OPEN ACCESS Top-Quality Science
Freely available online

Review of Communication Research
2013, Vol.1, No. 1, 69-84
doi: 10.12840/issn.2255-4165_2013_01.01_003

Quantitative Communication Research: Review, Trends, and Critique

Timothy R. Levine
Michigan State University, MI, USA
levinet@msu.edu

Highlights

- Trends in quantitative communication research are reviewed with a content analysis of articles published in leading journals.
- Health communication and computer mediated communication have become increasingly popular topic of communication research.
- Null hypothesis significance testing remains the dominant approach to statistical inference.
- Survey research and close-ended, self-report measurement are the most common methods in quantitative communication research.
- Hypotheses and findings involving statistical mediation and moderation have become increasingly common.
- Reporting the shapes of distributions, estimates of statistical power, and confidence intervals, unfortunately, remain uncommon.

Abstract

Trends in quantitative communication research are reviewed. A content analysis of 48 articles reporting original communication research published in 1988-1991 and 2008-2011 is reported. Survey research and self-report measurement remain common approaches to research. Null hypothesis significance testing remains the dominant approach to statistical analysis. Reporting the shapes of distributions, estimates of statistical power, and confidence intervals remain uncommon. Trends over time include the increased popularity of health communication and computer mediated communication as topics of research, and increased attention to mediator and moderator variables. The implications of these practices for scientific progress are critically discussed, and suggestions for the future are provided.

Suggested citation: Levine, T. R. (2013). Quantitative Communication Research: Review, Trends, and Critique. *Review of Communication Research*, 1(1), 69-84. doi: 10.12840/issn.2255-4165_2013_01.01_003

Key words: Quantitative Research; Statistics; Research Design; Measurement

Editor: Giorgio P. De Marchis, Universidad Complutense de Madrid, Madrid, Spain

Received: Jul.17, 2012 **Accepted:** Oct. 3, 2012 **Published:** Jan. 2013

Content

Research is about learning new things. It is about advancing knowledge. Good research satisfies curiosity, but it also leads to new and intriguing questions begging to be answered (Firestein, 2012). Quantitative, social scientific research provides one avenue for furthering our understanding about communication. This essay summarizes the nature of the quantitative approach to communication research, discusses trends in quantitative communication over the past 20 years, and expresses concerns over research practices with the goal of providing constructive suggestions for improving the state of the art.

This review focuses on empirical, data-based, and quantitative communication research. Just as the label implies, quantitative research involves numbers and statistical analyses. At minimum, it involves counting instances, and it may involve quantified behavioral observation, scaling of psychological concepts, or physiological measures to name just a few types of data. Statistics are typically used to summarize data, test statistical hypotheses, and/or to claim that a given finding is not attributable to mere chance or randomness.

Quantitative communication research is typically based on a highly conventionalized approach to social science. Science and research, like all human endeavors, are subject to social pressures and normative influence. Understanding quantitative research in communication, therefore, requires understanding that it is conventional. The conventions are socialized and reinforced by text book descriptions and explanations, the teaching of social science methods to new generations of researchers, the demands of tenure and promotion evaluations, and the peer review process. As with much of human behavior, quantitative communication researchers look to what others are doing and do much the same thing. The topic of research differs from study to study and researcher to researcher, but the choices of approach tend to converge within research traditions.

The philosophical core of the quantitative approach is sometimes labeled scientific realism (cf. Pavit, 2001). In my view, scientific realism presumes a physical reality that exists independently from individual human perception. The goal of research is to bring our understanding more closely in line with actual reality. This is obtained through research methods that strive for objectivity, and the quality of the research is, among other things, evaluated by objectivity-closeness and the safeguards against subjectivity. Most would agree that absolute objectivity and a perfect correspondence between understanding and reality is not feasible. Instead, these are viewed as ideals to strive for. This philosophical core is shared across disciplines in the quantitative social sciences, but different disciplines and research traditions differ in how reality-closeness and objectivity are sought.

In terms of data gathering, the most common approaches to quantitative communication involve survey research, lab behavioral experiments, and content analysis of various media. Much of the research involves college students as research subjects and most data are collected in the United States. The individual study published in a peer-refereed academic journal is the typical unit of research accomplishment.

