International Forum
Vol. 24, No. 1
June 2021
pp. 5-24

FEATURE

Deconstructing Myths in Quantitative Research

John Wesley Taylor V

Abstract. *A myth can denote a prevalent but erroneous idea or belief. This article examines nine of these beliefs within quantitative research, that although popular, are ultimately misconceptions. Beyond highlighting the fallacy of the myth, positions are proposed in each case that align more closely with a scientific approach to research..*

Keywords: research, quantitative research, myth, interpretation, sample size, statistical significance, conclusions, overgeneralization, causation, inconclusive results, proof

Introduction

The word “myth” comes from the Greek μῦθος and referred to a traditional story (Oxford Living Dictionaries, 2018; Honko, 1984). Over time, however, the term acquired a second meaning, that of a popular but erroneous idea or belief. Taking this definition, we realize that myths have existed, perhaps since long ago, and some may still persist—myths of marine monsters that entrap ships, of strange creatures that walk like men. And we have contemporary myths, such as those that appear in movies.[1]

Could it be that myths also exist in research? Certain ideas that we might think are true but that, although popular, are ultimately wrong. Erroneous beliefs, like weeds, flourish with little cultivation. And while myths certainly abound amongst those who are recently initiated in the realm of research, it is possible that fallacies may still linger in the subconscious mind in an almost imperceptive format.

In this article, we will consider a number of these research myths, with some having been identified years ago but still present today (Campbell & Stanley, 1963; Robinson, 1970). We will seek to be proactive, not only countering myth with evidence of its fallacy but endeavoring to trace a clear delineation of realities that must ultimately replace misconceptions.

Myth #1: Research Answers Every Question

What is research? If we take a formal definition, research is a systematic inquiry, based on gathering and analyzing information, designed to develop or contribute to generalizable knowledge (U.S. Department of Health and Human Services, n.d.). "Systematic inquiry" means that we are asking questions and seeking answers to unknowns, all in an organized, purposeful manner. "Generalizable knowledge" points to the fact that we are looking for answers that extend beyond the immediate.

We can, however, ask different types of questions, such as the following:

1. What is the symbol for iron?
2. How well has Ben mastered the objectives of this course?
3. How can we convey the strategic priorities of our organization?
4. Should physician-assisted suicide be legalized?

We answer the first question through learning. The second is through an evaluation process. The third is by developing something, in this case, a strategy or process. The fourth is through an ethical or policy analysis. The key point is that research is only one approach to find answers to problems that we encounter. Only one method among several, but certainly an important method. At the same time, we must recognize that research, while not omnipotent, can assist us in considering many non-research questions. For example, on the topic of physician-assisted suicide, research can help us better understand the following aspects:

1. What are the attitudes and beliefs of patients and healthcare providers regarding assisted suicide?
2. What factors contribute to requests for an assisted suicide?
3. What is the experience of families whose loved ones participate in physician-assisted suicide? (Beyea & Nicoll, 1998).

We can conclude that, although research is not an appropriate method to answer every important question, it can still clarify the context of the various questions that we define as important (Alvesson & Sandberg, 2013; Dillon, 1984; McCay-Peet & Quan-Haase, 2017; Sheikh et al., 2011; Thomas, 2013; White, 2017).

Myth #2: Data Speak for Themselves

Consider that we encounter four points as they appear at the beginning of Figure 1. What do they represent? Obviously, the corners of a square. Or do they define a circle? Or perhaps something in between? Or do they stand for an X? Or an arrow? What do those four points truly represent?

Our conclusion is based on our interpretation of the data. In other words, the data do not speak for themselves. They must always be interpreted (Angrosino, 2007; Hofmann, 2005; Fottrell et al., 2007; Leonelli, 2016; Sartor, Del Riccio, Dal Poz, Bonanni, & Bonaccorsi, 2020). And that interpretation is influenced by our worldview, our way of seeing and making sense of what surrounds us (Hammersley, 2013). Two people, therefore, can look at the same data, the same information in a study, and see different patterns, because they view them from two different perspectives (see Figure 2). This does not necessarily mean that one of those interpretations is necessarily more valid than the other, although in some cases, arguments could be assessed as more robust within the context of a given discipline.

Myth #3: A Larger Sample is Better

It is often heard that the greater the number of participants in the sample, the better the study will be (Megerdichian, 2019; Sedlmeier, 2009; Spielman et al., 2019; Thompson & Liu, 2007; Verdery, Weir, Reynolds, Mulholland, & Edwards, 2019). To some extent, that position makes sense. Many shortcomings in the research stem from the lack of an adequate sample size, which does not allow us to draw valid conclusions. (We will comment further on that point in a subsequent myth.)

