A Thesis
Submitted to the Faculty
of
Xavier University
in Partial Fulfillment of the
Requirements for the Degree of
Master of Arts
by
Sarah Iman
March 16, 2016

Approved:

Karl W. Stukenberg, Ph.D.

Karl W. Stukenberg, Ph.D., ABPP
Chair, School of Psychology

Morell Mullins, Ph.D.
Morell Mullins, Ph.D.
Thesis Chair
Publication Bias and Graduate Students’ Perceived Trust in the Literature
<table>
<thead>
<tr>
<th>Role</th>
<th>Name</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chair</td>
<td>Morell Mullins, Ph.D.</td>
<td>Professor of Psychology</td>
</tr>
<tr>
<td>Member</td>
<td>Dalia Diab, Ph.D.</td>
<td>Assistant Professor of Psychology</td>
</tr>
<tr>
<td>Member</td>
<td>Mark Nagy, Ph.D.</td>
<td>Associate Professor of Psychology</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to take this page to acknowledge several people who helped make this thesis possible. Thank you Dalia and Mark for serving as committee members, your thoughts and insight were critical. I want to thank my thesis chair and mentor Morrie. Your research was the starting point for my thesis and without you sharing your initial study my thesis would be much different. Also, thank you for encouraging me to push myself, to think about research in a new ways, and to keep asking questions. Often, in an academic career a person is lucky to have a few instructors who make a lasting impact. I can honestly say all three of you made a lasting impact on not only who I am as an I-O Psychologist, but as a person.

I would like to thank my husband, Josh, for helping me through this process. We have both made sacrifices along the way. I hope you know the sacrifices you made did not go unnoticed. Thank you for understanding when I needed to work late or change our plans to progress on my thesis. Also, thank you for knowing when I needed to take a break and encouraging (or sometimes forcing) me to do so. Without this help, I probably would not have loved the thesis process as much as I did.
Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>i</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>ii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>iii</td>
</tr>
<tr>
<td>List of Appendices</td>
<td>iv</td>
</tr>
<tr>
<td>Abstract</td>
<td>v</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. Review of the Literature</td>
<td>1</td>
</tr>
<tr>
<td>II. Rationale and Hypotheses</td>
<td>16</td>
</tr>
<tr>
<td>III. Method</td>
<td>20</td>
</tr>
<tr>
<td>IV. Results</td>
<td>29</td>
</tr>
<tr>
<td>V. Discussion</td>
<td>40</td>
</tr>
<tr>
<td>VI. Summary</td>
<td>50</td>
</tr>
<tr>
<td>References</td>
<td>62</td>
</tr>
<tr>
<td>Appendices</td>
<td>66</td>
</tr>
</tbody>
</table>
List of Tables

Table | Page
--- | ---
1. Sample Demographics | 22
2. Correlation Matrix | 35
3. ANOVA Testing Program Type Differences Across PTM Subscale | 38
<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.</td>
<td>Demographic Information</td>
<td>66</td>
</tr>
<tr>
<td>B.</td>
<td>Amazon Gift Card Contest Survey</td>
<td>69</td>
</tr>
<tr>
<td>C.</td>
<td>Perceived Trustworthiness Scale</td>
<td>70</td>
</tr>
<tr>
<td>D.</td>
<td>Distrust Scale</td>
<td>76</td>
</tr>
<tr>
<td>E.</td>
<td>Social Cynicism Scale</td>
<td>77</td>
</tr>
<tr>
<td>F.</td>
<td>Negative Affectivity Scale</td>
<td>78</td>
</tr>
<tr>
<td>G.</td>
<td>IRB Approval Letter (Sampling Strategy 1)</td>
<td>79</td>
</tr>
<tr>
<td>H.</td>
<td>Request for Participation Email (Sampling Strategy 1 and 2)</td>
<td>80</td>
</tr>
<tr>
<td>I.</td>
<td>Informed Consent Form</td>
<td>82</td>
</tr>
<tr>
<td>J.</td>
<td>Debrief</td>
<td>84</td>
</tr>
<tr>
<td>K.</td>
<td>IRB Approval Letter (Sampling Strategy 2)</td>
<td>85</td>
</tr>
<tr>
<td>L.</td>
<td>IRB Approval Letter (Sampling Strategy 3)</td>
<td>86</td>
</tr>
<tr>
<td>M.</td>
<td>Request for Participation Email (Sampling Strategy 3)</td>
<td>87</td>
</tr>
<tr>
<td>N.</td>
<td>Final Perceived Trustworthiness Measure</td>
<td>89</td>
</tr>
</tbody>
</table>
Publication bias and questionable research practices (QRPs) are a growing topic of concern (Jeffrey et al., 2015; Kepes & McDaniel, 2013; Koehler et al., 2015; Mullins, Crowe, & Wymer, 2015). Developing a scale to accurately measure the consequences of publication bias on outcomes such as perceived trustworthiness of the literature is critical to evaluate the extent to which publication bias, and QRPs, impact science, and is the goal of the present study. Items were developed using the available research information. It was hypothesized that there would be a negative correlation between graduate student awareness of publication bias and their level of trust in the literature, which was supported. Additionally, distrust, social cynicism, and negative affectivity (NA), were included in the study to assess the validity of the developed measure. It was hypothesized that all validation constructs would negatively correlate with the perceived trustworthiness scale. None of these were supported; however there are several explanations for this. These implications, and study limitations, are discussed further.
Chapter I

Review of the Literature

Publication bias, questionable research practices, and the general approach to organizational science has become a growing topic of concern both within the published research and at professional conferences (Jeffrey et al., 2015; Kepes & McDaniel, 2013; Koehler et al., 2015; Mullins et al., 2015). For example, at the Society for Industrial and Organizational Psychology (SIOP) 2015 conference, the Theme Track was “Rethinking Our Approach to Organizational Science,” featuring presentations that focused on “improving the peer review process” and “pursuing better science in organizational psychology” (Jeffery et al., 2015; Koehler et al., 2015). Publication bias is also a potential issue in all domains of research and science.

In the present study, an initial review addresses the definition of publication bias, the types and reasons why publication bias is a problem, consequences it brings, as well as the related topic of questionable research practices. I also review the issues originally raised by Kepes and McDaniel (2013) regarding how these practices affect the trustworthiness of the resulting literature, as well as a follow-up article by Mullins, Crowe, and Wymer (2015) about how the resulting literature is considered by graduate students. Mullins et al. took an exploratory approach to develop an initial measure of perceived literature trustworthiness. Although their study was a step towards understanding some potential consequences of publication bias, additional research to
develop a more robust scale with a larger and more diverse sample to assess the consequences of publication bias and related questionable research practices is needed.

Developing a tool to accurately measure the consequences of publication bias on outcomes such as perceived trustworthiness of the literature is critical to evaluate the extent to which publication bias, and related questionable research practices, impact science, and was the goal of the present study. Items for this scale derived from the literature review and the arguments drawn from Mullins et al. (2015) and Kepes and McDaniel (2013). Items also derived from concerns noted in the literature that may have an impact on how graduate students perceive literature trustworthiness. Additionally, I was interested in evaluating how graduate students intend to use research literature upon graduation and employment. This could potentially have some interesting, yet concerning, implications for science in the long term if practices such as publication bias and questionable research behaviors continue.

**Publication Bias**

Publication bias is the “selective publication of studies with a particular outcome” (Ferguson & Brannick, 2012, p. 120). Effects of publication bias are not specific to I-O psychology, or even to the field of psychology (Ferguson & Brannick, 2012; Kepes & McDaniel, 2013). Understanding the impact publication bias has is critical and needs further examination for many reasons. Scientific learning and research often starts with evaluating published work. For example, many graduate programs use published research journals in lieu of, or in addition to, textbooks to teach their students about their field of research. Furthermore, when developing a new study, even established academics use published work to not only support their hypotheses and research design,
but to also develop and focus their research questions. Practitioners evaluate research from their education and training to develop plans and solutions to organizations’ or clients’ problems. If published research inaccurately represents reality, then what students, academics, and practitioners are learning, researching, and applying is flawed. All sciences face these issues and need solutions to reduce and eliminate them (Kepes & McDaniel, 2013). Publication bias can be divided into two types, which is be outlined in the following section.

**Types of publication bias.** One type of publication bias is the “file drawer problem,” in which the researcher self-selects out of submitting a study for peer review or publication because of non-significant findings (Rosenthal, 1979). There are many reasons researchers self-select out. For example, Franco, Malhotra, and Simonovits (2014) categorized 26 detailed e-mail responses to understand why researchers’ self-select out. Emails were from researchers whose studies found non-significant results and who did not write a paper. Franco and colleagues found 15 researchers indicating they did not complete the project because they “believed that null results have no publication potential, even if they found the results interesting personally” (p. 1504), nine authors reducing the priority of the non-significant results and focusing more on other research, and two authors eventually publishing papers in support of their original hypothesis using different samples. This suggests that if a study shows non-significant findings, researchers will sometimes not write up the results because they believe the study is not publishable.

The second type of bias contributes to the first. This publication bias occurs when research is submitted, but is rejected because it lacks statistically significant findings and
therefore is thought of as having “no value,” being “insignificant,” or possibly being “invalid” (Greenwald, 1975). Publishing both significant and non-significant findings is important, as it will develop more robust sets of theories, creating a stronger science (Landis, James, Lance, Pierce, & Rogelberg, 2014). Franco et al. (2014) demonstrated the prevalence of this publication bias when they examined 249 Time-sharing Experiments in the Social Sciences studies to evaluate what percent of research published had significant results, mixed results, or null findings. In their sample, 60% of studies with significant results and 50% with mixed results were published, but only 20% with null results were published (Franco et al., 2014). This perpetuates the file-drawer type of publication bias, because there is statistical evidence non-significant research is less likely to be published.

Ferguson and Brannick (2011) found further support that publication bias is of great concern. They examined 91 meta-analyses for potential bias and evaluated if authors took any precautions to eliminate/reduce it. They found potential publication bias occurred in 41% of research studies, and in 25% of those, the authors describe the level of bias as “worrisome” (p. 3), though they did not operationally define what “worrisome” meant. Given the extent to which publication bias is possibly present in the literature, its potential consequences must have thorough consideration, further supporting the need for an assessment that measures the perceived trustworthiness of the literature.

**Publication bias consequences.** Publication bias has numerous consequences. Failure to report or publish null results can cause researchers to explore areas or relationships that others have found to be non-significant, thus wasting time, energy, and
resources (Franco et al., 2014). Research often begins with examining the current literature and determining which elements need further study. For example, if counterproductive work behaviors (CWBs) have been extensively researched with social cynicism, then researchers will look for a new perspective on CWBs, maybe by introducing a new predictor, such as optimism. If other researchers have already studied CWBs and optimism together and found non-significant results each time, but did not publish these results because of the existing bias against non-significant results, then conducting the same studies would waste both time and resources that could have been better spent had those non-significant findings been reported. If researchers lack access to both significant and non-significant research on a topic, they cannot know whether they are conducting research that will add to their scientific field. This is not to say that all non-significant research should be published; however, studies with a solid rationale and methodology but that find non-significant results should also have an equal chance at publication as studies with solid rationale and significant results. Studies with poor or under-developed rationales should rightly continue to be rejected based on their lack of a solid conceptual foundation for the research conducted.

Consequences of publication bias go beyond just wasting time and resources; it also skews the literature and increases the difficulty to correct Type I errors (Kepes & McDaniel; 2013; Mullins et al., 2015; Simmons, Nelson, & Simonsohn, 2012). If journals only publish significant findings, then scientists should be concerned that the published literature is not representative of all obtained results (Kepes & McDaniel, 2013). This can affect research for years, because it keeps science from self-correcting
the Type I errors that will occur, which when alpha is set at .05, is five percent of the time.

Replication work also helps identify errors and is important to further understand unique findings that dispute theories or when there are complex findings such as a three-way interaction (Leung, 2011). Failing to conduct, or publish, replication research also makes it more difficult to correct Type I errors that may be in the literature (Mullins et al., 2015). Furthermore, because the research literature represents a cumulation of knowledge, if the research is biased or has Type I errors to start, it will only continue to become more biased and error-filled as results cumulate (Simmons et al., 2012). This becomes a cyclical concern that will continue to perpetuate itself unless non-significant results are published as often as significant results.

Replication work is critical for science beyond correcting potential Type I errors. As noted by Koole and Lakens (2012), if researchers replicate others’ work, it improves the quality and reputation of their field by demonstrating which results are reliable. However, they noted that “individually, researchers are better off by conducting only original research, because this will typically yield more publications and citations, and thus ultimately greater rewards in terms of better jobs and more grant money” (Koole & Lakens, 2012, p. 610). In other words, researchers should be conducting and publishing replication work; however, because replication work is rarely published because of publication bias and the low citations replications generate, replication work is generally not conducted (Koole & Lakens, 2012). That predictably low number of citations has a negative effect on the publishing journal’s impact factor, which is “the average number of times an article published in the two preceding years was cited;” this number is taken to
reflect the level of prestige or importance of that journal (Zupanc, 2014, p 113). If
publishing replication studies predictably results in fewer citations (Koole & Lakens,
2012), and if fewer citations lower a journal’s impact factor (Zupanc, 2014), this reflects
a systemic problem that actively discourages the publishing of replication research.