The statistical approach adopted by most quantitative communication research typically rests on a conventional (and many believe logically problematic; see Levine, Weber, Hullett, Park, & Lindsey, 2008) approach to data analysis that Gigerenzer et al. (1990) call modern hybrid statistics. This approach involves testing substantive hypotheses against nil-null statistical distributions. In this view, science is about testing and confirming statistical hypotheses based on probabilistically discrediting a lack of findings using the $p < .05$ standard. Although this view has been the subject of intense criticism across the social sciences, it nevertheless is nearly universally practiced in published quantitative communication research as well as most other quantitative social sciences. It is expected in the peer review process and it is ubiquitous in the training of new quantitative communication researchers. All commonly used statistical software packages contribute to the dominance of the hybrid approach.

The hybrid approach is a merging of two distinct statistical frameworks, one by R. A. Fisher and the other by Neyman-Pearson (Gigerenzer & Murray, 1987). The approach specifies two statistical hypotheses, the null (H0) specifying the distribution of the sample statistic if there is no difference or association, and the alternative (H1) which is defined as not H0. If the probability of the data presuming H0 is less than or equal to the conventional .05 level, the null is rejected, and evidence for H1 is inferred (Levine et al., 2008).

Quantitative research resides in an interesting place in the field of human communication. Numerically, quantitative communications researchers represent a relatively small but highly visible minority of researchers in the field. Despite being a minority, quantitative research and researchers are disproportionally highly represented in virtually all analyses of scholarly output such as publication rates (e.g., Bolkan, Griffin, Holmgren, & Hickson, 2012), research funding (Levine, 2012), and citations (Levine, 2010).

Statistics and methodology aside for a moment, when I contemplate the current state of knowledge about human communication, I am struck by how little I think we know. I acknowledge, however, the opposite conclusion is also defensible. In many ways, communication research has evolved at a rapid pace.

In my opinion, there are three especially fundamental questions that are at the core of our discipline. These are: (1) how is that people communicate with each other, and (2) what constitutes communication competence and effectiveness, and (3) how can communication competence and effectiveness be enhanced.

After more than 20 years as a communication teacher and researcher, I find it troubling that I do not have deeper answers to these questions, especially the first. Having just sampled four dozen published communication articles for this essay, as well as in my more general in-field reading, I have the impression that communication research is highly fragmented and fails to yield much insight into core communication processes. In fact, I find it scandalous that no one is actually communicating in the data that is used in much communication research and that most data is static rather than process oriented. I worry that our theory and methods are ill suited to the task of achieving a deep understanding of communication.

To explain my concern, I provide an example that I believe is representative. I very recently reviewed a paper for publication in a communication journal looking at athletic coach verbal aggression and student athlete motivation. I found this an interesting question. It might be interesting to code coach half time speeches and look at subsequent performance. If one wanted to be experimental, one could get a coach to enact different types of communication and then measure motivation. But, what the study I reviewed did was have a sample of student athletes read either a brief aggressive or a non-aggressive "message" from a hypothetical coach about their performance in a hypothetical athletic event, and then fill out some scales about their motivation and their opinions of the hypothetical coach. I wonder things like: How well can people project how motivated they might have been if the situation had been real? Research such as this may have its place, but wouldn't it be nice if more communication research studied, as data, actual communication rather than imagined communication?

Methodologically, I can name two culprits that I think are especially responsible for slowing intellectual progress. My current targets are what I perceive to be an over-reliance on self-report survey methods and modern hybrid null hypothesis significance testing (NHST for short) as the preferred statistical tool. Surely, these two practices are not without merit and surely also many other factors could be legitimately blamed for slow progress. Certainly, too, much of my own research involves self-report methods, NHST, or both. Nevertheless, I hope my colleagues around the world will reflect on the implications of the field's reliance on self-report measurement and NHST as well as the advantages of studying communication as a behavioral process. This means observing communication over time.

