# The necessities for building a thesis statement

Nathan Hamlitsch
*Nagoya University*

A research paper is essentially about providing well-argued support and/or counterevidence for a claim or thesis statement. As such, it is hard to overstate the importance of understanding how to create a thesis statement. Yet constructing one from scratch can be most challenging for a new researcher. In this paper, we will look at the necessary elements for constructing a thesis statement, and the consequences if one of these elements is missing or mistaken for the thesis statement itself. Other methods have suggested steps (i.e. brainstorming) that a researcher could take leading up to the formation of a thesis statement. These approaches, although useful, do not discuss necessary elements for constructing a thesis statement. In the present work, I propose that before a thesis statement is built, it is essential to have: 1) an observed phenomenon, followed by 2) a research question. Creating these two elements first will lead directly to the construction of a meaningful thesis statement. Many new researchers have trouble progressing beyond the observed phenomenon stage. This often leads to a paper that devotes most of its pages to categorizing or describing parts of a phenomenon, which leaves the writer no choice but to end the paper with a summary, rather than a conclusion. Additionally, when a researcher stops short of creating a thesis statement but formulates a research question, the research question (in many cases stated as an objective) can be mistaken for the thesis statement, or the researcher inadvertently creates the risk of *post hoc* reasoning (i.e. data dredging or the Texas sharpshooter fallacy) when interpreting results.

## 1. Introduction

"I'll Show You My Underwear (or How to Tackle the Dreaded Thesis Statement)" is the title of an essay published in *Writing Magazine* (Frank 2005), a magazine published in the UK that covers, as the name suggests, all things writing. This essay takes the position that, just like underwear, showing one's thesis statement to the public is quite uncomfortable; it's embarrassing, inappropriate, and can make you feel vulnerable. Much of students' reluctance to produce (a clear) thesis statement stems from an uncertainty about how to create one. Some are even unsure whether the sentence they've produced is a thesis statement. This creates a situation where the student wants to keep anything that resembles a thesis statement well hidden and out of the public eye for fear of being ridiculed. In this paper, I propose that students need two things before they can produce a thesis statement. The first is an observed phenomenon. The second is a research question. If a student has these two elements, they can then proceed to produce a meaningful thesis statement with confidence.

There is some helpful research on the topic of thesis statement building. One example is Tajino, Stewart, and Dalsky's (2010) book *Writing for Academic Purposes*. In that work, it is proposed that students follow a four step procedure that will help them take a broad topic (such as "climate change"), and narrow it down to a specific claim about that topic. For convenience, I have reproduced a condensed version of their example in Table 1.

*NU Ideas* Volume 4.2. 2015. Nagoya University Institute of Liberal Arts & Sciences

| |
|---|
| 1) Brainstorm the topic |
| • Climate change |
| 2) Narrow the topic |
| • Climate destruction and the Kyoto Protocol |
| 3) Clarify your stance |
| • Local action supporting the Kyoto Protocol targets |
| 4) Make a specific assertion based on clearly stated support |
| • The situation with climate change is becoming critical for human civilization. In this paper, I will first argue that humans are causing climate destruction and that human activity must be checked. I will then provide a basic outline of the Kyoto Protocol of 1997 and describe how institutions in the Kansai region of Japan are taking action in support of the protocol. |

**Table 1.** Thesis Statement (Tajino, Stewart & Dalsky 2010, 79-80)

Table 1 shows the development of a thesis statement starting from brainstorming about the topic in 1, narrowing down the topic in 2, clarifying one's position in 3, and finally, turning the topic into a specific claim in 4. While this "brainstorming" approach can be very helpful, it may not be clear to the inexperienced researcher how one should proceed from one step to another, or how the steps are connected. For example, the stance supplied in 3 ("local action supporting the Kyoto Protocol targets") seems to be an outcome or reaction (or possibly some kind of background information) rather than a step toward the claim that "humans are causing climate destruction and that human activity must be checked" in 4. Additionally, although Tajino et al. (2010) do not deal with observed phenomena, they do make mention of research questions, stating "it is useful to organize your paper around research questions" (80). However, they do not take up this topic in detail, nor do they show how research questions are related to thesis statements.