As concerning as publication bias is, the problem is more of a concern when
considering the broader context of research publication. If researchers and journals
consistently favor/publish studies with statistically significant results, then one potential
(and, indeed, documented; see O’Boyle, Banks, & Gonzalez-Mulé, 2014) outcome of
such a bias is the adoption of what the literature has come to refer to as “questionable
research practices.” These practices need to be examined in conjunction with publication
bias to further assess the damage done to the trustworthiness of the research literature.

**Questionable Research Practices**

Kepes and McDaniel (2013) suggested publication bias may exist, in part,
because articles with significant findings tend be seen as more interesting by readers, thus
promoting readership and attention, which can enhance a journal’s reputation. That is,
publishing significant findings may have started because of readership interest, which
may have contributed in the greater issue of publication bias occurring today.

Additionally, O’Boyle et al. (2014) suggested that publications in top-tier journals are
“the currency in which our field trades” (p. 2), and are tied to critical outcomes, such as
salary and tenure for academics. As such, the system has created a reward structure that
relies on publications, creating an intense pressure for researchers to publish. These
pressures may push researchers to engage in questionable research practices (QRPs) to
increase their chance for publication (O’Boyle et al., 2014). Furthermore, a graduate
program’s rank is often contingent upon the number of publications, only increasing the “publish or perish” system for universities (Kepes & McDaniel 2013). This creates more pressure for academics to publish, because not only does their academic prestige rely on their publications, but so does their job. Therefore, research has suggested that researchers may engage in questionable practices to increase the likelihood of publication in order to meet the demands placed on their career from both their university and their field as a whole (O’Boyle et al., 2014). There are several types of QRPs, with HARKing and P-hacking having received substantial attention.

**HARKing.** One type of QRP is HARKing, or Hypothesizing After the Results are Known (Kerr, 1998). This is also sometimes described as “presenting *post hoc* hypotheses as *a priori*” (PPHA; Leung, 2011). Researchers who HARK make decisions about what to report and how to report it after the data are analyzed (Kerr, 1998). This is to downplay, or remove, any non-significant results, and to highlight significant but non-hypothesized findings.

There are multiple types and levels of HARKing outlined by Kerr (1998). “Pure HARKing” (Version 1) is when the researcher reports only hypotheses that are most consistent with the results obtained. Version 2 is “Pure HARKing + Straw Man,” in which the researcher has one or more post hoc hypotheses contradict the data to create the illusion of a competitive hypothesis test. In Version 3, “Suppress Loser Hypotheses,” researchers write hypotheses to be plausible as both a priori and post hoc and do not report non-significant hypotheses. Researches that engage in “HARKing using Post Hoc Plausibility + Necessity of Anticipation” (Version 4) expand a set of hypotheses, suppress non-significant, or as Kerr referred to them as, “loser hypotheses,” and frame
the hypotheses such that they could be anticipated and/or judged a priori. Lastly, Version 5, “Empirical Inspiration HARKing,” uses both a priori and post-hoc hypotheses that seem plausible based on the results obtained. Although Kerr outlined these types, he also noted HARKing is not mutually exclusive. Researchers can engage in variations of each type. There are also differences in each type’s potential influence on research. For example, Version 1 is a more extreme variation of Version 4, because in Version 1, researchers explicitly select what is and is not reported, whereas in Version 4, researchers expand their a priori hypotheses. Therefore, those who engage in “Pure HARKing” might been seen as less trustworthy than those who engage in “Post Hoc Plausibility + Necessity of Anticipation HARKing.”

The most notable ethical issue with HARKing is that it “violates a fundamental ethical principle of science: the obligation to communicate one’s work honestly and completely” (Kerr, 1998, p. 209). This can affect how readers use research because unless the author clearly articulates what is *a priori* and *post hoc*, detecting these behaviors is difficult, if not impossible. Not being able to detect if the presentation of research results has been manipulated can have implications for how other scientists trust and apply research. Being aware that some researchers HARK, but unable to distinguish between those who do and those who do not, may lead to a distrust of all research. Understanding what that potential distrust may affect is critical, and requires the development of a robust trustworthiness scale.

*P*-hacking. *P*-hacking is another questionable practice that might affect the perceived trustworthiness of the literature. Simmons, Nelson, and Simonsohn (2011) suggested this practice is done using “researcher degrees of freedom” (p. 1). Researcher
degrees of freedom represent the decisions researchers make throughout the research process that potentially influence whether or not they obtain statistical significance; such decisions include which variables are controlled, which are omitted or reported, when to stop collecting data, when to collect more, and so on. P-hacking involves consciously exploiting these decisions to obtain statistical significance. Scientists generally agree that it is difficult, or in some cases impossible, to decide everything before the study begins, but concern arises when researchers explore statistical alternatives, searching for “statistical significance” and only report what was significant (Simmons et al., 2011). The issue is one of Type I error, in which the likelihood of at least one analysis producing a false positive finding increases with the number of analyses conducted. When many analyses are run, an \textit{a priori} alpha level of .05 results in a Type I error rate much greater than 5%. Simmons et al. (2012) suggested a simple solution to identify if any of these practices occurred: report everything. If sample size was determined beforehand, researchers should state that. If variables are removed, researchers should note that as well. By reporting what was done, readers can fully understand and replicate the process, if they so choose.

Outlined above were just some of the consequences of publication bias. It is possible there are many unknown consequences of publication bias that will affect our field over the coming years. Kepes and McDaniel (2013) suggested we start evaluating the trustworthiness of the I-O psychology literature to understand what effects publication bias and other research behaviors have had on the credibility of the research literature. Mullins et al. (2015) examined perceived trustworthiness with respect to publication bias and found that graduate students were concerned about publication bias
and also reported being more likely to apply literature findings in a broad manner once in their career fields. However, conducting more research assessing how publication bias has affected the perceived trustworthiness of the literature and assessments about these practices must be conducted to begin to fully understand the implications.

**Trustworthiness**

Questions of perceived trustworthiness of a literature must be asked in the context of that literature’s readership. Mohammadi, Thelwall, Haustien, and Lariviere (2014) examined readership of scientific journals. Their results suggested readers of Clinical Medicine, Engineering and Technology, Social Science, Physics, and Chemistry literature mostly were Ph.D. students, with postgraduate students and postdoctoral researchers being the second most common. These groups are also most likely to “mine the research” to gain extensive knowledge on a topic. The question of how graduate students perceive the trustworthiness of the publication process and research journals is probably most critical because this population is likely new to the publication process and to the streams of literature presented in research journals, and graduate students are the future practitioners and academics, so understanding what affects their views and uses of the literature is important. Additionally, because graduate students are the future of their domains, it is possible to begin implementing new approaches and methods to reduce publication bias and QRPs through education and awareness.

**Measuring perceived trustworthiness.** Mullins et al. (2015) developed an exploratory scale to assess graduate students’ perceptions of trust in the literature. They found that graduate student distrust was mainly within the publication process, and that this distrust could lead to students using the literature in a more broad and generalized
manner, once they obtained their degree. In other words, rather than using explicit research findings, they reported they were more likely to use the research as an overall summary tool for solutions, meaning that rather than focusing on the research method used to obtain the results and the results found, graduate students focused on what the research found overall, regardless of the methods used. This finding is of particular concern because general summaries of research will not be as effective in solving problems as understanding and using specific findings and theories. Mullins et al. also suggested that the training in scientific principles and rigor is a primary benefit to hiring I-O psychologists, and if we begin to focus on the “overall message” rather than the specific details, this value is lost. This benefit extends to other programs and degrees as well, and is at risk of being lost if graduate students begin to only evaluate research for general messages.

Although Mullins et al.’s (2015) study showed interesting and promising results, there are a few concerns. Their research was mainly exploratory, with the goal of increasing the awareness and conversation about publication bias, and as such, it used a sample of 23 respondents from a single graduate program. As noted by the authors, the generalizability and power of their sample is questionable, and further research needs to use a variety of programs. For example, because publication bias is not limited to I-O psychology, collecting data from a variety of fields such as clinical, cognitive, social, and personality psychology is critical understanding the impact and severity of this problem. Additionally, because M.A./M.S., Psy.D., and Ph.D. programs teach using research journals, extending the program selection to include not only different areas of concentration but different program models will help to understand the effects
publication bias may have more broadly. The items developed in the present study are unique to this study; however, items drew inspiration from Mullins and colleagues. However, this study used a larger and more diverse population, as well as introduced more items to develop a more robust scale. Doing such helps educators better understand the impact publication bias has on graduate student trust and use of research, which can inform not only education, but also how the research literature may be used by practicing psychologists once graduate school is complete.

**Validation constructs.** When developing a scale, it is critical to include other validated scales to establish convergent and discriminant validity. However, because the area of publication bias and QRPs is relatively new, Mullins et al.’s (2015) scale is the only measure available. This is not directly useful as the goal of this study is to develop a more robust scale using Mullins et al.’s measure as a starting point. Additionally, the research is lacking on what theoretically should and should not be related to perceived trustworthiness of the research literature. Therefore, the following section explore potentially important theoretical areas that could provide convergent validity evidence. This study does not explore constructs that could provide potential discriminant validity because there is little research available on perceived trustworthiness. That is, there are no theoretical findings in prior research to guide tests of discriminant validity to further validate this scale.

**Distrust.** It makes sense to expect that a person’s general level of distrust should relate to their level of trust, or distrust, of specific domains. Distrustful people have a tendency to take a defensive stance on things in general (Lewicki, McAllister, & Bies, 1998). If people are generally distrustful, then it stands to reason they might naturally be
distrustful of the research literature, so to establish validity, including a distrust scale from the International Personality Item Pool (IPIP) is be beneficial (Goldberg et al., 2006). Because of the tendencies a distrustful person has, a person who scores high on the perceived trustworthiness scale should score low on the distrust scale.

**Social cynicism.** Social cynicism is the belief that institutions, people, etc. cannot be trusted (Leung et al., 2002). Social cynics have a negative view of human nature, have biased views towards groups, and generally are mistrustful. People who are social cynics would be more likely to distrust the literature and publication process because of their personality. This makes social cynicism relevant for validation purposes because the perceived trustworthiness scale should assess the processes used to publish and the institutions that publish that work. Social cynicism is more appropriate than general cynicism because understanding how respondents perceive institutions overall is more critical than their general cynicism towards people.

**Negative affectivity (NA).** Research has shown those high on NA are likely to be upset and have a negative self-view (Watson & Clark, 1984). Additionally, people with high NA tend to magnify mistakes, disappointments, and threats. Those high on NA might reasonably be expected to relate to lower levels of trust in the literature, since any perceived mistakes will be magnified by the reader’s level of NA, as will any disappointment resulting from violations of what has been taught as “proper research procedure.” Participants who score higher on the perceived trustworthiness scale should score lower on the NA scale, if the perceived trustworthiness scale is valid.

**Conclusion**
Overall, this study aims to develop a scale to assess the impact publication bias and questionable research practices have on graduate students’ trust in the research literature using issues outlined in previous research and conference topics (Kepes & McDaniel 2013; Kerr, 1998; Mullins et al., 2015; O’Boyle et al., 2014; Simmons et al., 2011; Simmons et al., 2012). It is critical to understand the underlying implications these practices have on research as a whole, and in specific areas. This study uses several personality measures, such as distrust, social cynicism, and negative affectivity, to establish convergent validity of the perceived trustworthiness scale being developed.
Chapter II

Rationale and Hypotheses

Publication bias crosses domains of research and science, and the literature outlined previously highlights some of the major issues and consequences faced because of publication bias and the related QRPs. Most notable is the need for research to examine what these practices have done to the perceived trustworthiness of science. Specifically, Kepes and McDaniel (2013) and Mullins et al. (2015) argued that additional research is critical to fully assess the impact these behaviors have on the perceived trust in the research literature. Kepes and McDaniel made the first call to action to examine these behaviors, and Mullins et al. took an initial step towards understanding some potential consequences of publication bias. However, although these studies have allowed for progress, further research to explore additional, new relationships not previously examined is needed. The scale developed in this study addresses the original concerns outlined by Kepes and McDaniel and Mullins and colleagues.

As such, there are several relationships worth exploring. Although some of these relationships were examined by Mullins et al. (2015), it is important to conduct replication work to assess these relationships with a broader sample. This provides further research to potentially address the concerns outlined by Mullins and colleagues. These relationships could also indicate the impact publication bias practices have, as well
as the validity of the scale developed in this study. New relationships, not explored by Mullins et al., are also worth examining and is outlined in this chapter as well.

A relationship between awareness of publication bias and graduate students trust in the literature should exist because as people become aware of biases and questionable practices that can be used to enhance the likelihood of publication, they may view the publication process, and the results of that process, as less trustworthy (Kepes & McDaniel, 2013). If respondents are aware of specific behaviors researchers use to boost the chances of publication, and are aware journals have a bias towards one particular type of statistical result, then they may be less likely to trust in the research process. As such, the following is hypothesized.