Before I address these and other issues in quantitate communication research, however, the results of a small-scale informal content analysis of communication research is offered. Impressions can offer insight, but it seems appropriate to collect and use some quantitative data. The general question guiding the data collection involves the trends and practices in quantitate communication research with attention to what has changed and what has not changed over the past 20 years. More specifically, it would be useful to offer data relevant to my claims regarding the ubiquity of self-report research and NHST. Therefore, it seemed reasonable to sample some recent communication research. It also seems reasonable to have a comparison or control group of older research. Finally, it was reasoned that the results would provide a useful way to frame and organize the current discussion.

Method

Sample and Sampling

The sample consisted of N = 48 randomly selected published articles reporting original quantitative research in leading communication journals. Four journals were selected: Human Communication Research (HCR), Communication Monographs (CM), Communication Research (CR), and Journal of Communication (JOC). In the author's opinion, these are the top 4 journals in the field. This opinion is based on current and historical citation patterns and centrality in network analysis (Feeley, 2008; Feeley & Moon, 2010; Levine, 2010; cf. Bolkan et al., 2012). Volumes corresponding to eight years for each of these journals were sampled. The years included 1988 to 1991 and 2008 and 2011. Twenty-four articles published between 2008 and 2011 were initially selected using random numbers generated on Random.org. Articles were randomly selected sequentially without replacement. To obtain each article, first one of the four journals was randomly selected. Then, one of the four volumes was randomly selected from within that journal. Then an issue was randomly selected from within the randomly selected volume. Finally, an article was randomly selected from within each issue. If the selected article did not report original quantitative research, the next article in the same issue that met the criteria was selected. This procedure was repeated until 24 articles were sampled. For each of the 24 articles, a matching article in the same journal from 20 years earlier was selected. So, for example, one article selected was in CM, 2008, the second article in issue four. The matching article was in CM, 1988, the second article in issue 4. All 48 articles that were selected were downloaded in PDF format for subsequent coding.

Coding

After reading all the selected articles, the author devised a coding system that captured the apparent trends, hunches about trends, insights gained from the reading, and the directions that the author hoped this essay would go. The articles were then re-read. The general topic of the research was recorded, along with the general method employed, the approach to measurement, and the types of statistical analyses. It was also recorded if the research was funded, if multiple studies were reported, if the research was limited to a college student sample, and if the data were entirely collected in the USA. All coding was done by the author. The coding was straight forward, and multiple coders were not deemed necessary to make the desired points.

Results

Table 1 provides a summary of the findings. Raw counts, percentages, and chi square tests are reported.

Table 1. Trends in Quantitative Communication Research over 20 years in Premiere Communication Journals

Trend	Time Period	
	1988-91	2008-11
	n = 24	n = 24
Topic [χ^2 (df = 5, n = 48) = 14.74, p = .01, φ = .55]		
Interpersonal – Group	11 (46%)	7 (29%)
Media	9 (38%)	6 (25%)
Organizational	3 (13%)	0
Health Communication	1 (4%)	5 (21%)
Com Tech / New Media	0	6 (25%)
Misc Article / Research Features		
Funded [χ^2 (df = 1, n = 48) = 0.76, p = .38, φ = .13]	4 (17%)	2 (8%)
Multiple Studies [χ^2 (df = 1, n = 48) = 0.12, p = .73, φ = .05]	5 (21%)	6 (25%)
Exclusively Student Sample [χ^2 (df = 1, n = 48) = 0]	12 (50%)	12 (50%)
Only USA data [χ^2 (df = 1, n = 48) = 5.7 8, p = .02, φ = .35]	22 (92%)	15 (63%)
Method [χ^2 (df = 7, n = 48) = 6.37, p = .50, φ = .36]		
Survey	15 (63%)	14 (58%)
Lab Experiment	6 (25%)	7 (29%)
Content Analysis	3 (13%)	1 (4%)
Other	0	3 (13%)
Measurement [χ^2 (df = 4, n = 48) = 8.21, p = .08, φ = .41]		
Self-Report Scales	17 (71%)	21 (88%)
Coding	13 (54%)	7 (29%)
Other (Physiological, reaction time)	1 (4%)	2 (8%)
Various Statistical Practices and Analyses		
Report Means [χ^2 (df = 1, n = 48) = 4.18, p = .04, φ = .30]	18 (75%)	23 (96%)
Standard Deviations [χ^2 (df = 1, n = 48) = 6.86, p = .01, φ = .38]	9 (38%)	18 (75%)
Distribution shape [χ^2 (df = 1, n = 48) = 3.20, p = .07, φ = .26]	0	3 (13%)
NHST [χ^2 (df = 1, n = 48) = 0]	23 (96%)	23 (96%)
Effect Sizes [χ^2 (df = 1, n = 48) = 3.60, p = .06, φ = .28]	14 (58%)	20 (83%)
Statistical Power	0	2 (8%)
Confidence Intervals [χ^2 (df = 1, n = 48) = 3.20, p = .07, φ = .26]	0	3 (13%)
EFA / PCA [χ^2 (df = 1, n = 48) = 0]	3 (13%)	3 (13%)
CFA [χ^2 (df = 1, n = 48) = 0]	3 (13%)	3 (13%)
Correlation [χ^2 (df = 1, n = 48) = 1.34, p = .24, φ = .17]	9 (38%)	13 (54%)
Multiple Regression [χ^2 (df = 1, n = 48) = 0.09, p = .76, φ = .05]	7 (29%)	8 (33%)
Chi Square [χ^2 (df = 1, n = 48) = 4.76, p = .03, φ = .38]	6 (25%)	0