Other notable works (Swales 1990; Swales 2004; Swales and Feak 2012) seem to assume that the researcher already knows how to build a thesis statement. Within their models such as CARS ('Create-a-research-space' Model; Swales 1990) they focus on where research questions or hypotheses should be placed, such as in an introduction, but they do not discuss building or generating thesis statements.

This paper aims to fill this gap by first showing the elements a researcher needs in order to build a thesis statement (the "what" part) in section 2, and then what can happen – and does happen – if the researcher fails to produce these elements (the "why" part) in section 3. Finally, conclusions are presented in section 4.

It should be noted here that the author is a Tutorial Specialist at Mei-Writing, the Academic Writing Department at Nagoya University, Japan. Mei-Writing works to help graduate students publish their research ideas by helping them make their research projects clear and convincing. Mei-Writing helps students at all stages of the academic writing process from creating a research goal, to finalizing journal articles or doctoral theses. The propositions laid out in this paper are based on my experience working with graduate students at Mei-Writing. As such, when I refer to 'students', reference is specifically to students at the graduate level although the same basic principles should be applicable to any level.

## 2. The necessities for building a thesis statement

This paper is an attempt to help students or new researchers build their thesis statements from scratch. This section describes the elements that are needed to build a thesis statement.

As stated in the previous section, before a thesis statement is built, it is essential to have two elements: 1) an observed phenomenon, and 2) a research question.

**2.1 Observed phenomenon**

I will first examine what is meant by observed phenomenon. In the context of research, observed phenomena can largely be classified into four distinct types[1]. It is up to the researcher to decide which kind of phenomenon their research will be based on. The four types are:

1. Something observed in the natural/physical world
2. Something found in a previous work, a database, government data, etc.
3. A phenomenon that one finds in a (preliminary) lab experiment or survey results
4. A past hypothesis, now well supported or proven

I will now present concrete examples of the four types. The first, "something observed in the natural/physical world," is perhaps the most intuitive and is exactly as it sounds. An example of this is Darwin's Finches. Charles Darwin famously discovered that finches on the Galápagos Islands had considerably different beak shapes and sizes depending on the island on which they lived and its unique environment. Referring to this discovery (among others), Carl Zimmer (2010) points out, "Darwin did not realize the full importance of his observations until he returned to England in 1836. At the Galápagos Islands, for example, Darwin had collected a number of birds that had dramatically different beaks. Some had massive beaks good for crushing seeds, while others had slender needle-like beaks for feeding on cactus plants" (27). In short, this was Darwin's observed phenomenon in the natural world.

The second phenomenon, "something found in a previous work, a database, government data, etc." is important to all students, but especially to humanities and social science students, some of whom may not be able to make an observation in nature or be involved in field work. An example taken from an economics' journal (Gasparini and Cruces 2013) makes the following observation: "Both poverty and income inequality decreased significantly in Latin America during the first decade of the twenty-first century. At the start of the century, 25 out of every 100 people in the region were living on less than $2.50 per day; today, only 14 out of every 100 are in that situation" (51). Although the data or statistics could be from any number of sources, in this particular work, the authors note, "Unless explicitly mentioned, the source for all statistics cited in this paper is the Socio-Economic Database for Latin America and the Caribbean (SEDLAC), a joint project by CEDLAS and the World Bank (sedlac.econo.unlp.edu.ar)" (Gasparini and Cruces 2013, 61). These data about poverty and income inequality in Latin America can be a solid basis upon which the student could formulate a research question.

The third type is "a phenomenon that one finds in a (preliminary) lab experiment or survey results." The following is an actual example based on a tutorial session that I had with a student who was conducting research in cognitive psychology.[2] First, as background to the student's research the following observation was made: "In many western countries (e.g. USA, Britain, Italy, etc.) Imagined Contact is effective for reducing intergroup bias". However, based on the results of interviews conducted by the student in Japan, the same effect could not be found. This observed phenomenon formed the basis for further research which led to a significant result once "controlling for individual experiences" was taken into account.