**Hypothesis 1:** There will be a negative correlation between graduate student awareness of publication bias and their level of trust in the literature.

Mullins et al. (2015) found that students who distrusted the literature were more likely to plan to use research in future contexts in a broad, generalized sense. That is, rather than focusing on underlying scientific principles and methods presented in the research literature, they indicated they would extract general messages without regard to the methodology behind how they were obtained. This finding is potentially concerning if replicated because, as noted by Mullins et al., if students are only applying general information from their training, the added value of hiring psychologists with extensive research-based training is lost. If this result is replicated in a larger and more diverse sample, it may indicate that the future application of scientific principles could be fundamentally changed by publication bias and QRPs. As such, the following hypothesis is presented.
Hypothesis 2: There will be a negative correlation between scores on the perceived literature trustworthiness scale and self-reported interest in general take-aways from research articles.

Validation Constructs

Graduate students who are more generally distrustful may also have a higher distrust of the research literature, and would therefore score lower on the perceived literature trustworthiness scale. Specifically, those who are distrustful have a tendency to take a protective action based on some assumed outcome; in other words, those who are distrustful are more likely to take a negative stance on things in general (Lewicki et al., 1998). Parallel logic should follow for a tendency toward social cynicism. Specifically, those respondents who are higher on social cynicism in general may be less trusting of the research literature based on the presence of publication bias and questionable research practices. This is reasonable, because many of the elements of social cynics include a negative view of human nature, a biased view towards groups of people, and generally mistrust others’ intentions (Leung et al., 2002). Finally, those high on negative affectivity may also be less trusting of the research literature, as those high on NA are likely to have more negative emotions, be more distrustful, and focus on negative events (Watson & Clark, 1984). As such, the following relationships are hypothesized.

Hypothesis 3: There will be a negative relationship between perceived trustworthiness of the literature and a general tendency to be distrustful.

Hypothesis 4: There will be a negative relationship between perceived trustworthiness of the literature and social cynicism.
Hypothesis 5: There will be a negative relationship between perceived trustworthiness of the literature and negative affectivity.

Exploratory Analyses

Additional exploratory relationships were be examined following data collection, including differences between graduate students’ academic program type (M.A./M.S. vs Psy.D./Ph.D.) and their level of perceived trust in the research literature. Although there is no prior research to drive the theory for this relationship, I posit those in longer, more research-focused programs (e.g., a Ph.D. program) might be less trusting of the literature because they have more exposure to the research and publication process, and potentially more direct experience with it as an author. This latter statement is also explicitly considered in further exploratory analyses.

Another potential exploratory relationship to examine is how familiar a graduate student is with the research and publication process and how trusting they are in the literature. For example, students who have submitted research for publication might be less trusting in the literature if they have directly been rejected for non-significant findings than students that have never submitted research for publication. Exploring if the publication process has any effect on graduate students’ perceived trust in the literature is important because they are the future of their fields, and if they are distrustful of the process and the system, the overall consequences might be greater than researchers first assessed.
Chapter III

Method

Participants

Participants were from M.A., M.S., Psy.D., and Ph.D. Programs in a variety of psychology programs, such as Cognitive, Human Factors, Clinical, and Social (for a complete list, see the demographic items in Appendix A). The sample was not limited to departments with I-O psychology programs because examining a diverse sample of departments provided a wider understanding of the impact of publication bias and QRPs and increased generalizability.

There are differing thoughts about proper sample sizes needed for a factor analysis. Spector (2002) stated that for an exploratory factor analysis, 100 to 200 participants are needed. When Costello and Osborne (2005) conducted a meta-analysis on current factor analysis practices, they found that the majority of studies examined used a subject to item ratio of 10:1. However, they also found that ratios ranged from 2:1 or less, all the way to 100:1. Overall, there is little agreement in practice and in theory about the required sample size for a factor analysis. Because exploratory factor analysis played a key role in data analysis, initially it was decided to use a minimum sample size of at least useable 200 participants (Maccallum, Widaman, Zhang, & Hong, 1999; Spector, 2002). However, collecting 200 participants was challenging and after three months of data collection, a minimum sample size of at least 120 participants was agreed-
upon by the thesis committee and myself. This sample size should ensure adequate power to detect medium effect sizes for correlational research questions (Cohen, 1992). Participants were anonymous during the study. The end of the study contained a link to an external survey where participants entered their contact information for a chance to win a $20 Amazon gift card for participation (Appendix B). Participant demographics are presented in Table 1.

Data were inspected for completion and quality before analyses were conducted. Removed first were cases that answered less than 50% of the survey items. This resulted in 52 participants removed. Next, participants were removed if they failed to answer entire scales in the study; 2 participants were removed. Questions about program type, field, and level were examined as the study was restricted to current graduate students in psychology programs. One response stated the respondent was an undergraduate student and was removed. Data were then examined for variability, to determine if respondents simply clicked through the study using the same response answer. No responses were removed on this basis. The final sample contained 123 responses.

Once the data were cleaned, the remaining responses were examined for formatting consistency. For example, the question “I have not submitted work for publication ever” was an open-ended response and no specific response format was required. As such, responses were in words and numbers. All responses were converted to 1 for yes and 0 for no. When ranges were provided, the average of the range was taken, for example, when asked “on average, how many research journals do you read for graduate school (GA/TA position, classes, Thesis/Dissertation) a week,” if the response was “2-6,” the average (in this case 4) was calculated and used.
Table 1.

Sample Demographics

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-28</td>
<td>101</td>
<td>84</td>
</tr>
<tr>
<td>29-39</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>40-50+</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>46</td>
<td>37.4</td>
</tr>
<tr>
<td>Female</td>
<td>75</td>
<td>61</td>
</tr>
<tr>
<td>Prefer Not to Respond</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian or Asian American</td>
<td>5</td>
<td>4.1</td>
</tr>
<tr>
<td>Black or African American</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>4</td>
<td>3.3</td>
</tr>
<tr>
<td>White (Non-Hispanic)</td>
<td>106</td>
<td>86.2</td>
</tr>
<tr>
<td>Multi-racial</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Prefer not to respond</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td>Missing</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td><strong>Program Field</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical</td>
<td>22</td>
<td>18</td>
</tr>
<tr>
<td>Cognitive</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Counseling</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Developmental</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Ecological</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Educational</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Experimental</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Human Factors</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Industrial-Organizational</td>
<td>50</td>
<td>41</td>
</tr>
<tr>
<td>Personality</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Social</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Other</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td><strong>Program Type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M.A.</td>
<td>27</td>
<td>22.3</td>
</tr>
<tr>
<td>M.S.</td>
<td>18</td>
<td>14.6</td>
</tr>
<tr>
<td>Psy.D.</td>
<td>9</td>
<td>7.3</td>
</tr>
<tr>
<td>Ph.D.</td>
<td>65</td>
<td>53.7</td>
</tr>
<tr>
<td>Other (Education Specialist)</td>
<td>2</td>
<td>1.7</td>
</tr>
<tr>
<td><strong>Years in Program</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 - 2 Years</td>
<td>87</td>
<td>70</td>
</tr>
<tr>
<td>3 - 4 Years</td>
<td>22</td>
<td>18</td>
</tr>
<tr>
<td>5 - 6 Years</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>7 - 8 Years</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Measures

**Perceived trustworthiness.** A questionnaire designed for the present study to assess the perceived trustworthiness of the research literature and the publication process is included as Appendix C. Participants responded on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). Based on a literature review, some potential items were as follows: “I perceive the research literature as ethical,” “I trust the results in research journals,” and “Authors always report exactly what occurs in their studies.” Internal consistency reliability was assessed for sets of items designed to be part of the same sub-scales, and is reported as part of the study results. These subscales were: Publication Bias, Questionable Research Practices, Statistical Significance, and Trust. The Publication Bias Subscale items focused specifically on practices within the publication process; an example item from the subscale before analysis was, “It is important as a field to stop publication bias.” The Questionable Research Practices Subscale items examined potentially ethically questionable practices that occur in the sciences. An example item in the pre-analysis scale was, “Altering findings to increase chances of publication is unethical.” The subsection on statistical significance included items reflecting issues with obtaining and publishing statistically significant and non-significant results. An example item from the subscale before analysis was, “Researchers view non-significant results as a form of personal failure.” Lastly, the Trust Subscale included items that specifically tapped into perceived trust of the research literature. An example item from the subscale before analysis was, “The research literature is reliable.” Due to the nature of the present study, reliabilities for the sub-scales are reported...
following a description of the factor analysis and item analysis results leading to the
selection of the final items to include as part of the study’s Results section.

**Distrust.** The distrust scale from the International Personality Item Pool (IPIP)
was utilized. This scale has 10 items, with an example item being “I often suspect hidden
motives in others” (Goldberg et al., 2006). The scale is similar to the items measured in
Tellegen’s Multidimensional Personality Questionnaire (MPQ). Participants responded
on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The scale has
a previously-reported Cronbach’s alpha reliability of .83 (International Personality Item
Pool). For the scale’s source information, see Appendix D. The Cronbach’s alpha for
this scale in the current study was .88.

**Social cynicism.** Social cynicism was measured using a subscale from the Social
Axioms Scale (SAS), which has 13 items (Leung et al., 2002). Respondents were to rate
each item on a 5-point scale, indicating the degree to which they believe each statement
to be true, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). An example item is,
“If one belongs to a marginal group, it is difficult to gain acceptance from the majority
group” (Leung et al., 2002). The Cronbach’s alpha for a US based sample is .63, but
alphas were .79, .73, and .78 for Hong Kong, Venezuela, and Japan, respectively (Leung
et al., 2002). Later, Leung et al. (2012) reported a Cronbach’s alpha of .79 for a US
sample. Source information for the SAS scale is included as Appendix E. The
Cronbach’s alpha for this scale in the current study was .75.

**Negative affectivity.** Negative affectivity was measured with a 5-item short form
scale from Thompson (2007). Respondents were given the following instructions:
“Thinking about yourself and how you normally feel, to what extent do you generally
Items were answered on a 5-point scale ranging from 1 (never) to 5 (always). An exam item is “upset.” Prior research demonstrated Cronbach’s alpha of .74 (Thompson, 2007). The negative affectivity scale’s source information is included as Appendix F. The Cronbach’s alpha for this scale in the current study was .70.

**Hypothesis 2 measurement.** Hypothesis 2 was intended to be measured with the following question in the demographics section: “In considering the research literature, I am more interested in general take-aways than specific methodologies (e.g., summarizing the results without focusing heavily on the method used).” Responses were to be rated on a 1 (strongly disagree) to 5, (strongly agree) scale. However, due to an oversight in the construction of the online survey, this item was left out. Therefore, it was not possible to test this hypothesis.

**Demographics.** The demographic information collected included information on age, sex, race, program type and field, number of years in the program, publication experience, and familiarity with research journals. Demographic questions are presented in Appendix A. Demographic findings are presented in Table 1. This was used to evaluate possible background-based differences (e.g., program type/field, publication familiarity) in graduate students’ level of perceived trust in the research literature.

**Procedure**

Because publication bias and QRPs are not limited to any specific field of psychology, examining a range of degree levels and fields was important. The first sampling strategy was conducted as follows. A list of 25 randomly-selected psychology graduate departments in the United States was developed, where the last 5 programs were backup schools if more participants were needed. To develop the list, the *American
Psychological Association’s 2015 Graduate Study in Psychology text served as the graduate program database. From there, a random number generator was used to select 25 page numbers that corresponded with a graduate program in the book. Generally, there was only one program per page, but if there was more than one on that page, a second random number was generated to identify the program for the list. For example, if page 155 was generated, and it had 2 programs, then another number ranging 1-2 was generated to select the program.

Once all programs were selected and Xavier University’s IRB approved (see Appendix G for the initial IRB approval letter), if it was possible to identify the professors who teach a research methods course, then they were contacted via email. If it was not possible, then the chair was contacted. This method was selected because department chairs probably receive a number of research participation request emails, so to increase forwarding potential contacting the research methods professor was viewed as preferable. Each department contact received an email informing them of the study’s purpose and asking if they were willing to forward the study’s information and directions (including the link) on to their graduate students (see Appendix H). If the department’s contact agreed, the graduate students received the survey and instructions with a requested date of completion to aid in timely data collection (see Appendix H). Prior to responding to any questions, participants read and indicated their agreement with an informed consent document (see Appendix I). Participants then completed the Perceived Trustworthiness items, the distrust scale, the social cynicism scale, the NA scale, and the demographic questions. Upon completion of the study, participants were debriefed (see Appendix J). Once they read the debrief, they had the option of entering their email
address in an external survey to win one of five $20 Amazon Gift Cards (see Appendix B).

After three weeks of data collection, only 27 usable responses were collected. With such a slow response rate, the initial IRB protocol was resubmitted with modifications. The first modification was to allow other faculty at the university to be contacted if the original faculty members contacted either did not respond or were unavailable (e.g., on sabbatical, no longer at the university). The second modification was to identify and contact an additional 20 programs to increase participation. These programs were selected using the same method as the first 25 programs. Xavier University’s IRB approved these modifications (Appendix K). After approval, the first list of programs were contacted again and the additional 20 schools were contacted. After two weeks of collection, faculty or chairs were contacted again to increase participation. After four weeks of data collection using this sampling strategy, only 42 total useable responses had been collected. Because the original target sample size was 200, the IRB protocol was modified again to increase participation.