We must recognize, however, that as we continue to increase the sample size, we arrive at a point of diminishing returns, in which what we receive from the process is less than what we invest in it. In essence, we lose efficiency without increasing effectiveness. Perhaps we can illustrate the problem. Imagine that we are sitting in a boat on a small lake. The greatest depth of the lake is 10 meters and the visibility from the surface is 5 meters. From the boat, we can look down and perceive certain characteristics of the bottom of that lake, contours that are evident between the surface and five meters deep. But we cannot perceive what the topography of the lake is like beyond that depth.

We only have one tool at our disposal to continue that exploration. We can turn on a pump that can draw water from the lake; and in doing so, the water level begins to drop. Sitting in the boat on the surface and always with a visibility of five meters, we now begin to distinguish new characteristics of the bottom of the lake, aspects that we did not see before. We continue to pump water from the lake until the surface is five meters above the deepest point. We can now distinguish all the important features of the topography of the bottom of that lake. We have made important discoveries. We know now what we did not know before.

But imagine that we continue the project of pumping out the water. Although we could further extract a good amount of water from the lake, what we see regarding its topography no longer presents us with important new information. The water pumped from the lake represents the sample. With a very small sample, we discover almost nothing beyond that which we already knew intuitively. Our "findings" are nothing more than those obvious things that we knew before starting the research, which was already visible from the surface. But as sample size increases, we move beyond the obvious and begin to discover things of importance, findings that were not detected from our position on the original surface of the lake. There comes a point, however, where what is truly important has already been identified and the small details that are now revealed are not of great consequence. We have moved from the important to the trivial. And when that happens, we have lost a positive yield. In a research study, which should be parsimonious, we have lost efficiency.

Furthermore, we must recognize that sample size is only one part of the equation. Representativeness is equally important. If we increase the size of a non-representative sample, part of another myth that we shall explore, our results will never become more accurate. But assuming for the moment that the sample is representative, what then should be its ideal size? That question leads us to the next myth.

Myth #4: Minimum Acceptable Sample Size is a Magic Number

Sometimes, in an endeavor to facilitate the process, some individuals indicate that the sample should be a certain proportion of the population, for example, 10%. Let us put aside, for the moment, the problem that we often do not know the size of the population. If the population is 10 people, interviewing one person (i.e., 10%) would obviously be too few to draw reliable conclusions. And if the population is 10,000 people, interviewing 1,000 people (again, 10%) would be needlessly high, and we will have spent much time and effort without significantly increasing the trustworthiness of important results.

In a desire to speed up the process, others take the position that the minimum sample size is a certain fixed number of participants, such as 100 persons. However, if the population is 50 people, trying to get a sample of 100 will be impossible. And if the population is 10,000 people, a sample of 100 will be statistically significant only for "findings" of what was already obvious without conducting the study.

What, then, should we do to determine the minimum required sample size? In short, we conduct a power analysis (Bujang & Adnan, 2016; Kyriazos, 2018; Lan & Lian, 2010; Lerman, 1996; MacCallum, Browne, & Sugawara, 1996; Schoemann, Boulton, & Short, 2017; Tomczak, Tomczak, Kleka, & Lew, 2014). In any study, there are four interrelated factors linked in such a way that when any three of them are defined, the fourth is determined as a result. These four factors are α, β, γ and n.

Consequently, to determine n, the minimum sample size, we need to set α, β, and γ. The factors α and β have to do with the probability of making a mistake of stating a conclusion incorrectly. And γ has to do with what the researcher considers to be important, as opposed to trivial.

Let us first consider the decision errors, whose limiting probabilities are represented by α and β. To do so, let us use an analogy—the decision to carry or not to carry an umbrella when heading out for the day. If I decide to bring an umbrella and it then rains, I made the right decision. But if I decide to carry an umbrella and then it does not rain, I have made a decision error (a Type I error, whose probability is related to α). Alternately, I can decide not to carry an umbrella. If it does not rain, I am fine. But if it rains, I have again made a decision error (a Type II error, whose probability is related to β).

Let us change the metaphor. If I am a physician and a patient arrives having a certain pathology, I must make a decision: Should I give a certain medicine or not? If I prescribe the medication and it is effective, I have done well. But if I prescribe the drug and that drug is not effective, I have made a decision error, with the needless expense and possible side effects. On the other hand, if I decide not to prescribe the drug and that drug would not have had any effect, I made the right decision. But if I did not prescribe the medication and it would have had a positive effect, I made another type of decision error.

In a research study in which we propose hypotheses, there are also potential errors. If, for example, we decide to reject the null hypothesis (and retain the alternate) when we should not have, we have made a Type I error (a false positive). And if we decide not to reject the null hypothesis, when in fact we should have, we have made a Type II error (a false negative). In research of this nature, α and β are the probabilities that we have defined as the acceptable limits of committing Type I and Type II errors.