The last phenomenon type is "a past hypothesis, now well supported or proven". One

[1] These four types are based on the author's observations and experience and may not be exhaustive. However, presenting these four possibilities to students seems to give them enough ideas to apply at least one of them to their own research.

[2] Some details have been changed to protect the identity of students and their work.

example of this is the Heliocentric Model of our solar system, which replaced the Geocentric Model. This hypothesis that the Sun is at the center of our solar system, instead of the Earth at the center, is trivial now that it has been proven, but this can be used as an observed phenomenon for further inquiry such as 'how the solar system formed in this manner', or 'whether other solar systems work in a similar way', etc.

The unifying feature of types 1-4 is offered in the following definition: An "observed phenomenon" is some (possibly trivial) truth that exists in the universe, be it natural/physical or social/cultural, and is put into one's mind using one or more of his/her senses. This definition provides the basis for how the term "observed phenomenon" is used in this paper.

Once a student has taken notice of or discovered some kind of observed phenomenon, they can then proceed to the next step: developing a research question.

**2.2 Research question**

After the student has a concrete foundation for their research project, the observed phenomenon, they can then proceed to the next step. This step is the research question: an inquiry of "how" or "why" that provides a research direction for the student.[3] The "how" question often indicates a mechanism which the student is interested in discovering, whereas the "why" question tends to indicate reasons, functions, factors or causes. Referring back to the example about Darwin's Finches in section 2.1, Zimmer (2010) continues, "Darwin was puzzled. Some naturalists of his day argued that species had been created where we now find them, well suited to their climate. But if the finches had all been created on the Galápagos Islands, why were they so different from one another" (27)? Upon asking himself questions such as this, Darwin at some point started formulating his own original claim or answer to this question, i.e. his thesis statement. "Darwin began to wonder if instead the birds had not been unchanging since creation. Perhaps they had evolved into their current forms. Darwin's finches helped lead him to conclude that all of life had evolved" (Zimmer 2010, 27). Likewise, once the student has an observed phenomenon, they can then make a "how" or "why" inquiry into the phenomenon itself. The student's answer to this research question is their thesis statement, which they can revise as many times as is necessary.

An example demonstrating this process can be seen in (1). This example is based on my own work in linguistics.

1) a. (Phenomenon) Japanese is borrowing abstract foreign loanwords for terms that already exist in the language (e.g. baransu/kinkou/tsuriai).

   b. (Research question: WHY) Why is Japanese borrowing abstract foreign loanwords for terms that already exist in the language? That is, what special function, if any, do they possess?

   c. (Thesis statement) Foreign terms that express abstract concepts compound productively with other native/Sino-Japanese/foreign elements to form new words or concepts. (This is something that perhaps the existing elements are not so good at doing.)

The thesis statement produced in (1)c may not be finalized, but it can be revised later as needed. At the very least, it is a clear claim that can be tested. Following these steps leads directly to the creation of a meaningful thesis statement because the sentence is based on an observed phenomenon, i.e. something that actually exists, and therefore is, in principle, a

[3] There are also interesting "what", "does" or "is" questions such as "What happens if…?" or "Does A react with B?" or "Is there an effect or not?", but these questions tend to have a straight forward answer such as "yes" or "no". Students who are interested in solving these types of questions tend to not have the same problems in formulating a thesis statement as those who are interested in the more open-ended "how" or "why" questions.

testable claim. The next section will discuss why these two elements are necessary for building a thesis statement by showing what can (and does) happen when one of these elements is missing or possibly mistaken for the thesis statement itself.