This third IRB modification shifted the sampling strategy to a snowball strategy in which I contacted personal connections to either complete the survey (if they met the demographic requirement of a current psychology graduate student) or pass it along to people they knew who fit the demographic (see Appendix L for IRB approval letter and Appendix M for the sampling strategy). The survey was posted on Facebook.com and Reddit.com within psychology-related pages or forums to increase participation. After an additional month of posting, emailing, and sharing the survey, data collection reached
120 useable responses, and the thesis committee agreed that this was sufficient to proceed with data analysis.

During the study, data were secured through Qualtrics, a survey collection website. The highest possible anonymity was selected from the available system selections. Once the study finished, data were secured in Dropbox, a password protected online data storage system, which is accessible through a password-protected laptop and on a password-protected work computer using a different password.
Chapter IV

Results

To analyze the factor structure of the Perceived Trustworthiness Scale, an exploratory principal components analysis (PCA) with varimax rotation to extract the factors was conducted on each of the subscales. To determine the number of factors in each subscale, the rotated component matrix was examined. Factor loadings were decided by examining the rotated component matrix for where the item loaded most highly. For example, if item one in a subscale loaded on factor one at .34, factor two at .24, factor three at .55, and factor four at .87, it was decided the item loaded onto factor four. Once the number of factors was identified and initial item cuts were made based on the PCA, all theoretically-constructed subscales were subjected to an item analysis. Based on the item analysis, if removing an item could increase Cronbach’s alpha, then the item was considered for removal. If an item was removed, the item analysis was re-run (without the removed item) and remaining items were re-examined for possible removal. This process continued until Cronbach’s alpha would either not increase by removing an item, or removing an item resulting in an increase in alpha so small (.01 - .02) that the increase did not make sense in the author’s judgment. That is, if the content of the item fit theoretically with the content of the remaining items, and the items removal would result in a small (.01 - .02) increase in alpha, the item was retained. Once the item analysis was completed, a PCA with varimax rotation was conducted on the final
scale to check for unidimensionality. A more detailed description of the scale analyses follows. The final measure is presented in Appendix N. All hypotheses then were tested using the final Perceived Trustworthiness Scale and Subscales.

**Analyses for Publication Bias Subscale**

The pre-analysis Publication Bias Subscale consisted of items 1-18 in Appendix C. Examining the rotated factor matrix initially indicated 6 possible factors. Factor one had 6 items, factor two had 3 items, factor three had 4 items, factors four and five had 2 items, and factor six had 1 item. Upon examining the content of these items, factors two-six had similar phrases or words (e.g., “results”), which probably resulted in the several potential factors. Only factor one was retained because it contained six items that related to the general theme of publication bias and statistical significance issues, and the largest majority of items loaded onto it. Those items were items 3, 4, and 15-18. Once the factor was identified, an item analysis was conducted to examine which items should be retained for the final subscale. Cronbach’s alpha for items 3, 4, 15-18 was .829. When item 17 was removed, Cronbach’s alpha increased to .859. Removing any additional items would have decreased Cronbach’s alpha. The final publication subscale includes 5 items 3, 4, 15, 16, 18, with a Cronbach’s alpha of .859. Retained items for all sub-scales are presented in Appendix N.

**Analyses for Questionable Research Practices Subscale**

Before analysis, the Questionable Research Practices Subscale contained items 19-50. The rotated factor matrix indicated 11 possible factors. Factor one included 7 items, factor two 5 items, factor three 4 items, factors four and five 2 items each, factors six, seven, and nine 3 items each, and factors eight, ten and 11 had 1 item each. Factors
four, five, eight, and 11 were removed first because they had few items that loaded onto the factor. Items loading on factor six included themes about university publish/perish systems and items loading on factor seven had themes about reporting data. Both factors were removed. Factors one, two, and three were then examined as potential factors because of the number of items that loaded on each and the content of the items. Factor one included items 20, 23, 24, 38, 44, 45, and 49, with common item themes involving research practices, research ethics, teaching about QRPs, and publication bias. Factor two included items 30-32, 42, 46 and involved themes about hypothesis testing and research practices about hypotheses. Factor three included items 25-27 and 29, reflecting item themes such as data manipulation and data reporting.

Each sub-factor was analyzed with an item analysis. Sub-factor one (containing items 20, 23, 24, 38, 44, 45, and 49) had a Cronbach’s alpha of .677. When item 44 was removed Cronbach’s alpha increased to .740. After re-running the item analysis without item 44, removing item 20 increased Cronbach’s alpha to .766. Alpha could not be increased by removing any additional items. The final QRPs scale was finalized with items 23, 24, 38, 45, and 49 (Cronbach’s alpha = .766).

Sub-factor two (containing items 30-32, 42, and 46) had a Cronbach’s alpha of .674. Upon further item analysis, removing item 31 increased Cronbach’s alpha to .687. After re-running the item analysis, removing item 42 would have increased Cronbach’s alpha to .710. Cronbach’s alpha could be increased further by removing item 46.
Examining the possible final scale, Cronbach’s alpha was at the low end of acceptability and only reflected two items (30, 32). Therefore, it was decided that this sub-factor was not a meaningful factor; these items were excluded from the final scale. Sub-factor three
(containing items 25-27 and 29) demonstrated a Cronbach’s alpha of .631. The item analysis indicated removing item 29 would increase Cronbach’s alpha from .627 to .628. Running the reliability analysis again showed that no further increase in alpha was possible. The final alpha for this factor would be .628, which is too low to be acceptable in research or practice, as such, this factor was discarded (Spector, 1992). In summary, the final scale contains only one scale with items 23, 24, 38, 45, and 49 (Cronbach’s alpha = .766).

Interestingly, the final QRPs Subscale items tap into more general ethical issues such as the demand to publish, altering findings to help publication chances, conducting science in an ethical and responsible manner, and teaching ethics to new psychologists, rather than specific items related to QRPs. This subscale contained the most pre-analysis items (31 items) and several of these items specifically related to QRP topics such as P-hacking, HARKing, post hoc hypotheses, and data manipulation. However many of these items were not retained in the final measure. Because of the subscale’s final content, it makes sense to rename this subscale with a more general name such as Research Ethics Subscale (this name will be used going forward).

**Analyses for Statistical Significance Subscale**

Before analysis, the Statistical Significance Subscale contained items 51-62. The rotated factor matrix indicated 4 possible factors. Factor one had 5 items, factor two had 3 items, and factors three and four had 2 items each. Factors two through four had low loadings for most of the items (alphas ranged from .293 - .336) and were removed as possible factors. Items 53, 56, 57, 60, and 61 loaded onto factor one. When an item analysis was done with factor one items, Cronbach’s alpha was .763. An item analysis
was conducted, and item 60 was removed (Cronbach’s alpha = .779). The final scale contains items 53, 56, 57, and 61.

**Analyses for Trust Subscale**

Prior to analysis, the Trust Subscale contained items 63-75. After the PCA with varimax rotation, four possible factors were identified. Factors one and three each had 4 items. Factor two had 3 items and Factor four had 2 items. Only factors one and three were subjected to item analyses because they had equal items load onto each factor. Factor three had a Cronbach’s alpha of .356 and was not explored further as the alpha could only increase to .380, which is not an acceptable alpha for research (Spector, 1992). An item analysis was then conducted on the items that loaded on factor one (items 67, 68, 69, and 73) and showed a Cronbach’s alpha of .798. Upon re-running the item analysis, removing item 73 would increase alpha to .817. Removing any additional items would not increase Cronbach’s alpha. The final scale contains items 67, 68, and 69, with a Cronbach’s alpha of .817.

**Hypothesis Testing**

The hypotheses were to be tested using the total Perceived Trustworthiness Measure. However, once the subscales were factor analyzed, using a total scale score did not make sense. First, only one of the four subscales’ final items assesses trust. The remaining three subscales (Publication Bias, Research Ethics, and Statistical Significance) reflect an evaluation of the extent to which different facets of broader issues (publication bias and research ethics) are perceived to affect overall research literature. Although these three subscales have some theoretical overlap (for example, some statistical significance questions in the Publication Bias Subscale), an exploratory
factor analysis (EFA) with a maximum likelihood rotation was conducted to further assess the factor structure and item loadings of the subscales. The rotated component matrix showed that were four factors. Further examination showed that each item in the subscale loaded only on its respective subscale, suggesting that each subscale is conceptually distinct. That is, items 1 – 5 loaded only on the first factor, items 6 – 10 loaded on factor 2, items 11 – 14 loaded on factor 3, and items 15 – 17 load on only factor four. Therefore, each subscale should be considered theoretically and psychometrically distinct.

Additionally, on the Research Ethics Scale, high scores indicate the perception of a problem, whereas on the Publication Bias and Statistical Significance Subscales, high scores indicate a belief that ethics are important. Therefore, aggregating these subscales into one total measure does not make theoretical sense based on item content and what high scores mean on each subscale. Cronbach’s alpha for the aggregated Perceived Trustworthiness measure was .59, which is too low for acceptable use in research and practice (Spector, 1992) and is almost certainly a reflection of the multidimensional nature of the aggregate measure. To aggregate the subscales (alphas ranging from .76 - .85) into one measure and reduce alpha is psychometrically problematic and creates conceptual ambiguity as to the meaning of the resulting correlations.

Although for the reasons outlined it does not make sense to use an aggregate score for the total measure in future work, the present study originally hypothesized relationships with the total Perceived Trustworthiness Measure, which now will be referred to as PTM. As such, the hypotheses were tested using both the PTM and the most conceptually appropriate subscale. The most conceptually appropriate subscale was
decided by examining the hypothesis and the subscales, then selecting which made theoretical sense. For example, Hypothesis 1 predicted a negative correlation between graduate student awareness of publication bias and their level of trust in the literature. Because the hypothesis should theoretically test trust and awareness, the trust subscale was more theoretically related and was selected for analysis. The correlation matrix is presented in Table 2.
Table 2.

Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>M (SD)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Negative Affectivity</td>
<td>2.48 (0.52)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Social Cynicism</td>
<td>2.61 (0.48)</td>
<td>.278 **</td>
<td>(.745)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Distrust</td>
<td>2.35 (0.60)</td>
<td>.564 **</td>
<td>.371 **</td>
<td>(.888)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Perceived Trustworthiness</td>
<td>4.33 (0.25)</td>
<td>-.095</td>
<td>-.122</td>
<td>-.148</td>
<td>(.595)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Publication Bias</td>
<td>4.34 (0.57)</td>
<td>-.052</td>
<td>-.187 *</td>
<td>-.109</td>
<td>.647 **</td>
<td>(.859)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Research Ethics</td>
<td>4.72 (0.29)</td>
<td>.027</td>
<td>-.116</td>
<td>-.035</td>
<td>.666 **</td>
<td>.424 **</td>
<td>(.766)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Trust</td>
<td>3.50 (0.68)</td>
<td>-.011</td>
<td>-.072</td>
<td>.026</td>
<td>.094</td>
<td>-.357 **</td>
<td>-.066</td>
<td>(.817)</td>
<td></td>
</tr>
<tr>
<td>8 Statistical Significance</td>
<td>4.24 (0.60)</td>
<td>-.112</td>
<td>.135</td>
<td>-.131</td>
<td>.552 **</td>
<td>.115</td>
<td>.160</td>
<td>-.258 **</td>
<td>(.779)</td>
</tr>
</tbody>
</table>

** = correlation is significant at the .01 level. * = correlation is significant at the .05 level.

Note: Numbers in parentheses are coefficient α reliabilities.
Hypothesis 1 predicted a negative correlation between graduate student awareness of publication bias and their level of trust in the literature and was tested using the Publication Bias Subscale (final items 1-5) and the Trust Subscale (final items 16-19). This hypothesis was supported, $r(119) = -.357, p < .001$.

Hypothesis 2 predicted a negative correlation between scores on the PTM and reported intent to use the research for general purposes in the future was not tested. Due to an error during survey creation, the question intended to test this hypothesis was not included in the final survey.