(continued)

Table 1. Trends in Quantitative Communication Research over 20 years in Premiere Communication Journals (continued)

Trend	Time Period	
	1988-91	2008-11
	n = 24	n = 24
Raw Percentages [χ^2 (df = 1, n = 48) = 0.60, p = .44, φ = .11]	5 (21%)	3 (13%)
ANOVA [χ^2 (df = 1, n = 48) = 1.61, p = .20, φ = .18]	5 (21%)	9 (38%)
t-tests [χ^2 (df = 1, n = 48) = 1.50, p = .41, φ = .18]	5 (21%)	2 (8%)
MANOVA [χ^2 (df = 1, n = 48) = 0.0]	3 (13%)	3 (13%)
SEM [χ^2 (df = 1, n = 48) = 2.40, p = .12, φ = .11]	2 (8%)	6 (25%)
Log Linear Analysis [χ^2 (df = 1, n = 48) = 3.20, p = .07, φ = .26]	3 (13%)	0
Multiple Discriminant Analysis	1 (4%)	0
Network Analysis	1 (4%)	0
Multi-Level Modeling	0	2 (8%)
Mediation and/or Moderation	5 (21%)	18 (75%)
[χ^2 (df = 3, n = 48) = 14.56, p = .002, φ = .55]		
Tested Moderators	4 (17%)	14 (58%)
Tested Mediators	1 (4%)	6 (25%)

Chi square tests were not reported in situations with very small counts. Both percentages and chi squares should be interpreted with caution given the small sample size and uneven distributions. The reader should note that not all percentages sum to 100% because not all categories were mutually exclusive. For example, one article involved research using self-report data, behavioral observation, and a physiological measure and was coded as involving all three. Most articles reported more than one statistic.

Twenty years ago, two broad topics of research (interpersonal and media) accounted for more than 80% of the articles sampled from the four leading communication journals. Twenty years later, that proportion dropped to just over 50%. Communication technology and health communication research are now prevalent combining to account for 46% of the 2008-2011 sample. The shift from old (e.g., newspapers, radio, and television) to old and new media is not surprising given the rise of e-mail, texting, social networks and the like. The increased trendiness of health communication research has been noted in other recent research. In a larger sample of published research, health communication emerged as the single most studied topic in communication journals (Levine, 2012). It is speculated that the increase in health as a topic stems from the increased pressure on faculty in the United States to seek external research funding (Levine, 2012).

In spite of increased pressure for funding, the proportion of sampled articles that were funded declined (although not significantly so) over time. Levine (2012) also reported a lack of statistically significant differences in the proportion of published research that was funded. Funded research remains atypical in published communication research, even in the best journals.

The proportion of articles that reported multiple studies was similar over time with between 20 and 25% of articles reporting multiple studies. Multiple studies are common in some top psychology journals such as Journal of Personality and Social Psychology. This trend has apparently not gained momentum in communication.