**3. When a paper lacks an observed phenomenon or a thesis statement**

Section 3.1 will first demonstrate what happens when a paper lacks a phenomenon, that is, the paper itself becomes a journey of discovering the phenomenon, and therefore lacks a conclusion. In these cases, only a summary of the paper is provided in the closing section, among other issues. I will then discuss in 3.2 the post hoc fallacies (the Texas sharpshooter and data dredging) that can arise when a paper has an observed phenomenon and a research goal, but no thesis statement, which often occurs because the objective is accidently construed as being a thesis statement, or the student relies on a data analysis program to find a pattern, which often finds one even if a meaningful pattern doesn't exist.

**3.1 The 'way of the phenomenon' paper**

In this section, I will demonstrate what often occurs when a paper lacks an observed phenomenon and thereby also lacks a research question and a thesis statement. When this happens, the paper usually becomes a journey of discovering the phenomenon itself, and concludes with statements of the discovered phenomena followed by a summary. I'll call these papers "the way of the phenomenon" papers. When students come to a tutorial session, I look for the following kinds of statements:

2) What are the similarities/differences between A and B?
3) This paper attempts to discuss/consider…
4) The relation between A and B is…
5) We will present an overview of…

The sentences in (2) - (5) can be extremely useful elements to have in a paper. However, if something similar to one of these sentences seems to be the closest I can get to finding a claim within the paper, that's a sign that it may need some reworking or improvements. Example (6) is based on the stated objective of the thesis paper of a student who was working in linguistics.

6) This paper makes a comparison of the Chinese expressions [A] and [B] and the Japanese [C] and [D], and attempts to make clear their similarities and differences.

I have edited the specific expressions in favor of alphabet characters (i.e. [A]), but the structure or basic idea of the sentence is the same. The sentence in (6) states that the author wishes to shed light on the similarities and differences of expressions in the two languages, but hasn't yet offered any of those similarities or differences. The paper, in many cases, will end up being a list of those similarities and differences, each one a phenomenon unto itself. Indeed, this paper spends nearly 70 pages describing observed phenomena, many of which would be an interesting independent research paper for deeper inquiry. For instance, one of the many phenomena the paper lists is in (7).

7) Most of the [C] examples that express spatial movement correspond to [A], but nearly all those which correspond to temporal/psychological movement do not.

The phenomenon in (7) seems like an interesting example from which to make an inquiry such as *why* it is the case that only spatial movement corresponds to [A], but temporal/psychological movement does not. Unfortunately, this does not happen in this paper. In fact, the whole structure of the paper is drastically altered because of this fundamental flaw. Figure 1 compares a fairly standard structure of an academic paper (on the left), to the structure of "the way of the phenomenon" papers (on the right).

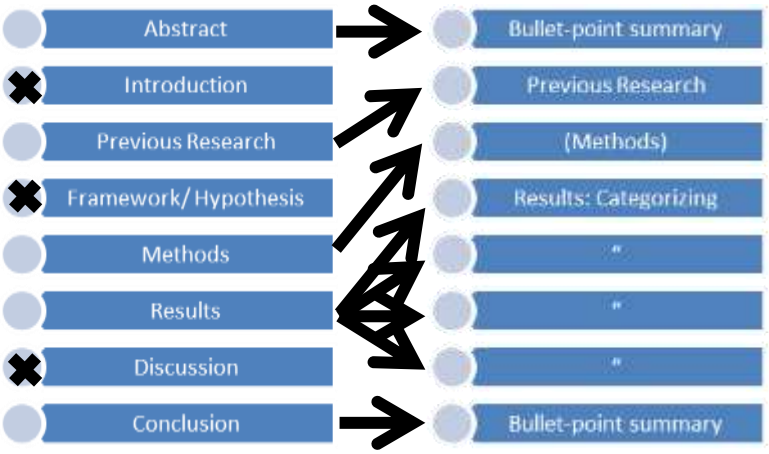



**Figure 1.** Comparison of the structure of a typical academic paper (left) and 'the way of the phenomenon' papers (right). X indicates elements that are typically missing in 'the way of the phenomenon' papers.