Hypothesis 3 predicted a negative correlation between perceived trustworthiness of the literature and a general tendency to be distrustful. When tested with the PTM and the Distrust Scale, the hypothesis was not supported, $r(199) = -.143, p = .115$. When tested with the Trust Subscale items and the distrust scale the hypothesis was also not supported, $r(119) = .030, p = .744$. Hypothesis 4 predicted a negative correlation between perceived trustworthiness of the literature and social cynicism. When tested with the PTM and cynicism scales, the hypothesis was not supported ($r(119) = -.122, p = .183$). When tested with items from the Trust Subscale and the social cynicism scale, the hypothesis was also not supported, ($r(119) = -.072, p = .433$). Lastly, Hypothesis 5 predicted a negative correlation between perceived trustworthiness of the literature and negative affectivity. When tested with the PTM and the negative affectivity scale, the hypothesis was not supported, ($r(119) = -.095, p = .299$). When tested with the Trust Subscale and the negative affectivity scale, the hypothesis was not supported ($r(119) = -.011, p = .909$).
Exploratory Analyses

Exploring if the publication process has any effect on graduate students’ perceived trust in the literature is important because they are the future of their fields, and if they are distrustful of the process and the system, the overall consequences might be greater than researchers anticipate. As such, additional exploratory relationships were examined. The first exploratory relationship examined differences between graduate students’ academic program type (M.A., M.S., Psy.D., Ph.D.) and their level of perceived trust in the research literature using the entire PTM. Although there is no prior research to drive the theory for this relationship, I posited that students in longer, more research-focused programs (e.g., a Ph.D. program) might be less trusting of the literature. This is because they are more likely to be exposed to the research and publication process as their programs are longer and potentially expect graduate students to have experience with publication process as an author.

To test this, a one-way analysis of variance (ANOVA) was conducted using the participant’s program type and all of the subscales. All means, standard deviations, and F-tests are reported in Table 3. The Trust Subscale was tested first. There was not a significant difference across program types, $F(3, 117) = .508, p = .677$. A one-way analysis of variance (ANOVA) was then conducted using program type (M.A./M.S./Psy.D./ Ph.D.) as an independent variable (IV) and the Publication Bias Subscale as the dependent variable (DV). There again was a not significant difference based on program type, $F(3, 117) = .511, p = .676$. 
Table 3.

ANOVA Testing Program Type Differences Across PTM Subscales

<table>
<thead>
<tr>
<th>Scale</th>
<th>Program</th>
<th>M, SD</th>
<th>F-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trust</td>
<td>M.A.</td>
<td>3.61, .72</td>
<td>$F(3, 117) = .508, p = .677$</td>
</tr>
<tr>
<td></td>
<td>M.S.</td>
<td>3.44, .68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Psy.D.</td>
<td>3.62, .35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ph.D.</td>
<td>3.45, .68</td>
<td></td>
</tr>
<tr>
<td>Research Ethics</td>
<td>M.A.</td>
<td>4.76, .04</td>
<td>$F(3, 117) = 3.71, p = .014^*$</td>
</tr>
<tr>
<td></td>
<td>M.S.</td>
<td>4.55, .10*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Psy.D.</td>
<td>4.60, .08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ph.D.</td>
<td>4.77, .26*</td>
<td></td>
</tr>
<tr>
<td>Statistical Significance</td>
<td>M.A.</td>
<td>4.12, .64</td>
<td>$F(3, 117) = 3.61, p = .015^*$</td>
</tr>
<tr>
<td></td>
<td>M.S.</td>
<td>3.93, .56*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Psy.D.</td>
<td>4.19, .41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ph.D.</td>
<td>4.39, .59*</td>
<td></td>
</tr>
<tr>
<td>Publication Bias</td>
<td>M.A.</td>
<td>4.41, .52</td>
<td>$F(3, 117) = .511, p = .676$</td>
</tr>
<tr>
<td></td>
<td>M.S.</td>
<td>4.23, .52</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Psy.D.</td>
<td>4.29, .38</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ph.D.</td>
<td>4.39, .57</td>
<td></td>
</tr>
</tbody>
</table>

* Indicates Significant Finding
Significant differences were found, however, when ANOVAs were conducted on the Statistical Significance and Research Ethics Subscales. A Tukey Honestly Significant Difference (HSD) post hoc test was conducted to explore the differences between group means for the Statistical Significance Subscale. There was a significant mean difference between M.S. students and Ph.D. students, suggesting that Ph.D. students were more concerned with the focus on obtaining statistical significance in research than M.S. students. A Tukey HSD post hoc test for the Research Ethics Scale indicated a significant mean difference between M.S. students and Ph.D. students, suggesting that Ph.D. students were more concerned with ethics and Research Ethics.

Another exploratory analysis examined whether a difference in trust in the literature might exist as a function of participation in the publication process (operationalized as having submitted work for publication). It is possible that students who have experience submitting research for publication would be less trusting in the literature if they were rejected for non-significant findings, than students who had never submitted research for publication. An independent samples $t$-test was used to analyze this, however, no significant difference was found ($t(119) = 1.87, p = .064$). This suggests there is no difference in trust between those who had participated ($M = 3.53, SD = .73$) and had not participated in the publication process (defined as respondents who had experience with the publication process), $M = 3.46, SD = 0.61$. 
Chapter V

Discussion

The primary objective of this study was to develop a scale to assess graduate students’ perceived trustworthiness of the research literature. Using theoretical arguments about the damaging effects publication bias and QRPs could have and one exploratory article about graduate students’ perceived trust in the research literature, this study generated potential items to develop a more comprehensive scale. Once items were constructed and organized, four possible subscales were identified (Publication Bias, Research Ethics, Statistical Significance, and Trust). The PCA and item analysis of each of these subscales identified a unidimensional set of items for each subscale with an acceptable alpha level (.76-.85). An EFA was then conducted to assess the entire scale. EFA results indicated items only loaded onto their respective subscales, indicating the subscales are psychometrically distinct. The final Publication Bias, Statistical Significance, and Trust Subscales included construct-relevant items. For example, the Publication Bias Subscale items reflect issues that occur in the publication process and specific publication biases. The Trust Subscale’s items reflect trust-related behaviors such as worrying about the accuracy of results and distrusting research because of publication bias and QRPs. Lastly, the Statistical Significance Subscale items reflect the idea that an over-emphasis statistically significant results is problematic. Additionally,
the final PTM scale did not make theoretical sense as a composite measure, and should be
treated as individual scales in future research.

Hypothesis 1 predicted a negative correlation between graduate student awareness
of publication bias and their level of trust in the literature and was tested using the Trust
Subscale and the Publication Bias Subscale. This hypothesis was supported, \( r(119) =
-0.357, p < .001 \). This suggests when people become aware of biases and questionable
practices that can be used to enhance the likelihood of publication, they may view the
publication process and the results of that process as less trustworthy (Kepes &
McDaniel, 2013). This result should be of interest to psychologists because if these
behaviors continue, new psychologists (i.e., graduate students) are likely to be less
trusting of the research presented. This may be problematic because if people are less
likely to trust the research literature when they know about these behaviors, they may be
less likely to use the resulting research. Particularly concerning is that if new
psychologists do not trust the research literature, they may be less likely to use it in future
applied work, affecting the quality of their work and potentially lowering the value of
hiring psychologists in applied settings.

Hypothesis 2 was unable to be tested due to researcher oversight, but as a
replication of the Mullins et al. (2015) findings, should be considered for future research.

Hypothesis 3 predicted a negative correlation between perceived trustworthiness
of the literature and a general tendency to be distrustful. When tested using both the total
PTM and the Trust Subscale, this hypothesis was not supported. Hypothesis 4 predicted
a negative correlation between perceived trustworthiness of the literature and social
cynicism. This hypothesis was also tested using both the Perceived Trustworthiness
Measure and the Trust Subscale, and was not supported. Lastly, Hypothesis 5 predicted a negative correlation between perceived trustworthiness of the literature and negative affectivity was tested using both the PTM and the Trust Subscale and was not supported. I posit two explanations for these findings.

The general distrust, social cynicism, and negative affectivity scales were included as validation scales for the PTM developed in the present study. None of these validation hypotheses were supported using either the PTM or the Trust Subscale. These results may suggest that although theoretically these constructs should be related to the PTM or the Trust Subscale, the scope of the new scales may be too narrow to be related to the broad constructs of general distrust, social cynicism, and negative affectivity. In other words, the Trust Subscale evaluates perceived trust of the research literature as it relates to publication bias, QRPs, and ethical practices. This scale is highly specific in what it assesses, whereas the scales used for distrust, NA, and social cynicism are overall scales containing generalized items. Using a scale with more specific items for these constructs might result in different findings, such as a scale that assesses professional ethical beliefs. Additionally, as noted previously, using the full PTM to test the hypotheses did not make sense theoretically after the factor analysis was conducted. These non-significant results further support that the full PTM is not as conceptually meaningful as its subscales.

Moreover, a lack of correlation when using the Trust Subscale is not necessarily a problematic finding. It suggests that even though someone is generally distrustful, cynical, or high in NA, they do not automatically distrust the research literature, and that people who distrust the research literature likely do not have a significantly more
distrustful, negative, or cynical personality than a member of the general population would be expected to have. That is, there is no halo effect where general predispositions influence specific attitudes towards publication bias and research ethics. This suggests that the distrust in the research literature is due to behaviors such as publication bias and research ethics and not the reader’s personality. Although this is not what was originally hypothesized, this explanation of the data makes sense and indicates that the behaviors occurring in psychology are problematic, and that the perception of the problem is not a function of readers who are simply jaded or predisposed to negative emotional states. This further supports prior suggestions (e.g., Jeffrey et al., 2015; Kepes & McDaniel, 2013; Koehler et al., 2015; Mullins et al., 2015) that action should be taken to address publication bias and ethics in the research literature. Such actions would be less likely to succeed if the problem arose from individual dispositions than if the problem were demonstrably a function of various forms of research malpractice, as the present data would seem to suggest.

**Exploratory Analyses**

Exploratory relationships were examined, including differences between graduate students’ academic program type (M.A., M.S., Psy.D., Ph.D.) and their level of perceived trust in the research literature. Although no prior research drives the theory for this relationship, I posited students in longer, more research-focused programs (e.g., Ph.D. programs) might be less trusting of the literature because they have more exposure to the research and publication process, and potentially more direct experience with it as an author. To test this, an ANOVA was conducted using the participant’s program type and all of the subscales. The Trust Subscale and the Publication Bias Subscale ANOVAs
showed no significant difference based on program type, but, the Statistical Significance Subscale and the Research Ethics Subscale ANOVAs showed significant differences between M.S. students and Ph.D. students, with Ph.D. students having higher means. This may suggest Ph.D. students are more concerned with the emphasis put on obtaining statistical significance and the ethics used in research than are M.S. students, at least among those taking part in the present study. Whether this reflects any systematic difference in education based on program type is beyond the scope of the present data to identify.

These findings may support the notion that students in longer, more research focused programs may be more concerned with statistical significance and research ethics issues than at least some other students, potentially because of their experience and exposure with the research issues and publication process. I posit two explanations for this. First, issues such as publication bias and questionable research ethics may not be discussed equally in all graduate programs, which could explain why there were no significant differences between those in M.A. programs versus those in Ph.D. or M.S. programs. A second explanation is that this difference may occur because those in Ph.D. programs are likely training to become academics and researchers, whereas students in M.S. programs are likely to become practitioners. Those planning to become practitioners may have less concern with research practices and statistical significance, whereas future academics regard these as more problematic issues; differences across program types should, for all of these reasons, be viewed as very preliminary.

Overall, these findings are an interesting pattern; however, more research should be done to further explore possible differences between the different program types.
Specifically, replication with larger sample sizes is needed before any interpretation of the observed differences becomes practical or meaningful, especially because of the high means for all of the groups.

Another exploratory analysis examined how familiar respondents were with the research and publication process and effects this might have on trust of the literature. For example, students who have submitted work for publication might be less trusting in the literature if they have been rejected for non-significant findings than students who have never submitted research for publication. This was tested with an independent samples $t$-test (using the Trust Subscale as the DV and experience with the publication process as the IV). There was no significant difference on trust of the research literature between those who had experience with the publication process and those who did not. This finding is important, as it suggests that any distrust in the research literature could be attributed to knowledge of publication bias and issues in research ethics, not from any negative experiences they may have had with the publication process. This further supports the findings in the validation hypotheses that attitudes towards publication bias are not a result of predispositions or, in this case, negative experience with the publication process, and are actually indications of a belief that these behaviors are problematic and should be addressed.

**Contributions and Implications**

The measure developed in this study can be used for future research assessing issues such as searching for statistically significant results to increase publication chances, ethical issues in research, biases in the publication process, and trust in the research literature. Further research studying the implications of these biases and
behaviors have will help broaden researchers’ understanding of the problems. Additional research on this topic can use the measure developed in the present study, making future research easier to conduct.

Additionally, this study found support that publication bias and questionable research ethics affect graduate students’ perceived trust in the research literature. This study empirically supported what most research has suggested in only theoretical terms (e.g., Jeffrey et al., 2015; Kepes & McDaniel, 2013; Koehler et al., 2015). That is, other than Mullins et al. (2015), few studies have examined the effects of these behaviors (e.g., publication bias) have on the amount of trust graduate students have in the literature. The Mullins et al. study, however, was primarily exploratory and used a small sample from one graduate program, limiting generalizability. The present study is one of the first to examine the actual effects publication bias and QRPs/ethics have on attitudes about the research literature using a broader approach to developing a measure to assess the effect publication bias and research ethics violations have on graduate students’ perceived trust in the research literature. These findings can be used to support the need for changing how publications are selected and how researchers conduct research.