The frequency of use of college student data was identical in both samples with 50% of studies reporting only student data. Data collected only in the United States was more common than the use of student data. More than 90% of the research in the 1988-1991 sample involved exclusively USA data. There has been a statistically significant internationalization of communication research over the past 20 years. Still, however, nearly two-thirds of the 2008-2011 sample was limited to data collected within the United States. Communication remains a USA-centric academic discipline, but internationalization is likely to continue.

In terms of the method employed, the similarities over time were more striking than the differences. Survey methodology remained the most prevalent method. Lab experimentation was evident in 25% to 30% of the research. Content analysis was less common (8% overall; declining from 13% to 4% over time). In the 2008-2011 sample, one article each employed non-meta-analytic secondary data analysis, meta-analysis, and naturalist behavioral observation.

With regard to measurement, as expected, self-report scaling was ubiquitous (79% overall; increasing from 71% to 88% over time). Behavioral or media coding (e.g., coding actual communication whether mediated or interpersonal) was reported in just over 40% of the articles samples, declining from 54% to 29%. These differences were marginally significant.

There were several noteworthy observations regarding statistical analyses and statistical reporting. First, in terms of descriptive statistics, the reporting of central tendency in the form of the arithmetic mean was very common (85% over all; increasing from 75% to 96% over time). Reporting dispersion was also common, although less so. Standard deviations or variances were reported in 56% of sampled articles. It is encouraging to note that the rate increased from 38% to 75% over time. Unfortunately, reporting distribution shapes remains atypical (6% overall; increasing from 0% to 13%). Readers can only know if the mean is a meaningful description of central tendency when the distribution is reported. Many uses of NHST also make assumptions about distributions. Finally, sometimes the shapes of distributions are substantively informative. Thus, it is unfortunate that most articles fail to report how the data are distributed.

In terms of inferential statistics, as expected, NHST was nearly universal being reported 96% of the sampled articles. No other methodological or statistical practice was as ubiquitous as NHST. The good news is that NHST were accompanied by effect size estimates in nearly three-quarters (71%) of the selected articles and that the reporting of effect sizes increased from 58% to 83% over time. In the reporting of effect sizes, communication may be well ahead of many of the other social sciences. Less encouraging were the findings that confidence intervals remain unusual and occurred in less than 10% of the articles examined. Further, although scales are commonly used, factor analysis remains uncommon.

In terms of the types of NHST used, zero-order correlations (46%), multiple regression (31%) and ANOVA (29%) were common. The prevalence of chi squares, t-tests, and log linear analysis are apparently in decline while structural equation modeling has become common (25% in the more recent set of articles). The use of MANOVA, multiple discriminant analysis, network analysis, and multi-level modeling were also observed in the sampled articles.

Discussion

This review focuses on trends in quantitative communication research. Forty-eight published articles reporting original quantitative research were sampled and content analyzed in an effort to provide an empirical foundation for the current discussion. Half of the articles analyzed were recent while the other half was twenty to twenty-four years old. All articles were randomly sampled from leading communication journals.

Among the noteworthy findings was a shift in the topics of research. Not surprisingly, as the nature of communication media and technology has evolved over time, communication research has followed. Research on newspapers, radio, and television has not been abandoned, but there has been some shift in focus to newer media such as computer-mediated communication, video games, and cell phone communication. A major tension in this research is between communication research that involves new technology, and research on new technology that involves social considerations. The challenge for communication researchers interested in technology is to maintain the primacy of communication processes in theory and research. Nevertheless, the shift from old to new media seems a natural response to the changing ecology of human communication.

The second major topical shift is the increased research on health communication. Health, of course, is not a new concern and the interplay between health and communication is multifaceted and important. Still, the sheer amount of research on health communication seems disproportional to its centrality to the field. Increased pressure for external funding in universities in the United States seems to be behind this trend. Public funding of higher education in the United States has been constant or declining while administrative costs have skyrocked. University administrators have responded by pressuring faculty to seek external research funding so that the universities can reap the overhead on grants. Since health is presumably the most fundable topic of communication research in the United States, universities have increasingly created and expanded health communication research, and the results of this trend are reflected in the increased proportion of communication articles that address health issues. Interestingly, however, the increase in health related research has not produced a corresponding increase in published funded research. In the future, it will be interesting to see if the increase in health communication is associated with a

2013 , *1* (1), 69-84

healthier population.