As Figure 1 shows, in many cases the way of the phenomenon papers omit much of the introduction, framework/hypothesis and discussion sections. The introduction section is omitted because it usually provides background into the research topic, much of which is observed phenomena. This section is used instead for presenting, in my experience, some previous research. The lack of a thesis statement explains the lack of not only the hypothesis section, but also the discussion section because there are no results to discuss or compare with one's thesis statement. Additional problems include having no conclusion because there is no claim. The end of the paper is a summary, in some cases bullet points, of the whole paper (i.e. all the listed categories). This happens because the body of the paper almost exclusively categorizes all the similarities and differences. The abstract also ends up being a near carbon copy of the end summary. This situation can be avoided if detected early on in the research process by making sure the student has an observed phenomenon and a research question.

### 3.2 The Texas Sharpshooter and Data Dredging

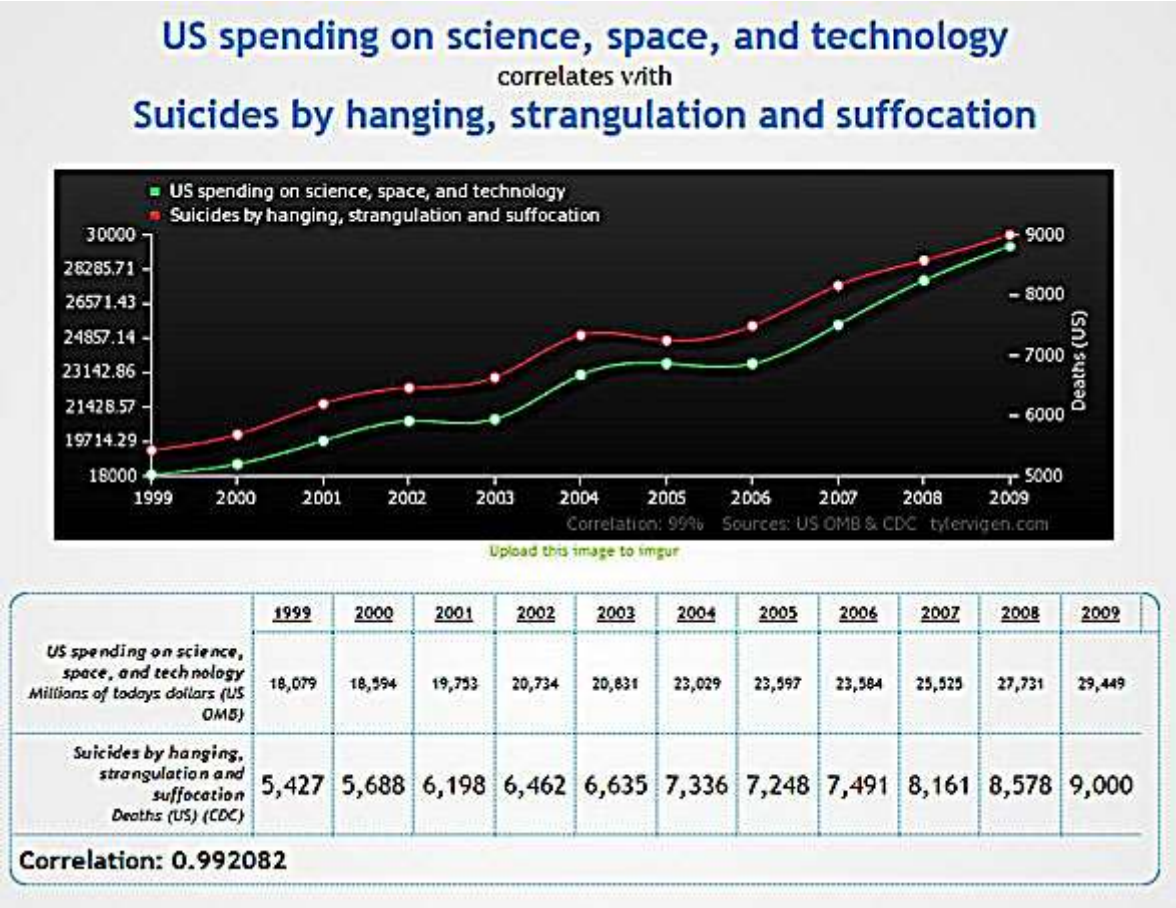
This section will bring attention to another case: when two elements are present, an observed phenomenon and research question, but the research question is stated as a goal/objective.