Furthermore, the results suggest that trust in the research literature may not be affected by one’s personality or experience with the publication process. That is, there was a significant correlation between trust in the research literature and knowing about publication bias, but there were no significant correlations with distrust, NA, social cynicism, and trust. Therefore, it is reasonable to suggest that distrust in the research literature is not due to personality or publication experience, but actual knowledge of and concern about the practices occurring (i.e., publication bias). This is important because it
demonstrates that changing the way research is selected for publication and conducting research more ethically (i.e., not mining for statistically significant results) could increase perceived trust in the research literature.

**Limitations and Future Directions**

One limitation of the present study is the high means for some of the scales developed in this study. The Research Ethics Subscale, the Publication Bias Subscale, and the Statistical Significance Subscale had means that ranged from $M = 4.33$, $SD = 0.25$ to $M = 4.72$, $SD = 0.29$ on a 5-point scale. These indicate a potential ceiling effect. Because the subscales measure thoughts and attitudes on ethical practices, respondents may answer higher, because of their high ethical standards. The Trust Subscale had a lower mean ($M = 3.50$, $SD = .68$), indicating no ceiling effect. This subscale measures actual publication bias and QRP behaviors respondents are aware of, rather than their thoughts and attitudes. A lower mean here may indicate respondents are aware of publication bias and issues with research ethics, which affects their trust. Taken together, these means may suggest that although respondents have strong feelings on the proper ethical and research procedures, when research does not align with these ideals, they are less likely to trust the research literature.

Another limitation is the response rate and sampling strategy issues. Initially, graduate Methods instructors or department chairs were asked to forward the study to their graduate students. Few programs actually forwarded the study, and those that did had few complete it. Because the study was not forwarded as initially expected, additional sampling strategies were needed. The second sampling strategy contacted additional departments and faculty in those departments. Participation only increased
slightly using this procedure; therefore, a third sampling procedure was required. The third procedure used a snowball sampling method where the study’s information was posted on Facebook, Reddit threads related to psychology, and emailed directly to people in psychology graduate programs. This provided enough responses to begin data analysis.

Students who decided to participate may differ from those who did not participate, especially those who found the survey via Facebook or Reddit. This is not something that can be controlled, and self-selection into a sample is a common issue. These possibilities mean there could be differences between those who participated and those who did not. If there is a difference between participants and non-participants, the results may not depict a full understanding of the impact publication bias has on the perceived trustworthiness of the research literature. Additionally, those who decided to participate might have stronger feelings about publication bias and the perceived trustworthiness of the literature, which might also explain the high means in several of the scales. This might affect the generalizability of the study; however, this is a limitation for all research studies.

Another limitation is Hypothesis 2. Due to an oversight in survey development, the question intended to measure Hypothesis 2 was not included. Therefore, the hypothesis could not be tested. Not being able to test this hypothesis limits the understanding of publication bias. Specifically, how publication bias affects how graduate students plan to use the research literature post-graduation is worth exploring in future studies.
Lastly, because this study produced one of the first systematically-designed scales to assess perceived trustworthiness of the literature, there is little available against which to validate this scale. Future research should examine the topics of publication bias and questionable research practices with additional validation measures. Future research should also focus on using the measure developed in this study with another psychology graduate student population. The sampling issues in this study may impact generalizability and the overall psychometrics of this scale. Testing this scale again will further assess its reliability and validity. Future research should also examine psychologists’ perceived trustworthiness of the research literature to understand how publication bias and questionable research practices affect the community of working psychologists.

Overall, this study intended to develop a measure to assess graduate students’ perceived trust in the research literature when graduate students are aware of publication bias and QRPs and provide empirical evidence about these behaviors. The aggregate PTM was, however, judged to be less useful than the subscales on their own. Hypothesis 1 was supported, which found a negative correlation between awareness of publication bias and perceived trust. That is, students who were aware of publication bias were less likely to trust the research literature. This finding confirms that publication bias is a problem and should be addressed. None of the validation hypotheses were supported. However, this suggests that to the extent to which the sample might have been generally distrustful, cynical, or high in NA, it did not meaningfully affect scores on the newly-developed measure. Taken together, these findings imply that publication bias and
engaging in questionable research ethics are problems that should be addressed in psychology.
Chapter VI

Summary

Review of the Literature

Publication bias, questionable research practices (QRPs), and the general approach to organizational science has become a growing topic of concern (Jeffrey et al., 2015; Kepes & McDaniel, 2013; Koehler et al., 2015; Mullins, Crowe, & Wymer, 2015). Developing a tool to accurately measure the consequences of publication bias on outcomes such as perceived trustworthiness of the literature is critical to evaluate the extent to which publication bias, and QRPs, impact science, and was the goal of the present study. Items for this scale derived from the literature review and the arguments drawn from Mullins et al. (2015) and Kepes and McDaniel (2013).

Publication Bias

Publication bias is the “selective publication of studies with a particular outcome” (Ferguson & Brannick, 2012, p. 120). One type is the “file drawer problem,” in which the researcher self-selects out of submitting a study for peer review or publication because of non-significant findings (Rosenthal, 1979). The second occurs when research rejected because it lacks statistically significant findings and thought of as having “no value” or possibly “invalid” (Greenwald, 1975).

Publication bias consequences. Publication bias has numerous consequences. Failure to publish null results can cause researchers to explore areas or relationships that
others have found to be non-significant, wasting time and resources (Franco, Malhotra, & Simonovits, 2014). It also skews the literature and increases the difficulty to correct Type I errors (Kepes & McDaniel; 2013; Mullins et al., 2015; Simmons, Nelson, & Simonsohn, 2012). Additionally, if researchers and journals consistently favor/publish studies with statistically significant results, then one outcome of this bias is the adoption of what the literature has come to refer to as “questionable research practices.” These need to be examined with publication bias to further assess the damage done to the trustworthiness of the research literature.

**Questionable Research Practices**

Kepes and McDaniel (2013) suggested publication bias may exist, in part, because articles with significant findings tend be seen as more interesting by readers, thus promoting readership and attention, enhancing a journal’s reputation, increasing the demand for academics to publish (O’Boyle et al., 2014). There are several types of QRPs, with HARKing and *P*-hacking having received substantial attention.

**HARKing.** One type of QRP is HARKing, or Hypothesizing After the Results are Known (Kerr, 1998). This is also described as “presenting post hoc hypotheses as a priori” (PPHA; Leung, 2011). Researchers who HARK decide what to report and how to report it after the data are analyzed to downplay, or remove, any non-significant results, and to highlight significant but non-hypothesized findings (Kerr, 1998).

**P-hacking.** *P*-hacking is another questionable practice that might affect the perceived trustworthiness of the literature. Simmons, Nelson, and Simonsohn (2011) suggested this is done by consciously exploiting decisions to obtain statistical significance or “researcher degrees of freedom” (p. 1).
Outlined just some of the consequences of publication bias. Conducting research assessing how publication bias has affected the perceived trustworthiness of the literature and assessments about these practices must be conducted to understand the implications.

**Trustworthiness**

Questions of perceived trustworthiness of a literature must be asked in the context of that literature’s readership. How graduate students perceive the trustworthiness of the publication process and research journals is likely most critical, because this population is likely new to the publication process and to the streams of literature in research journals.

**Measuring perceived trustworthiness.** Mullins et al. (2015) developed an exploratory scale to assess graduate students’ perceptions of trust in the literature. Although Mullins et al.’s (2015) research was exploratory, and the authors noted the generalizability and power of their sample is questionable. The primary goal of this study is to develop a more robust Perceived Trustworthiness Scale and further research.

**Validation constructs.** When developing a scale, it is critical to include validated scales to establish convergent validity. The following section will explore potential areas that could provide convergent validity evidence.

**Distrust.** Distrustful people have a tendency to take a defensive stance in general (Lewicki, McAllister, & Bies, 1998). If a person is generally distrustful, it stands to reason they might naturally be distrustful of the research literature, so to establish validity, including a distrust scale from the International Personality Item Pool (IPIP) will be beneficial.

**Social cynicism.** Social cynicism is “the overall belief that people, institutions, etc. cannot be trusted” (Leung et al., 2002). This makes social cynicism relevant for
validation purposes because the perceived trustworthiness scale will assess the processes used to publish and the institutions that publish that work.

**Negative affectivity (NA).** Research has shown those high on NA are likely to be upset and have a negative self-view (Watson & Clark, 1984). Additionally, people with high NA tend to magnify mistakes, disappointments, and threats.

**Conclusion**

Overall, this study aimed to develop a scale to assess the impact publication bias and questionable research practices have on graduate students’ trust in the research literature using issues outlined in previous research and conference topics (Kepes & McDaniel 2013; Kerr, 1998; Mullins et al., 2015; O’Boyle et al., 2014; Simmons et al., 2011; Simmons et al., 2012).

**Rationale and Hypotheses**

The literature highlights some of the major issues and consequences faced because of publication bias and QRPs. Most notable is the need for research to examine what these practices have done to the perceived trustworthiness of science.

The following relationship should exist because when people become aware of biases and questionable practices that can be used to enhance the likelihood of publication, they may view the publication process and the results of that process as less trustworthy (Kepes & McDaniel, 2013). As such, the following is hypothesized.

*Hypothesis 1: There will be a negative correlation between graduate student awareness of publication bias and their level of trust in the literature.*
Validation Constructs

Graduate students who are more generally distrustful may also have a higher distrust of the research literature, parallel logic should follow for social cynicism and those high on negative affectivity. As such, the following relationships are hypothesized.

Hypothesis 2: There will be a negative relationship between perceived trustworthiness of the literature and a general tendency to be distrustful.

Hypothesis 3: There will be a negative relationship between perceived trustworthiness of the literature and social cynicism.

Hypothesis 4: There will be a negative relationship between perceived trustworthiness of the literature and negative affectivity.

Method

Participants

Participants were from M.A., M.S., Psy.D., and Ph.D. Programs in a variety of psychology programs. Because exploratory factor analysis played a key role in data analysis, initially it was decided to use a minimum sample size of at least useable 200 participants (Maccallum, Widaman, Zhang, & Hong, 1999; Spector, 2002). However, collecting 200 participants was challenging and after three months a sample size of 120 participants was agreed-upon by the thesis committee.

Measures

Perceived trustworthiness. A questionnaire designed for the present study to assess the perceived trustworthiness of the research literature and the publication process. Participants responded on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). Based on a literature review, some potential items were as follows: “I
perceive the research literature as ethical” and “Authors always report exactly what
occurs in their studies.”

**Distrust.** The distrust scale from the International Personality Item Pool (IPIP)
was utilized. This scale has 10 items, with an example item being “I often suspect hidden
motives in others” (Goldberg et al., 2006). Participants responded on a 5-point scale
ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The scale has a previously-
reported Cronbach’s alpha reliability of .83 (International Personality Item Pool).

**Social cynicism.** Social cynicism was measured using a subscale from the Social
Axioms Scale (SAS), which has 13 items (Leung et al., 2002). Respondents were to rate
items on a 5-point scale, indicating the degree to which they believe each statement to be
true, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). An example item is, “If
one belongs to a marginal group, it is difficult to gain acceptance from the majority
group” (Leung et al., 2002). Leung et al. (2012) reported a Cronbach’s alpha of .79 for a
US sample.

**Negative affectivity.** Negative affectivity was measured with a 5-item short form
scale from Thompson (2007). Respondents were given the following instructions:
“Thinking about yourself and how you normally feel, to what extent do you generally
feel...” Items were answered on a 5-point scale ranging from 1 (*never*) to 5 (*always*). An
exam item is “upset.” Prior research demonstrated Cronbach’s alpha of .74 (Thompson,
2007).

**Demographics.** The demographic information collected included information on
age, sex, race, program type and field, number of years in the program, publication
experience, and familiarity with research journals. This was used to evaluate possible
background-based differences (e.g., program type/field, publication familiarity) in graduate students’ level of perceived trust in the research literature.

**Procedure**

The sample was obtained through two waves of contact with instructors of graduate research methods courses, who were asked to send the survey link to their students, and a snowball sampling strategy via social media (e.g., Reddit) and graduate student contacts.

**Results**

To analyze the factor structure of the Perceived Trustworthiness Scale, an exploratory principal components analysis (PCA) with varimax rotation to extract the factors was conducted on each of the subscales. To determine the number of factors in each subscale, the rotated component matrix was examined. Once the factors were identified and initial item cuts were made based on the PCA, all theoretically-constructed subscales were subjected to an item analysis. Based on the item analysis, if removing an item increased Cronbach’s alpha, then the item was considered for removal. If an item was removed, the item analysis was re-run (without the removed item) and remaining items were reexamined for possible removal. This continued until Cronbach’s alpha would either not increase by removing an item, or removing an item for the amount alpha would increase did not make sense in the author’s judgment. Once the item analysis was completed, a PCA with varimax rotation was conducted on the final scale to check for unidimensionality. Based on the above methodology, the final Publication Bias Subscale contains 5 items with a Cronbach’s alpha of .859. The final QRPs Subscale 5 had items which a Cronbach’s alpha of .766, however, based on the final item content the name was
changed to Research Ethics Subscale (this will be used going forward). The final Statistical Significance Subscale contains 4 items, with a Cronbach’s alpha of .779, and the completed Trust Subscale contains 3 items, with a Cronbach’s alpha of .817.