Another noteworthy finding was that the frequency of college student data was identical in both samples and that only half of studies examined reported exclusively student data. The author found it surprising that student data were not used exclusively in a majority of studies, nor was the use of student data different over time. The use of expedient student data is somewhat controversial and is conventionally considered a limitation. The actual issue is much more nuanced. For many core communication processes, students might not be meaningfully different than non-students. In my area of research (deception detection), there is ample evidence that students do not produce different results than non-students in traditional research designs. But, age, education, socio-economic status, and living in a college student environment certainly affect some communication processes and outcomes. Further, the extent to which findings might be different if a different sample was used is not well understood in many areas of communication research. Collecting data with different types of populations requires conceptual and measurement equivalence to make meaningful comparisons. Absent such equivalence, simply using harder to collect samples is unlikely to provide added value (Levine, Park, & Kim, 2007). Theory provides a better path to generality than methodological strategies (Levine, 2011).

Reporting and interpreting descriptive statistics is essential (Levine, Weber, Park, & Hullett, 2008). In this regard, communication research has improved substantially over time, but further improvement is still needed. Substantial proportions of articles in leading communication journals report means, standard deviations, and estimates of effect size (typically in units of zero-order correlations, standardized regression coefficients, multiple correlations, and eta squared). Where improvement is most needed is in reporting the shape of distributions. This, I believe, is one area where improvement is both needed and easy to accomplish.

There are at least three reasons why reporting the shape of distributions is valuable. First, noting the shape of distribution can have important substantive implications. For example, Serota, Levine and Boster (2010) recently observed that lying rates are not normally distributed and that most lies are told by a few prolific liars. Second, when distributions are not normal, the median and mode may be more informative than the mean, and the mean can be misleading. Therefore, readers need information about distributions in order to understand central tendency. Third, many significance tests rest on assumptions about the nature of statistical distributions. Even though statistics may be robust to violated assumptions or corrections may be reported, reporting distributions is informative.

Regarding the actual reporting of shapes of distributions, there are many ways this can be done, and the options depend both on the nature of the data and the goals of the research. If the data are approximately normally distributed, then researchers should say this and report at least means and standard deviations. If the data are substantially skewed, this should be noted, and it may make sense to report the mean, median, and mode(s) as central tendency. If there is more than one mode, the multi-modal nature of the data should certainly be reported. Graphing distributions with histograms or stem-and-leaf plots can be very informative and is easy to do.

Another desirable but infrequent aspect of statistical reporting is providing confidence intervals around estimates of effect sizes. Confidence intervals were reported in only 3 of the 48 reports examined. Information and examples for calculating confidence intervals are provided in Levine, Weber, Park, and Hullett (2008). Reporting effect sizes and the confidence intervals around the effect sizes would vastly improve reporting practices and overcome many of the limitations stemming from hybrid significance testing.

A surprising result was how infrequently power analyses are reported. Less than 10% of the articles sampled mentioned statistical power. This low rate of reporting is surprising because the reporting of statistical power is often specified in journal's instructions to authors. Simple rules of thumb like "anyways report estimates of statistical power" however are problematic.

Statistical power is one of the more complex and confusing issues addressed in this review. For one thing, statistical power does not exist in modern hybrid NHST. Power is a Neyman-Pearson idea and requires the specification of a precise H1. In hybrid NHST, H1 is simply defined as not H0. To calculate power, the sample size, the alpha level, and the effect size must be known. The problem is that the effect size is usually not known, and, if it was known, then there might be no need to the research (because the effect was already well documented). As a consequence, power is most often calculated based on arbitrary effect size levels making the results of power analyses also arbitrary. This makes power a confusing issue. But, power is also a critical issue because the lower the power, the more likely statistical inference errors. Statistical power can be improved by using larger sample sizes and by increased reliance on meta-analysis.