If there's a phenomenon, there's usually a research question. But in many cases, the research question is stated as a goal, and can be mistaken for a thesis statement. If the research question is left as a goal, and not turned into a question (at least temporarily for the sake of generating hypotheses), I find that students will be hesitant to postulate possible answers. However, if the student can restate their goal as a research question, it works as an effective hypothesis generating activity which the student can use to focus on one or two hypotheses before they continue into data collection or experimentation. Turning a research goal into a research question is generally quite straight forward as in (8).

8) a (Goal) This study aims to investigate the cause of the recent increase of Acute Leukemia in children.
   b (Research question) Why has there been a recent increase of Acute Leukemia in children?
   c (Possible thesis statement 1) Pre-natal exposure to ionizing radiation is the main cause in a recent increase of Acute Leukemia in children.
   d (Possible thesis statement 2) Work in occupations exposing mothers to chemicals such as benzene is the main cause in a recent increase of Acute Leukemia in children.

The goal stated as (8)a is turned into a "why" question in b. This question can then be used to generate possible thesis statements such as (8)c or d. Those claims may not turn out to be true, or may need substantial revision. The important point is to have a specific direction of the kinds of things which may be relevant to the research, while keeping in mind that it is also acceptable and useful to the scientific community if negative results are obtained.

If a student is not aware of the necessary elements presented in section 2 and their relation to each other, they could be put in a situation where they feel as though they have to force a conclusion after collecting all their data. The problem with this is obvious. The conclusion may be false or may just happen to correlate with something completely unrelated as in Figure 2.




| | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *US spending on science, space, and technology Millions of todays dollars (US OMB)* | 18,079 | 18,594 | 19,753 | 20,734 | 20,831 | 23,029 | 23,597 | 23,584 | 25,525 | 27,731 | 29,449 |
| *Suicides by hanging, strangulation and suffocation Deaths (US) (CDC)* | 5,427 | 5,688 | 6,198 | 6,462 | 6,635 | 7,336 | 7,248 | 7,491 | 8,161 | 8,578 | 9,000 |

**Correlation: 0.992082**

**Figure 2.** US spending on science, space, and technology correlates with suicides by hanging, strangulation, and suffocation. (Source: http://www.tylervigen.com/)

As Figure 2 shows, two completely unrelated sets of data points, regardless of whether there is a real connection, can appear to be correlated with each other. In this case, there is a 99% correlation between US spending on science, space, and technology with suicides by hanging, strangulation, and suffocation. Finding a connection where there isn't one isn't difficult to do, especially for sophisticated software programs that automatically find data clusters or patterns in the data, independent of the researcher's background knowledge of the subject.

One popular instantiation of this is known as the Texas sharpshooter fallacy, sometimes called the 'clustering illusion' (Carroll 2003, 77–78). The Texas sharpshooter fallacy is a type of *post hoc* explanation that sees meaning in a randomly clustered set of data points. The image the fallacy draws upon is that of a "sharpshooter" who randomly shoots many holes in the side of a barn, and then draws a target around a group of them that are densely clustered

together. This gives the illusion that the sharpshooter is an accurate shot. In reality, we would expect the sharpshooter to have a target set up before any shots are fired, in the same way a researcher would have some idea of what they are aiming for before collecting or analyzing their data.

There are a few high profile cases where researchers unknowingly fell victim to the Texas sharpshooter fallacy (Smith 2014; see also Thompson 2009). One such case is known as the "cancer cluster scare," where researchers were looking for a cause to explain cancer deaths of victims who were 19 years old or younger. The researchers noticed that many of the victims were living near large power lines, and subsequently drew the conclusion that living near these power lines was causing these deaths. Copious amounts of research later looking into this connection could not replicate its findings or find support for its conclusions (and is no longer taken to be a credible claim), but in the meantime it set off a needless national panic about living near large power lines (Smith 2014, 46–48). The researchers drew this conclusion because they were looking for a connection, any connection, and they found one. If the researchers had some other plausible explanations in mind before analyzing the data, the novel and unprecedented connection between cancer and power lines should have surprised them into making rigorous follow-up studies to see if the same results could be obtained. Unfortunately, this is not how the story unfolded.