In a study we want both α as well as β to be small probabilities (for example, no more than 5%), but we must also recognize that they can never be set at a probability of zero because if they were, we could never reach that probability, and consequently, we could never reach a conclusion. (More about that later.) We must also recognize that α and β are inversely related. That is, in trying to decrease the probability of making one, we increase the probability of the other.

What does this all mean? Alpha has to do with statistical significance, with our level of confidence. Generally, α is set at .05 or .01. That is, we can have a reasonable 95% or 99% confidence in stating that there is a difference or that there is a relationship. Beta has to do with power, with the sensitivity of research. Generally, power (1-β) is set at .80 or .90. That is, we can be reasonably sure of detecting a truth that is at least as great as the effect size (γ) that we have determined.

And that brings us to the third factor, effect size. Gamma has to do with meaningfulness. It is the minimum difference or the minimum relationship that will be of importance, of practical value to us. For example, a minimum difference can be defined in terms of a proportion of σ. For example, a γ value defined as .3σ indicates that we have defined as important any difference between means greater than about a third of a standard deviation. If we are talking about relationships, the minimum relationship that we consider important can be defined in terms of variance explained. A γ defined as $r = .30$, for example, represents an r^2 equal to .09, which means that any *r-value* that represents more than 9% of the variance we will consider as important.

How then do we determine the minimum effect size? Simply by deciding what is the minimum difference or the minimum relationship that will be important to us. And we make that decision based on theory, prior research, experience, and logic. There are, of course, statistical packages that can help us perform this analysis. For a question of difference, for example, the minimum sample size (n) required in a two-tailed test, with an infinite (or unknown) population and a power of .90, would be as shown in Table 1.

Table 1

Sample Size Required for Comparative Studies According to Significance Level and Effect Size

Significance	**Effect size**				
level	$\gamma = .5\sigma$	$\gamma = .3\sigma$	$\gamma = .25\sigma$	$\gamma = .2\sigma$	$\gamma = .1\sigma$
$\alpha = .05$	44	119	170	264	1050
$\alpha = .01$	63	168	241	374	1487

Differences of .5s or greater would probably be rather obvious before conducting a study. On the other side of the continuum, differences of .1s or less may not represent a difference of practical importance. In between, we find those findings that are neither obvious nor trivial but rather of practical importance. Cohen (1977), for example, defined such approximate effect sizes as large, medium, and small. Similarly, for a question of relationship, the required minimum sample size would be as shown in Table 2. For a given population, then, the smaller α, the smaller β (and the greater the power, 1-β), and the smaller γ (the effect size that we consider important), the larger must be n, the minimum required sample size.

Table 2

Sample Size Required for Correlational Studies According to Significance Level and Variance Explained

Significance level	Correlation (variance explained)				
	r = .5 (25%)	*r* = .4 (16%)	*r* = .3 (9%)	*r* = .2 (4%)	*r* = .1 (1%)
$\alpha = .05$	38	62	113	258	1044
$\alpha = .01$	53	86	158	364	1477

In summary, the required sample size is not a magic number, such as a set number for all studies or a proportion of the population. Rather, it is a specific number determined through power analysis, where we seek to balance α and β, while taking γ into account.

Myth #5: If Something is Statistically Significant, It is Important

Stanley Milgram (1967) proposed the theory of six degrees of separation, which indicates that every person in the world is not more than six relationships removed from any other person. With the advent of social networks, it is estimated that there are now only four degrees of separation (Backstrom, Boldi, Rosa, Ugander, & Vigna, 2012). Conclusion? Everything is interconnected. Everything is related.

At the same time, it is logical to recognize that no two things are identical in all their characteristics. It is impossible, for example, for two entities to have the same form and function, occupying the same place at the same time. Conclusion? Every entity is somewhat different from every other entity. In other words, no two things are completely alike in the real world.

Therefore, everything that exists is in some way related to, but different from, every other entity. Consequently, if we increase sample size towards infinity, two variables will always give evidence of a relationship that is statistically significant. And the difference between the two categories of a variable will always be statistically significant. Why, then, do we conduct research if everything is related and everything has a difference?

We must first remember that statistical significance has to do with confidence, with the fact that we want to be quite certain that results are trustworthy. However, not every difference nor every relationship is of practical importance. In evaluating an experimental teaching strategy with a large sample, for example, we may be able to identify a 0.1% performance difference as statistically significant. But this is not likely to be a meaningful difference in terms of educational value or the cost of the program.