From the author's point of view, one of the most unfortunate findings was the frequent use of survey methodology (in 60% of the articles sampled) and self-report measures (in 79% of articles, increasing from 71% to 88% over time). Survey methodology and self-report measurement are clearly useful approaches to research design and measurement, but given the subject of human communication, the prevalence of surveys and self-reports seem disproportionate to their utility. Communication has a prominent behavioral aspect, whereas surveys and self-reports tend to get at cognitions and affect.

Generally, surveys and self-reports are maximally useful under two jointly necessary conditions: People must be willing and able to answer the questions. That is, they must know the answers and they must be willing to accurately communicate responses to the researchers. When limitations of self-reports are discussed, it is usually in reference to the second of the criteria. Researchers worry about things like social desirability distorting answers. While I suspect that subjects are not always honest in response to self-reports, the ability issue is usually the greater concern for me. I often doubt that people have the self-awareness and meta-communicative wherewithal to accurately answer what is asked of them. A recent meta-analysis of verbal aggression and argumentativeness, for example, suggests that self-report and behavioral studies do not converge and that correlation between self-reports and behavioral observation is low (Levine, Kotowski, Beatty, & Van Kelegom, 2012). The lack of association may stem from unreliability in behavior. Unfortunately, it is also possible that people lack the objective self-awareness to accurately uncouple their desired traits, projections of their own behavioral predispositions, and what they actually do.

As I write this review, I am working my way through a book titled The Folly of Fools: Deceit and Self-deception in Human Life (Trivers, 2011). The focus of Trivers' book is on explaining self-deception from an evolutionary biology perspective and his main thesis is that the primary function of self-deception is in the service of other deception. By deceiving ourselves, Trivers argues, we more effectively deceive other people thereby gaining advantage for ourselves and our offspring. This is possible, according to Trivers, because much human brain functioning happens without conscious awareness.

Even if Trivers is wrong about self-deception functioning to advance other deception, he is almost certainly correct that much human functioning, including many communication processes and outcomes, are not subject to conscious awareness, and, therefore, ill-suited to study with self-report methods. Research shows that when asked, people typically will give answers even when they do not know the answer (Schwarz, 1999). The excessive reliance on self-report measures in quantitative communication research limits our knowledge to aspects of communication that can be accurately and consciously known by our research subjects.

A second concern I have with self-reports is that much self-report research uses scales that are of questionable validity. As I have an interest in both individual differences and measurement validation, I have from time to time conducted validation research on previously published and supposedly already validated measures. More often than not my own data (e.g., see Levine et al. 2003; Levine, Kotowski, Beatty, & Van Kelegom, 2012) suggest serious validity problems. Scores on the scales do not seem to measure the constructs they were designed to assess. Consequently, the conclusions drawn from research using these measures are dubious. It seems to me that valid measurement is a prerequisite for genuine knowledge advancement and that highly fallible measurement will like lead to empirical dead ends and confusion. Readers and reviewers should demand better and stronger evidence of validity including confirmatory factor analysis and evidence of behavioral prediction, if relevant.

To sum up my major concerns with self-report research in communication, I worry that much knowledge critical to understanding communication is of a sort that cannot be understood with self-reports methods. Communication researchers would be well served by devoting more efforts in observing actual communication as it happens and less time having people recall or imagine communication. Second, even for topics well suited to self-report measurement, I worry that the scales used to measure communication variables are not properly validated and yield scores that measure something other than intended. The net result, I believe, is a slowing of progress. Much published research does not tell us very much, or worse yet, some research actively provides misinformation about communication.

Besides self-reports, another major concern is with the dominance of modern hybrid NHST. As the current content documented, NHST is pervasive and was reported in all but a couple of the articles sampled.

My concerns have been expressed elsewhere in more technical detail (see Levine, Weber, Hullett et al., 2008), but I will raise a couple of basic issues here. These should be sufficient to explain why I think NHST retards scientific progress.

One concern is that what a small p-value for a NHST tells us is that the finding was unlikely given that the nil null statistical hypothesis was true. The nil null hypothesis specifies no differences between groups or no association among variables. In the social sciences, including communication, the nil null is almost never literally true, regardless of the plausibility or viability of the thinking that gives rise to the alternative hypothesis. Things are never exactly equal. The variables are almost never perfectly uncorrelated. So, a significant p-value lets us reject this implausible nil null. But, so what? Knowing that a findings is not zero provides little new knowledge or understanding. This is why I advocate for effect sizes with confidence intervals.