The hypotheses were to be tested using the total Perceived Trustworthiness Measure. However, once the subscales were factor analyzed, using a total scale score did not make sense. Although, it does not make sense to use an aggregate score for the total measure in future work, the present study originally hypothesized relationships with the total Perceived Trustworthiness Measure, which now will be referred to as PTM. As such, the hypotheses were tested using both the PTM and the appropriate subscale.

Hypothesis 1 was tested using the Trust Subscale (final items 16-19) and the Publication Bias Subscale (final items 1-5). This hypothesis was supported, $r(119) = - .357, p < .001$. Hypothesis 2 was tested with the PTM and the Distrust Scale and was not supported ($r(119) = -.155, p = .088$) and was not supported when tested with the Trust Subscale and the distrust scale ($r(119) = -.030, p = .744$). Hypothesis 3 was tested with the PTM and cynicism scales was not supported ($r(119) = -.122, p = .183$) and was not supported when tested with the Trust Subscale and the social cynicism scale, ($r(119) = -.072, p = .433$). Lastly, Hypothesis 4 was tested with the PTM and the negative affectivity scale and was not supported, ($r(119) = -.095, p = .299$) for either trust measure, ($r(119) = -.011, p = .909$).

**Discussion**

The primary objective of this study was to develop a scale to assess graduate students’ perceived trustworthiness of the research literature. This study generated potential items to develop a more comprehensive scale. The PCA and item analysis of
each of the subscales identified a unidimensional set of items for each subscale with an acceptable alpha level (.76-.85).

Hypothesis 1 predicted a negative correlation between graduate student awareness of publication bias and their level of trust in the literature and was tested using the Trust Subscale (final items 16-19) and the Publication Bias Subscale (final items 1-5) and was supported. This suggests when people become aware of biases and QRPs, they may view the publication process and the results of that process as less trustworthy (Kepes & McDaniel, 2013).

The general distrust, social cynicism, and negative affectivity scales were included as validation scales for the PTM developed. Hypotheses two – four predicted a negative correlation between the validation scale and the PTM. None of these hypotheses were supported with the PTM or the Trust Subscale. These results may suggest that although theoretically these constructs should be related to the PTM or the Trust Subscale, the new scale may too narrow to be related to the broad validation constructs. Using a scale with more specific items for these constructs might result in different findings. Moreover, a lack of correlation when using the Trust Subscale is not necessarily a problematic finding. It suggests that even though someone is generally distrustful, cynical, or high in NA, they do not automatically distrust the research literature. Although not originally hypothesized, this explanation makes sense and indicates that the behaviors occurring in psychology are problematic, and that the perception of the problem is not a function of readers who are simply jaded or predisposed to negative emotional states. This further supports prior suggestions (e.g., Jeffrey et al., 2015; Kepes & McDaniel, 2013; Koehler et al., 2015; Mullins et al., 2015).
that action should be taken to address publication bias and ethics in the research literature. Such actions would be less likely to succeed if the problem arose from individual dispositions than if the problem were demonstrably a function of various forms of research malpractice, as the present data would seem to suggest.

**Contributions and Implications**

The measure developed in this study can be used for future research assessing issues such as searching for statistically significant data to increase publication chances, ethical issues in research, biases in the publication process, and trust in the research literature. Further research studying the implications of these biases and behaviors have will help broaden researchers’ understanding of the problems. That future research is now easier to conduct because a reliable measure has been developed.

Additionally, this study found support that publication bias and questionable research ethics affect graduate students’ perceived trust in the research literature. This study empirically supported what most research has suggested in only theoretical terms (e.g., Jeffrey et al., 2015; Kepes & McDaniel, 2013; Koehler et al., 2015). The present study is one of the first to examine the actual effects publication bias and QRPs/ethics have on attitudes about the research literature using a broader approach to developing a measure to assess the effect publication bias and research ethics violations have on graduate students’ perceived trust in the research literature. These findings support the need for changing how publications are selected and how researchers conduct research.

**Limitations and Future Directions**

Every study has potential limitations. One limitation of the present study is the high means for some scales developed in this study. The Research Ethics Subscale, the
Publication Bias Subscale, and the Statistical Significance Subscale had means that ranged from $M = 4.33$ to $M = 4.72$ on a 5-point scale, indicating a potential ceiling effect. Because the subscales measure thoughts and attitudes on ethical practices respondents may answer higher, because of their high ethical standards. The Trust Subscale had a lower mean ($M = 3.50$, $SD = .68$), indicating no ceiling effect. This subscale measures actual publication bias and QRP behaviors respondents are aware of, rather than their thoughts and attitudes. A lower mean here may indicate respondents are aware of publication bias and issues with research ethics, which affects their trust. Taken together, these means may suggest that although respondents have strong feelings on the proper ethical and research procedures and when research does not align with these ideals, they are less likely to trust the research literature. Another limitation is the response rate and sampling strategy issues. Initially, graduate Methods instructors or department chairs were asked to forward the study to their graduate students. Few programs actually forwarded the study, and those that did had few complete it. Because the study was not forwarded as initially expected, additional sampling strategies were needed.

Lastly, because this study produced one of the first systematically-designed scales to assess perceived trustworthiness of the literature, there is little available against which to validate the scale. Future research should examine the topics of publication bias and QRPs with additional validation measures. Future research should also focus on using the measure developed in this study with another psychology graduate student population. The sampling issues may impact generalizability and the overall psychometrics and testing the scale again will further assess its reliability and validity.
Overall this study intended to develop a measure to assess graduate students’ perceived trust in the research literature when graduate students are aware of publication bias and QRPs and provide empirical evidence about these behaviors. Additionally, only hypothesis one was supported, which found a negative correlation between awareness of publication bias and perceived trust. That is, students who were aware of publication bias were less likely to trust the research literature. This finding confirms that publication bias is a problem and should be addressed. None of the validation hypotheses were supported. This suggests that to the extent the sample might have been generally distrustful, cynical, or high in NA, it did not meaningfully affect scores on the newly-developed measure. Taken together, these findings that publication bias and engaging in questionable research ethics are problems that should be addressed in psychology.
References


science and practice breakdowns. Presentation at the Thirtieth Annual Meeting of the Society for Industrial and Organizational Psychology, Philadelphia, PA.


Koehler, T., Vandenberg, R.J., Rogelberg, S.G., Cortina, J.M., Landis, R.S., & Tonidandel, S. (2015, April). Pursuing Better Science in Organizational Psychology There is greater recognition that the current incentive structure in organizational science needs to be improved. Presentation at the Thirtieth Annual Meeting of the Society for Industrial and Organizational Psychology, Philadelphia, PA.


Appendix A

Demographic Information

Age: ______

Sex:

_____ Male

_____ Female

_____ Prefer not to respond

In terms of race/ethnicity, how do you describe yourself? (Please select the option that best describes you)

_____ American Indian or Alaska Native

_____ Hawaiian or Other Pacific Islander

_____ Asian or Asian American

_____ Black or African American

_____ Hispanic or Latino

_____ White (Non-Hispanic)

_____ Multi-racial

_____ Other: _______________________

_____ Prefer not to respond

Program Type

_____ MA

_____ MS

_____ Psy.D.

_____ Ph.D.
___Ed.D.
___Other (please specify) __________________

Program Field
___: Clinical
___: Cognitive
___: Counseling
___: Developmental
___: Ecological
___: Educational
___: Experimental
___: Human Factors
___: Industrial-Organizational
___: Neuroscience
___: Personality
___: Physiological
___: Quantitative
___: Social
___: Other __________

Number of year(s) enrolled in your program: ____

For the following questions, please consider papers for which you have been either the first author or a co-author.

How many papers have you submitted to peer-reviewed journals? ____

How many papers have you had accepted for publication in peer-reviewed journals? ____
How many times have you had papers rejected by peer-reviewed journals? ____

How many times have you been offered the opportunity to revise and resubmit a paper at a peer-reviewed journal? ____

How many times have you decided not to submit a paper for publication because the results were not statistically significant? ____

On average, how many empirical research articles do you read for graduate school (GA/TA position, classes, Thesis/Dissertation) a week?

In considering the research literature, I am more interested in general take-aways than specific methodologies (e.g. summarizing the results without focusing heavily on the method used). (Scale 1-5, strongly disagree to strongly agree)

Have you ever observed someone engage in HARKing (Hypothesizing After the Results are Known)?
   ____: Yes
   ____: No
   ____: Unsure

Have you ever observed someone engage in P-Hacking (consciously exploiting research decisions to obtain statistical significance)?
   ____: Yes
   ____: No
   ____: Unsure
Appendix B

Amazon Gift Card Contest Survey

Please enter your email address for a chance to win one of five $20 Amazon gift cards.

Thank you for participation.

For issues contact Sarah Iman, at mclaughlins1@xavier.edu.

Email Address:
Appendix C
Perceived Trustworthiness Scale

Directions and Scale Anchors:
Respondents are asked to rate on a scale of 1-5, strongly disagree to strongly agree and are provided the following definitions of publication bias, HARKing, and P-Hacking.

In case you aren’t familiar with “publication bias,” “HARKing,” or “P-hacking,” here are some brief definitions.

Publication bias is the application of any decision rule that tends to specifically exclude certain types of research or research findings (e.g., a tendency to only publish statistically significant results).

HARKing is Hypothesizing After the Results are Known, and as the name suggests involves the writing of hypotheses based on data analyses already conducted. Lastly, P-Hacking is consciously exploiting research decisions to obtain statistical significance.

Items are grouped by potential subscales. Reverse-scored items are marked with an “(R)”.

Publication Bias Subscale Items 1-18

1. Researchers sometimes choose not to submit studies for publication if their results are non-significant.

2. Researchers are trained to believe that only statistically significant results are worthy of publication.

3. It is important as a field to stop publication bias.
4. Journals should publish both significant and non-significant research findings.

5. The publication process is unbiased. (R)

6. The published research literature is accurate. (R)

7. The published research literature shows an accurate representation of all research findings. (R)

8. Replication work is valuable.

9. Replication research should be published.

10. Replication work provides information that is important.

11. Without replication work, errors in the literature cannot be corrected.

12. Replication work helps correct errors from prior research.

13. The published literature available is representative of all research done in my field. (R)

14. If the published literature available was not representative of all research conducted in my field, I would know.

15. Non-significant results are important to publish.

16. Only statistically significant results should be published. (R)

17. Publication bias makes research less useful for specific problems.

18. Publication bias hurts the overall field.

**Questionable Research Practices Subscale Items 19-50**

19. Publication bias can affect research for years to come.

20. There is a high demand for academics to publish.

21. Universities put too much pressure on academics to publish.
22. The demand for publication has caused researchers to engage in unethical research practices.

23. Altering findings to increase chances of publication is unethical.

24. Performing science ethically is important.

25. Manipulating the presentation of results to obtain significance is a problem.

26. Data should be reported as they were collected, with any alterations fully reported.

27. It is unethical to manipulate the presentation of results to obtain statistical significance.

28. Because findings may be inaccurate, practitioners should be careful when using results to solve specific problems.

29. All hypotheses should be reported regardless of the findings or results.

30. It is acceptable to add hypotheses after data are analyzed. (R)

31. Mixing a priori and post hoc hypotheses without explaining which is which is ethical. (R)

32. Adding hypotheses that fit the post hoc data is acceptable. (R)

33. Authors always report exactly what occurs in their studies. (R)

34. It is hard to tell what research is HARKed or P-Hacked.

35. Prior to this study I was familiar with HARKing.

36. Prior to this study I was familiar with P-Hacking.

37. Researchers only need to report "what worked" in their study. (R)

38. Scientists have a duty to uphold ethical scientific practices.
39. Scientists who engage in practices like P-Hacking and HARKing should face some type of consequence.

40. Students are adequately educated about issues such as publication bias and questionable research practices. (R)

41. Journals should publish a mix of significant, non-significant, and replication work.

42. Journals should be able to publish whatever research they like. (R)

43. Students should be made aware of publication bias and questionable research practices like HARKing and P-Hacking in graduate school.

44. Issues of publication bias and questionable research practices should be discussed more in undergraduate psychology education.

45. Questionable research practices hurt the entire field.

46. Only testing the hypotheses made before the data are analyzed is ethically appropriate.

47. Hypotheses made after the data are collected should be noted as such.

48. Unethical research done today will harm research in the future.

49. Teaching graduate students ethical research practices is important.

50. Researchers should be allowed to conduct research in any way they want, regardless of the ethics. (R)

**Statistical Significance Subscale Items 51-62**

51. Applied work is negatively affected by issues such as publication bias and questionable research practices.

52. University researchers view non-significant results as a form of personal failure.
53. University researchers focus on publishing only significant results because outcomes such as salary and tenure depend on publications.

