.se

Kate Yaw, Ph.D.
Department of World Languages, University of Southern Florida
kyaw@usf.edu