It is true that the p-value (i.e., the probability of Type I Error) and the effect size (γ) have a certain relationship. Statistical significance is necessary, but insufficient, to enable us to speak of importance (Cano-Corres Sanchez-Alvarez, & Fuentes-Arderiu, 2012; Khalilzadeh & Tasci, 2017; Nakagawa & Cuthill, 2007; Thompson, 1998; Van Calster, Steyerberg, Collins, & Smits, 2018). In other words, it is necessary to first establish that something is trustworthy in order to talk about its importance. But the mere fact that something is trustworthy ($p < .05$, for example) does not necessarily imply that it is of practical importance (Cutter, 2020; Fan, 2010; Gliner, Vaske, & Morgan, 2010; Kalinowski & Fidler, 2010; Peeters, 2016; Rothman, 2016; Sullivan & Feinn, 2012; Vacha-Haase, Nilsson, Reetz, Lance, & Thompson, 2000).

The trustworthiness of a result and the importance of that result are simply two different things. With large samples, the vast majority of relationships or differences examined are likely to be statistically significant; but without considering effect size, one cannot state that something is important. Given that statistical significance does not equal importance, we must be careful not to allow the fascination of accurately measuring minutiae to distract us from what is truly important.

Myth #6: If There is No Statistical Significance, I can Conclude That There is No Difference or No Relationship

If I did not find a statistically significant difference, can I declare that there is no difference? And if I did not identify a statistically significant relationship, can I then declare that there is no relationship?

As we have mentioned, statistical significance gives us confidence that the results are trustworthy. Therefore, when statistical significance is not reached, we do not have confidence in the fidelity of the results.

When a statistical process does not return a probability of Type I error less than the error limit we have defined (that is, $p < \alpha$), what can we conclude? Nothing! No conclusion can be made regarding a matter if the findings on which that conclusion would be based are not trustworthy. What shall we say then? The only thing we can state is that "the results are inconclusive." This means that we know nothing more now than before we started the research.

What is the reason for this sad situation? If we properly conducted the power analysis to determine the minimum number required for the sample and if we obtained that sample size, the statistical processes will give us a good probability (e.g., a power of at least 90%) that the relationships or differences larger than γ (the minimum effect size that will be of practical importance) have been detected. The "inconclusive results" situations are common because the sample was too small, less than the minimum that a power analysis would have indicated (Blume, McGowan, Dupont, & Greevy 2018; Flather, Farkouh, Pogue, & Yusuf, 1997; Hawkins, 1990; Hernandez, 2021; Miladinovic, 2013). The solution? For future research, be sure to obtain the minimum sample size indicated by a power analysis.

Myth #7: If I Find a Significant Difference Between A and B, and A is Greater Than B, I Can Say That A is Greater Than B Overall

For example, when I find in my study a statistically significant difference between the linguistic ability of boys and girls, favoring girls, will I be able to declare that girls, in general, have a greater linguistic ability than boys? When I obtain a sample, that sample must be representative of the given population. The best way to ensure that it is representative in its various aspects is by having a sufficient number in the sample, as we have discussed, and by using a random process to derive that sample from the population. However, we should point out that what we refer to as "the population" is really the sampling frame (all preschool children in a certain town, for example), from which we derive the sample.

By having a representative sample, I can then confidently generalize my conclusions to that sampling frame (see Figure 3). The problem, however, is that many times I would like to generalize to the total population, that is, to all the preschool children that exist. But that is something I cannot do, simply because I do not know if the preschool children in that certain town are representative of all preschool children worldwide. And there is probably no way of knowing.[2]

If I endeavor to generalize to a total population, of which my sample is not representative, I have made an over-generalization error (Bamber, Christensen, & Gaver, 2000; Ercikan & Roth, 2009; Ercikan & Roth, 2014; Pham & Triantaphyllou, 2008; Wiernik, Raghaven, Allan, & Denison, 2020). What then can I say? If the sample was representative, I can only state that A is greater than B in the population (i.e., the sampling frame) of my study. What can I not say? That A is greater than B, for all the A and B in general.

By the way, what if I could not obtain a random sample?[3] Then, the sample is not representative of anything, and I cannot generalize. I can only comment on the characteristics of that group of individuals who participated in the study. But that does not make much sense because it does not allow me to extend the findings at all.

And a research study should result, to at least some degree, in generalizable knowledge.

Myth #8: If I Discover a Relationship Between X and Y, I Can Say That X has an Effect on Y

If I find that 99% of heroin addicts drank milk as children, can I conclude that drinking milk leads to addiction? If I find that there is a strong relationship between years of being married and the probability of death, can I conclude that marriage is the cause of death? Obviously, those conclusions seem outlandish, but only because we know better. But what happens when we do not know?

A correlation is necessary but insufficient to establish causation. In other words, all causality must be based on an underlying correlation, but we cannot conclude causality simply because we have found a correlation (Ahn & Bailenson, 1996; Peyrot, 1996; Sassower, 2017; Tindale, 2007). Why is this the case?