54. Consumers of research are only interested in significant results.

55. It is more important to report results accurately than it is to obtain statistical significance.

56. Journals only want to publish significant results.

57. Journal are not interested in research with non-significant findings.

58. Hypotheses without significant findings should not be reported. (R)

59. I would submit non-significant results without worrying they will be rejected.

60. Findings in research journals matter to my field.

61. Graduate students are trained to value obtaining statistical significance.

62. Research that is non-significant is also in-significant. (R)

**Trust Items: 63-75**

63. Mining the data for any significance it contains is important. (R)

64. If I learned a researcher engaged in unethical research practices, I would distrust the work.

65. If I learned that multiple researchers engaged in unethical research practices, my trust in the field as a whole would be diminished.

66. I do not worry about the accuracy of the results in the research literature.

67. I perceive the research literature as ethical.

68. I trust the results in research journals.

69. The research literature is reliable.

70. Findings in research journals are something I pay close attention to.
71. Because of publication bias, I distrust research literature. (R)

72. Because of questionable research practices (e.g. HARKing/ P-Hacking), I distrust the research literature. (R)

73. I trust research to be generalizable.

74. Knowing that publication bias occurs makes me trust all research less. (R)

75. Readers of research must be aware of potential biases in the literature. (R)
Appendix D

Distrust Scale

These items are not presented in order to protect the rights of the original copyright holder. They were obtained for this study from IPIP (Goldberg et al., 2006).
Appendix E

Social Cynicism Scale

These items are not presented in order to protect the rights of the original copyright holder. They were obtained for this study from Leung et al. (2002).
Appendix F

Negative Affectivity Scale

These items are not presented in order to protect the rights of the original copyright holder. They were obtained for this study from Thompson (2007).
Appendix G
IRB Approval Letter (Sampling Strategy 1)

September 3, 2015

Sarah Iman

Re: Protocol #15-022, Publication Bias and Graduate Students’ Perceived Trust in the Literature

Dear Ms. Iman:

The IRB has reviewed the materials regarding your study, referenced above, and has determined that it meets the criteria for the Exempt from Review category under Federal Regulation 45CFR46. Your protocol is approved as exempt research, and therefore requires no further oversight by the IRB. We appreciate your thorough treatment of the issues raised and your timely response.

If you wish to modify your study, including the addition of data collection sites, it will be necessary to obtain IRB approval prior to implementing the modification. If any adverse events occur, please notify the IRB immediately.

Please contact our office if you have any questions. We wish you success with your project!

Sincerely,

Kathleen J. Hart, Ph.D., ABPP
Vice Chair, Institutional Review Board
Xavier University

KJH/sb
Appendix H

Request for Participation Email (Sampling Strategy 1 and 2)

Dear Dr. [name],

My name is Sarah Iman and I am a Master’s candidate in I-O psychology at Xavier University. I am working on my thesis examining how research and publication practices affect graduate student views of the research literature. I am emailing you to ask if you would forward this email to the graduate students in your department, to invite them to take part. I obtained your contact information from your department’s web page, which identified you as someone who teaches graduate students research methods and/or statistics. Your university was chosen randomly from the population of universities granting graduate degrees in psychology, based on the American Psychological Association 2015 Graduate Study in Psychology. Understanding how the current state of the publication process might affect how the next generation of psychologists views and uses the empirical literature is important, and I hope that you can assist me by sharing the information below with your students. I would be happy to answer any further questions you might have about my study, and will of course share a summary of my results with you if you are interested in receiving one.

Here is the link to the survey and the information for participants.

[https://xavier.co1.qualtrics.com/SE/?SID=SV_bwTfBnuSE6Ye9L]
Greetings participants,

My name is Sarah Iman and I am a Master’s candidate in I-O psychology at Xavier University. I am working on my thesis examining how research and publication practices affect graduate student views of the research literature. I am asking you to take part in my study. Understanding how the current state of the publication process might affect how the next generation of psychologists views and uses the empirical literature is important, and I hope that you will participate.

Participation is voluntary. The survey is approximately 120 items online and should take no more than 30 minutes. At the end of this survey, you may provide contact information in a second survey to be entered to win one of five $20 Amazon Gift Cards. Based on the number of responses data collection will close six weeks after data collection begins.

If you have any questions at any time during the study, you may contact Sarah Iman at mclaughlins1@xavier.edu or her research supervisor, Dr. Morrie Mullins at mullins@xavier.edu.

Thank you.

Sarah Iman
Appendix I

INFORMED CONSENT FORM

My name is Sarah Iman and you are being asked to volunteer to participate in a project conducted through Xavier University as part of my master’s thesis research.

The purpose of this study is to examine how specific research and publication practices affect your views and usage of the research literature. This topic was selected because of the need to understand the consequences of specific behaviors becoming more common in science. You will answer approximately 120 items online and the entire study should take no more than 15-30 minutes. I understand that this is a lot of items, but I am in the validation stage for this measure, and as a graduate student in psychology I am sure that you understand the importance of working from a large item pool. There are no anticipated risks/discomforts greater than what you would normally experience when being on a computer. At the end of this survey, you may provide contact information in a second survey (one item) to be entered to win one of five $20 Amazon Gift Cards.

The only personal information that will be collected (your email address) will be in a separate one-item survey linked from the main survey once it is completed. This information is only collected if you would like a chance to win an Amazon Gift Card for participation. Collecting your contact information in a separate survey ensures, to the best of my ability, that all survey responses will be anonymous (the highest level of anonymity/privacy will be selected in the Qualtrics control panel) and cannot be linked to
you whether or not you enter the Amazon Gift Card drawing. Refusal to participate in 
this study will have NO EFFECT ON ANY FUTURE SERVICES you may be entitled to from the 
University. You are FREE TO WITHDRAW FROM THE STUDY AT ANY TIME WITHOUT 
PENALTY.

If you decide to participate in the project, please click the I AGREE box below. You 
may print or save a copy of this form to keep. If you have any questions at any time 
during the study, you may contact Sarah Iman at mclaughlins1@xavier.edu or her 
research supervisor, Dr. Morrie Mullins at mullins@xavier.edu. Questions about your 
rights as a research participant should be directed to Xavier University’s Institutional 
Review Board at (513) 745-2870 or at irb@xavier.edu.

I have been given information about this research study and its risks and benefits and 
have had the opportunity to ask questions and to have my questions answered to my 
satisfaction. By clicking “I AGREE” I freely give my consent to participate in this 
research project.

Click “I AGREE” to participate. If you do not agree to participate, you may exit the 
survey now. In either case, thank you for your time.
Appendix J

Debrief

This study examined the effects of publication bias and questionable research practices on the perceived trustworthiness of the research literature through developing a scale to assess the consequence of these behaviors.

If you would like full details about the study’s hypotheses, you can email the principal investigator, Sarah Iman, at mclaughlins1@xavier.edu, and I will send a more detailed explanation after the study is complete. If you have any questions or would like a copy of the results when data analysis is complete, you may contact the principal investigator, Sarah Iman, at mclaughlins1@xavier.edu.

Questions about your rights as research subject should directed to Xavier University’s Institutional Review Board at 513-745-2870 or irb@xavier.edu.

Thank you for your participation.

If you would like to be considered for the Amazon Gift Card drawing, please click the link below to an external survey. This survey will say: please enter your email address for a chance to win a $20 Amazon gift card. Thank you for your participation. For issues contact Sarah Iman, at mclaughlins1@xavier.edu.

[https://xavier.co1.qualtrics.com/SE/?SID=SV_6RmrjwowTyBzZjf]
September 16, 2015

Sarah Iman

Re: Protocol #15-022, Publication Bias and Graduate Students’ Perceived Trust in the Literature

Dear Ms. Iman:

The IRB has reviewed the request to modify your study, referenced above. We understand that you have revised the recruitment procedure and will be contacting 20 new schools. We are able to continue to approve your study based on the information you provided. Therefore, your above-referenced study, as modified, continues to be approved in the Exempt category under Federal Guidelines 45CFR46.

Please note that if you wish to further modify your study, it will be necessary to obtain IRB approval prior to implementing the modification. If any adverse events occur, please notify the IRB immediately.

We truly appreciate your efforts and attention to compliance within the spirit of human subject’s protection. We wish you great success with your research.

Sincerely,

Kathleen J. Hart, Ph.D., ABPP
Vice Chair, Institutional Review Board
Xavier University

KJH/sb
Appendix L
IRB Approval Letter (Sampling Strategy 3)

October 21, 2015

Sarah Iman

Re: Protocol #15-022, Publication Bias and Graduate Students’ Perceived Trust in the Research Literature

Dear Ms. Iman:

The IRB has reviewed the request to modify your study, referenced above. We understand that you will be modifying your sample collection method. We are able to continue to approve your study based on the information you provided. Therefore, your above-referenced study, as modified, continues to be approved in the Exempt category under Federal Guidelines 45CFR46.

Please note that if you wish to further modify your study, it will be necessary to obtain IRB approval prior to implementing the modification. If any adverse events occur, please notify the IRB immediately.

We truly appreciate your efforts and attention to compliance within the spirit of human subject’s protection. We wish you great success with your research.

Sincerely,

Kathleen J. Hart, Ph.D., ABPP
Vice Chair, Institutional Review Board
Xavier University

KJH/sb
Appendix M

Request for Participation (Sampling Strategy 3)

Here is a description of what I provide people when contacting people about my study.

Email to direct contact I know:
Hi (insert name),

As you know I am working on my master’s thesis at Xavier University. I need participants and was wondering if you would be willing to help with my study. It is 120 items, and there is a chance to enter a raffle to win 1 of 5 $20 Amazon Gift Cards for taking part. Please feel free to send this study along to friends in any psychology graduate program in the United States.

Here is the link to the study, as well as information you can use to introduce it to people you may send it to.
https://xavier.co1.qualtrics.com/SE/?SID=SV_bwTfBnuSE6Ye9L

Dear fellow students,

My name is Sarah Iman and I am a Master’s candidate in I-O psychology at Xavier University. I am working on my thesis examining how research and publication practices affect graduate student views of the research literature. I am asking you to take part in my study. Understanding how the current state of the publication process might affect how the next generation of psychologists views and uses the empirical literature is important, and I hope that you will participate.

Participation is voluntary. The survey is approximately 120 items online and should take no more than 30 minutes. At the end of this survey, you may provide contact information in a second survey to be entered to win one of five $20 Amazon Gift Cards. Based on the number of responses data collection will close six weeks after data collection begins.

If you have any questions at any time during the study, you may contact Sarah Iman at mclaughlins1@xavier.edu or her research supervisor, Dr. Morrie Mullins at mullins@xavier.edu. This study has been reviewed and approved by the Xavier University IRB.

Thank you.
Sarah Iman

Email to a direct faculty contact I know:
Hi (insert name),
I hope you are doing well.

As you know I am working on my master’s thesis at Xavier University developing a scale to assess graduate students’ perceived trust in the research literature. I need participants and was wondering if you would be willing to send this study along to psychology graduate students you know and faculty members in other psychology graduate programs to ask if they would ask their students to take part.

Below is the link to the study and the information you can use to introduce the study.
https://xavier.co1.qualtrics.com/SE/?SID=SV_bwTfBnuSE6Ye9L

Dear fellow students,

My name is Sarah Iman and I am a Master’s candidate in I-O psychology at Xavier University. I am working on my thesis examining how research and publication practices affect graduate student views of the research literature. I am asking you to take part in my study. Understanding how the current state of the publication process might affect how the next generation of psychologists views and uses the empirical literature is important, and I hope that you will participate.

Participation is voluntary. The survey is approximately 120 items online and should take no more than 30 minutes. At the end of this survey, you may provide contact information in a second survey to be entered to win one of five $20 Amazon Gift Cards. Based on the number of responses data collection will close six weeks after data collection begins.

If you have any questions at any time during the study, you may contact Sarah Iman at mclaughlins1@xavier.edu or her research supervisor, Dr. Morrie Mullins at mullins@xavier.edu. The study has been reviewed and approved by the Xavier University IRB.

Thank you,
Sarah Iman
Appendix N

Final Perceived Trustworthiness Measure

Publication Bias Subscale (Original Item Numbers 3, 4, 15, 16, 18)
Cronbach’s alpha .859

1. It is important as a field to stop publication bias.
2. Journals should publish both significant and non-significant research findings.
3. Non-significant results are important to publish.
4. Only statistically significant results should be published. (R)
5. Publication bias hurts the overall field.

Research Ethics Subscale (Original Item Numbers 23, 24, 38, 45, 49.)
Cronbach’s alpha .766

6. Altering findings to increase chances of publication is unethical.
7. Performing science ethically is important.
8. Scientists have a duty to uphold ethical scientific practices.
9. Questionable research practices hurt the entire field.
10. Teaching graduate students ethical research practices is important.

Statistical Significance Subscale (Original Item Numbers 53, 56, 57, 61)
Cronbach’s alpha .779

11. University researchers focus on publishing only significant results because outcomes such as salary and tenure depend on publications.
12. Journals only want to publish significant results.

13. Journals are not interested in research with non-significant findings.

14. Graduate students are trained to value obtaining statistical significance.

Trust Subscale (Original Item Numbers 67, 68, 69)

Cronbach’s alpha = .817.

15. I perceive the research literature as ethical.

16. I trust the results in research journals.

17. The research literature is reliable.