Related to the sharpshooter fallacy is data dredging. Data dredging, as it sounds, is much more involved than the Texas sharpshooter. The Texas sharpshooter just needs to draw a target around the bullet holes. With data dredging, on the other hand, the bullet holes are not obvious (see Milloy 1996). Geoff Norman aptly put it as "data dredging leads to a lot of mud but very little gold" (Norman 2014, 5). This can occur essentially because the researchers have not formulated any hypotheses before doing other parts of research and therefore interpret any seemingly meaningful pattern as a conclusion. Although this case arguably bears a deeper connection to the importance of having a thesis statement before proceeding with other parts of research, it is also important to highlight it in the context of it often being mistaken for the research objective. In short, the problems outlined here can be largely avoided if the researcher first has 1) an observed phenomenon and 2) a research question, followed by possible hypotheses (thesis statements) based on 1) and 2).

## 4. Conclusion

This paper proposed in section 2 that the necessary elements for constructing a thesis statement are 1) an observed phenomenon, followed by 2) a research question. Students and new researchers who have these two elements will be able to generate meaningful thesis statements. In section 3, it was demonstrated that failure to have these two elements will produce the highly problematic "the way of the phenomenon" paper. Furthermore, failure to have a meaningful thesis statement before proceeding to data collection or analysis creates a risk of *post hoc* reasoning when interpreting results. The two *post hoc* fallacies discussed in 3.2 were the Texas sharpshooter and data dredging. This often occurs because the research objective is mistaken for the thesis statement. This paper proposed that these issues can be avoided by first having the two necessary elements for building a thesis statement.

## References

Carroll, Robert Todd. 2003. *The Skeptic's Dictionary: A Collection of Strange Beliefs, Amusing Deceptions, and Dangerous Delusions*. John Wiley & Sons.

Frank, Steven. 2005. "I'll Show You My Underwear (or How to Tackle the Dreaded Thesis Statement)." *Writing* 27 (4): 14–17.

Gasparini, Leonardo, and Guillermo Cruces. 2013. "Poverty and Inequality in Latin America:

A Story of Two Decades." *Journal of International Affairs* 66 (2): 51–63.
Milloy, Steven J. 1996. "Science without Sense: The Risky Business of Public Health Research." June. http://www.junksciencearchive.com/news/sws/sws-chapter5.html.
Norman, Geoff. 2014. "Data Dredging, Salami-Slicing, and Other Successful Strategies to Ensure Rejection: Twelve Tips on How to Not Get Your Paper Published." *Advances in Health Sciences Education: Theory and Practice* 19 (1): 1–5.
Smith, Gary. 2014. "How to Lie With Statistics." *Skeptic* 19 (4): 44–48.
Swales, John M. 1990. *Genre Analysis : English in Academic and Research Settings*. Cambridge University Press.
———. 2004. *Research Genres : Explorations and Applications*. Cambridge University Press.
Swales, John M., and Christine Feak. 2012. *Academic Writing for Graduate Students, 3rd Edition: Essential Tasks and Skills*. 3rd Edition. University of Michigan Press/ELT.
Tajino, Akira, David Dalsky, and Tim Stewart. 2010. *Writing for Academic Purposes*. Tokyo: Histuji Shobo.
Thompson, William C. 2009. "Painting the Target around the Matching Profile: The Texas Sharpshooter Fallacy in Forensic DNA Interpretation." *Law, Probability & Risk* 8 (3): 257–76.
Zimmer, Carl. 2010. *The Tangled Bank: An Introduction to Evolution*. Roberts and Co. Publishers.