First, because often, we do not know the direction of the relationship. Imagine that for the first time, we come across a windmill. Looking closely, we realize that the speed of rotation of the windmill and the speed of the wind correlates. A potential conclusion, based on the data, could be that windmills are huge fans that produce wind. Or, to switch the illustration, imagine that we come upon a car that has hit a tree. The vehicle has a blown tire. Did the crash produce the blowout? Or could it be that the blowout resulted in the wreck? The key point is that arriving after the fact, either option could be viable.

Another consideration is the possible influence of a third variable. In a study, for example, we find that as ice cream sales rise and fall over the course of the year, the death rate due to drowning also similarly rises and falls. We might be tempted to conclude that ice cream consumption precipitates drowning. In reality, however, both indices increase and decrease because we are entering and leaving the summer, where the change in environmental temperature influences both the consumption of ice cream, as well as the probability that people are swimming and therefore, the possibility of drowning.

Given these circumstances, what can we say when we find a correlation? We can state that X is related to Y. Or simply that there is a correlation between X and Y, or that X is associated with Y. What we cannot declare is that one of the variables precipitates leads to or produces changes in the other variable. Nor can we speak of impact, effect, or influence. Nor regarding result, outcome, or consequence.

Myth #9: Research Proves, Research has Confirmed

If we want certainty, that is, total confidence, we must set $\alpha = .000$. And if we set α at .000, we always will find that p will be greater than α, indicating that there is no statistical significance. Therefore, as we have seen, the only thing we can say is that the results are inconclusive. If we set α to some other value, such as .05, then there is a 5% chance that our conclusions could be incorrect, even when we achieve statistical significance. As a result, in research, nothing is ever proved. Scientifically, we live in a world of probabilities and possibilities, not certainties (Weiss, 2003).

Research findings may, indeed, lend support to a premise or research hypothesis. But a theory is never fully supported. In fact, a successful theory is merely one that has been tested and has escaped being discarded, for now (Robinson, 1970).

To put it another way, the results of the research probe but do not prove a theory. There is never a final word in research, only the latest word, which we hope points in the right direction. Just by tracing the winding history of science, we realize that research has no absolute answers (DeWitt, 2018).

Conclusion

With so many myths in research that can entrap us, what can we do? The fundamental problem often lies in our innate desire to be something beyond who we really are, to say something more comprehensive and authoritative than what we can legitimately say.

Our circle of knowledge is surrounded by the vast universe of our ignorance, by all that we do not know. The problem is that, for most of that universe, we do not know that we do not know. In fact, the only things that we realize that we do not know are the points in that universe that touch the perimeter of our circle of our circle of knowledge (see Figure 4).

When the circle is small, the perimeter is also small. And we are tempted to think that, although there are still a few things that we do not know, we already know almost everything. However, as we learn new things, including through research, the circle of our knowledge grows. But the perimeter is also expanding. And when that happens, we start to come into contact more and more with that which we still do not know.

So, research invariably raises more questions than it answers. And that is why, the more we learn, the more we realize how much we do not know. And the more humble and teachable we are to become.

Figures

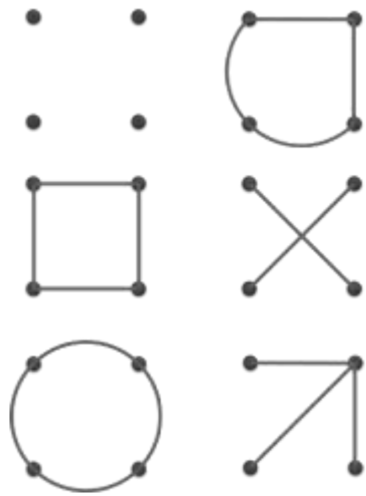

Figure 1. Interpreting data.

Figure 2. The same data, two perspectives.
Source:http://www.wolfescape.com/Humour/NonMedPicts/
MarriageBeforeAfter.gif

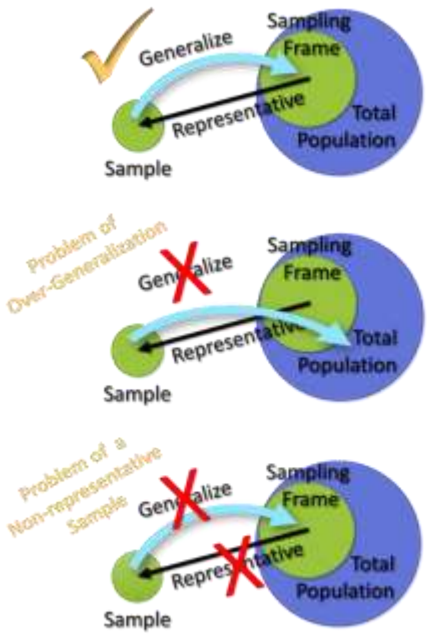

Figure 3. The problems of generalization.

.

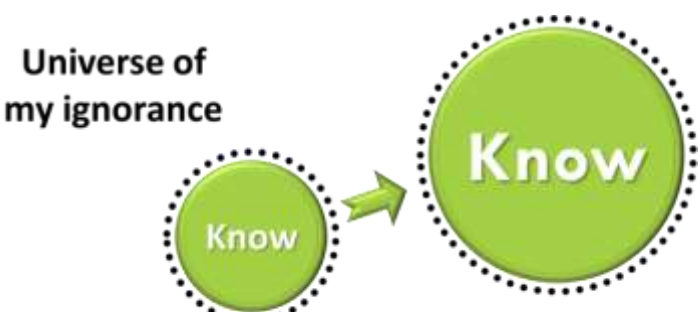

Figure 4. What we know and do not know

Notes

1. For example, the Star Wars sequels: “The Force Awakens,” “The Last Jedi,” and “The Rise of Skywalker.”

2. Some researchers try to address this problem by stating something like this: “To the extent that the sampling frame in this study may be representative of the population overall, the conclusions of the study may be applicable to the general population.” That stance, however, places the responsibility of generalization on the reader, who is not in any better position to determine whether or not the sampling frame is, in fact, representative of the global population. Given this situation, some (such as Futoma, et al., 2020) have proposed that generalizability itself should be called into question.

3. Some propose that a stratified sample can be used to ensure that the sample is representative of a population. However, one can establish correspondence (or proportions) only on a very limited number of factors. Consequently, it is still unknown whether the sample is representative in terms of other factors not considered in the stratification process, which could be important for the topic under study.

Note: This is a companion article to the previously published essay by the same author, “Deconstructing myths in qualitative research” (*International Forum Journal*, *20*(1), 18-30), available at https://journals.aiias.edu/iforum/article/view/323.

References

Ahn, W., & Bailenson, J. (1996). Causal attribution as a search for underlying mechanisms: An explanation of the conjunction fallacy and the discounting principle. *Cognitive Psychology, 31*(1), 82-123. doi:10.1006/cogp.1996.0013

Alvesson, J., & Sandberg, J. (2013). *Constructing research questions: Doing interesting research*. London, England: Sage.

Angrosino, M. (2007). *Doing ethnographic and observational research*. London, England: Sage. doi:10.4135/9781849208932.n6

Backstrom, L., Boldi, P., Rosa, M., Ugander, J., & Vigna, S. (2012). Four degrees of separation. In *Proceedings of the 4th Annual ACM Web Science Conference* (pp. 33-42). New York, NY: ACM.

Bamber, L., Christensen, T., & Gaver, K. (2000). Do we really 'know' what we think we know? A case study of seminal research and its subsequent overgeneralization. *Accounting, Organizations, and Society*, *25*(2), 103-129. doi:10.1016/S0361-3682(99)00027-6

Beyea, S. C., & Nicoll, L. H. (1998). Dispelling the myth that research answers all questions. *AORN Journal*, *68*(6), 1044-1047. doi:10.1016 /S0001 -2092(06)62145-5

Blume, J., McGowan, L., Dupont, W., & Greevy, R. (2018). Second-generation p-values: Improved rigor, reproducibility, and transparency in statistical analyses. *Plos One*. doi:10.1371/journal.pone.0188299

Bujang, M., & Adnan., T. (2016). Requirements for minimum sample size for sensitivity and specificity analysis. *Journal of Clinical and Diagnostic Research*, *10*(10), YE01-YE06. doi:10.7860/JCDR/2016/18129.8744

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental design for research in teaching. In N. L. Gage (Ed.), *Handbook of research in teaching: A project of the American Educational Research Association* (Chapter 5). Chicago, IL: Rand McNally.

Cano-Corres, R., Sanchez-Alvarez, J., & Fuentes-Arderiu, X. (2012). The effect size: Beyond statistical significance. *Journal of the International Federation of Clinical Chemistry and Laboratory Medicine*, *23*(1), 19-23.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.

Cutter, G. (2020). Effect size or statistical significance, where to put your money. *Multiple Sclerosis and Related Disorders*, *38*, 101490. doi:10.1016/j.msard.2019.101490

DeWitt, R. (2018). *Worldviews: An introduction to the history and philosophy of science*. Hoboken, NJ: John Wiley.

Dillon, J. (1984). The classification of research questions. *Review of Educational Research*, *54*(3), 327-361. doi:10.3102/00346543054003327

Ercikan, K., & Roth, W. (Eds.). (2009). *Generalizing from educational research: Beyond qualitative and quantitative polarization*. New York, NY: Routledge.

Ercikan, K., & Roth, W. (2014). Limits of generalizing in education research: Why criteria for research generalization should include population heterogeneity and uses of knowledge claims. *Teachers College Record, 116*(5), 1-28.

Fan, X. (2010). Statistical significance and effect size in education research: Two sides of a coin. *The Journal of Educational Research*, *94*(5), 275-282. doi:10.1080/00220670109598763

Flather, N., Farkouh, M., Pogue, J., & Yusuf, S. (1997). Strengths and limitations of meta-analysis: Larger studies may be more reliable. *Controlled Clinical Trials*, *18*(6), 568-579. doi:10.1016/S0197-2456(97)00024-X

Fottrell, E., Byass, P., Ouedraogo, T., Tamini, C., Gbangou, A., Sombié, I., . . . Graham, W. (2007). Revealing the burden of maternal mortality: A probabilistic model for determining pregnancy-related causes of death from verbal autopsies. *Population Health Metrics*, *5*, 1. doi:10.1186/1478-7954-5-1

Futoma, J., Simons, J., Panch, T., Doshi-Velez, F., & Celi, L. (2020). The myth of generlisability in clinical research and machine learning in health case. *Lancet Digital Health, 2,* 489-492. doi:10.1016/S2589-7500(20)30186-2

Gliner, J., Vaske, J., & Morgan, G. (2010). Null hypothesis significance testing: Effect size matters. *Human Dimensions of Wildlife*, *6*(4), 291-301. doi:10.1080/108712001753473966

Hammersley, M. (2013). *The myth of research-based policy and practice*. London, England: Sage.

Hawkins, D. (1990). Clinical trials—meta-analysis, confidence limits, and “intention to treat” analysis. *Journal of Obstetrics and Gynaecology*, *10*(4), 259-260. doi:10.3109/01443619009151189

Hernandez, H. (2021). Optimal significance level and sample size in hypothesis testing: Large samples. *ForsChem Research Reports, 6*, 2021-08.

Hofmann, T. (2005). From bits and bytes to information and knowledge. *Proceedings of the 14th ACM International Conference on information and knowledge management.* doi:10.1145/1099554.1099557

Honko, L. (1984). The problem of defining myth. In A. Dundes (Ed.). *Sacred narrative: Readings in the theory of myth*. Los Angeles, CA: University of California Press.

Kalinowski, P., & Fidler, F. (2010). Interpreting significance: The differences between statistical significance, effect size, and practical importance. *Newborn and Infant Nursing Reviews*, *10*(1), 50-54. doi:10.1053/j.nainr.2009.12.007

Khalilzadeh, J., & Tasci, A. (2017). Large sample size, significance level, and the effect size: Solutions to perils of using big data for academic research. *Tourism Management*, *62*, 89-96. doi:10.1016/j.tourman.2017.03.026

Kyriazos, T. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology*, *9*(8), 2207-2230. doi:10.4236/psych.2018.98126

Lan, L., & Lian, Z. (2010). Application of statistical power analysis: How to determine the right sample size in human health, comfort, and productivity research. *Building and Environment*, *45*(5), 1202-1213. doi:10.1016/j.buildenv.2009.11.002

Leonelli, S. (2016), The philosophy of data. In L. Floridi (Ed.). *The Routledge handbook of philosophy of information* (pp. 191-202). New York, NY: Routledge.

Lerman, J. (1996). Study design in clinical research: Sample size estimation and power analysis. *Canadian Journal of Anaesthesia, 43*, 184-191. doi:10.1007/BF03011261

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130-149.

McCay-Peet, L., & Quan-Haase, A. (2017). What is social media and what questions can social media research help us answer? In Sloan, L, & Quan-Haase, A. (Eds.). *The SAGE handbook of social media research methods* (pp. 13-26). London, England: Sage.

Megerdichian, A. (2019). Economic analysis of optimal sample size in legal proceedings. *International Journal of Statistics and Probability*, *8*(4), 13-24. doi:10.5539/ijsp.v8n4p13

Miladinovic, G., Mhaskr, R., Hozo, I., Kumar, A., Mahonoy, H., & Djulbegovic, B. (2013). Optimal information size in trial sequential analysis of time-to-event outcomes reveals potentially inconclusive results because of the risk of random error. *Journal of Clinical Epidemiology*, *66*(6), 654-659. doi:10.1016/j.jclinepi.2012.11.007

Milgram, S. (1967). The small world problem. *Psychology Today, 1*(1), 61-67.

Nakagawa, S., & Cuthill, I. (2007). Effect size, confidence interval, and statistical significance: A practical guide for biologists. *Biological Reviews*, 82(4), 591-605. doi:10.1111/j.1469-185X.2007.00027.x

Oxford Living Dictionaries. (2018). *Myth*. Accessed at https://en .oxforddictionaries.com/definition/ myth

Peeters, M. (2016). Practical significance: Moving beyond statistical significance. *Currents in Pharmacy Teaching and Learning*, *8*(1), 83-89. doi:10.1016 /j.cptl.2015.09.001

Peyrot, M. (1996). Causal analysis: Theory and application. *Journal of Pediatric Psychology, 21*(1), 3-24. doi:10.1093/jpepsy/21.1.3

Pham, H., & Triantaphyllou, E. (2008). The impact of overfitting and overgeneralization on the classification accuracy in data mining. In O. Maimon & L. Rokach (Eds.). *Soft computing for knowledge discovery and data mining* (pp. 391-431). doi:10.1007/978-0-387-69935-6_16

Robinson, G. (1970). Exploring some myths in research design. *The Journal of School Health*, *40*(7), 335-338. doi: 10.1111/j.1746-1561.1970.tb05618.x

Rothman, K. (2016). Disengaging from statistical significance. *European Journal of Epidemiology, 31,* 443-444. doi:10.1080/00031305.2016.1154108

Sartor, G., Del Riccio, M., Dal Poz, I., Bonanni, P. & Bonaccorsi, G. (2020). COVID=10 in Italy: Considerations on official data. *International Journal of Infectious Diseases*, *98*, 188-190. doi:10.1016/j.ijid.2020.06.060

Sassower, R. (2017). Causality and correlation. In B. Turner, B. (Ed.), *The Wiley Blackwell Encyclopedia of Social Theory*. John Wiley. doi:10.1002 /9781118430873.est0585

Schoemann, A., Boulton, A., & Short, S. (2017). *Social Psychological and Personality Science*, *8*(4). doi:10.1177/1948550617715068

Sedlmeier, P. (2009). *Improving statistical reasoning: Theoretical models and practical implications*. Mahwah, NJ: Lawrence Erlbaum Assoc.

Sheikh, K., Gilson, L., Agyepong, I., Hanson, K., Ssengooba, F., & Bennett, S. (2011). Building the fiel of health policy and systems research: Framing the questions. *PLoS Med, 8*(8), e1001073. doi:10.1371/journal.pmed.1001073

Spielman, R., Dumper, K., Jenkins, W., Lacombe, A., Lovett, M., & Perlmutter, M. (2019). *Psychology: OpenStax*. Retrieved from https://opentextbc.ca /psychologyopenstax/

Sullivan, G., & Feinn, R. (2012). Using effect size—or why the p value is not enough. *Journal of Graduate Medical Education*, *4*(3), 279-282. doi:10.4300/JGME-D-12-00156.1

Thomas, G. (2013). The myth of rational research. *British Educational Research Journal, 24*, 141-161. doi:10.1080/0141192980240203

Thompson, B. (1998). Statistical significance and effect size reporting: Portrait of a possible future. *Research in the Schools*, *5*(2), 33-38.

Thompson, P., & Liu, Y. (2007). Intricacies of statistical inference and teachers' understandings of them. In J. Lovett & P. Shah (eds.). *Thinking with data*. New York, NY: Psychology Press. doi:10.4324/9780203810057

Tomczak, M., Tomczak, E., Kleka, P., & Lew, R. (2014). Using power analysis to estimate appropriate sample size. *Trends in Sport Sciences*, *4*(21), 195-206.

Tindale, C. (2007). *Fallacies and argument appraisal*. New York, NY: Cambridge University Press.

U.S. Department of Health and Human Services, The Office of Research Integrity. (n.d.). *Module 1: Introduction: What is research?* Retrieved from https://ori.hhs.gov/content/module-1-introduction-what-research

Vacha-Haase, T., Nilsson, J., Reetz, D., Lance, T & Thompson, B. (2000). Reporting practices and APA editorial policies regarding statistical significance and effect size. *Theory and Psychology*, *10*(3), 413-425. doi:10.1177/0959354300103006

Van Calster, B., Steyerberg, E., Collins, G., & Smits, T. (2018). Consequences of relying on statistical significance: Some illustrations. *European Journal of Clinical Investigation*, *48*(5), e12912. doi:10.1111/eci.12912

Verdery, A., Weir, S., Reynolds, Z., Mulholland, G., & Edwards, J. (2019). Estimating hidden population sizes with venue-based sampling. *Epidemiology*, *30*(6), 901-910.

Weiss, C. (2003). Expressing scientific uncertainty. *Law, Probability, and Risk*, *2*(1), 25-46. doi:10.1093/lpr/2.1.25

White, P. (2017). *Developing research questions*. London, England: Palgrave.

Wiernik, G., Raghaven, J., Allan, T., & Denison, A. (2020). Generalizability challenges in applied psychological and organizational research and practice. *Brain and Behavior Sciences*. doi:10.1017/S0140525X20001685

John Wesley Taylor V, PhD
Associate Director of the Seventh-day Adventist Education System
General Conference, Silver Spring, Maryland, USA
taylorjw@gc.adventist.org