Further, when findings are not significant, not much is learned either. In NHST, the null is never accepted. Non-significant does not mean that there is no difference or effect. Instead, it means that the difference or effect was simply not large enough given the sample size. The net result is that the outcome of NHSTs is typically not substantively satisfying. Increased reliance on descriptive statistics and on effect sizes with confidence intervals are more desirable alternatives.

The other point I want to make is that NHST as a decision can have very high error rates. Type I errors (i.e., false positive results; giving $p < .05$ when the null is true) are improbable unless large numbers of tests are produced and culled, so only the significant outcomes are reported. This appears to be a growing scandal in Psychology were p-value farming (culling significant findings from larger collections of non-significant findings) and other questionable research practices are becoming recognized as threats to scientific inference.

Type II rates (i.e., false negatives; $p > .05$ when the null is false), however, can be common. Type II errors happen at a rate of 1.0 minus statistical power. In practice, I might offer a guess of a 30% type II error in quantitative communication research using NHST given typical sample and effect sizes. If statistical power is, on average, .7, and if the nil null is almost always literally false, than the 30% type error II rate is expected. Solutions to lower this error rate include increasing sample sizes, greater reliance on meta-analysis, and focusing more on effect sizes and confidence intervals.

An implication of substantial Type II error rates and Type I errors produced by p-value farming is that virtually all literatures in quantitative communication research can be summarized as providing a confusing set of "mixed" results. Valid hypothesis are sometimes supported and sometimes not, and the same is true of invalid theories and hypotheses. The use of NHST and laws of probability guarantee this outcome. Fortunately, meta-analysis can help sort things out, but absent that, it is often hard to make sense what some set of studies tell us.

With regard to reducing Type I errors, there is one approach that is common that I believe is actually counterproductive. MANOVA is often used as a "gatekeeper" test for the purpose of reducing the risk of Type I errors. Univariate effects are only reported if the multivariate effect is statistically significant. The problem is that in communication research hypotheses are usually univariate in nature leading researchers to report both the multi- and the univariate effects. By necessity, this practice produces more, not fewer, significance tests, and therefore seems to make the problem worse. Although I think it is usually unwise to use MANOVA, if MANOVA is used, it makes most sense to do so only when (a) the dependent variables are highly inter-correlated, (b) the hypothesis is genuinely multivariate, and (c) there is some reason not to just factor analyze the dependent variables first. If one's dependent variables are highly inter-correlated, it makes more sense to me to use either confirmatory factor analysis or path analysis to model how the variables are related.

My point about NHST is that such tests get it wrong much of the time (due to statistical power and p-value farming) and even when they get it right, the substantive yield is meager. Not-zero is not a very high bar to test a hypothesis or theory against and knowing that an effect is not zero tells us little about what the effect is. NHST is all about trying to negate nullities and at the end of the day knowing "not zero" means a paper might be publishable, but it does not make the findings particularly informative. Increasing the use of a full complement of descriptive statistics and reporting confidence intervals around effect sizes would go a long way toward minimizing these concerns and facilitating progress.

I want to close by saying that I do both self-report research and NHST. But, I do not just do self-reports and just do NHST. I try to use self-reports when I think they are the best method, and I try to take measurement validity and validation seriously. As for NHST, I typically report them in my research, but I have come to increasingly focus on descriptive statistics with special attention to looking at how my data are distributed.

Quantitative communication research is highly conventional. Understanding it requires knowing the conventions. Communication research can be improved by considering which conventions serve us well and which ones impede progress. It is hoped that this review constructively helps toward this desired end.

References

Bolkan, S., Griffin, D. J., Holmgren, J. L., & Hickson, M. (2012). Prolific scholarship in communication studies: Five years in review. Communication Education (on-line first version).

Feeley, T. H. (2008). A bibliometric analysis of communication journals from 2002_2005. Human Communication Research, 34, 505-520.

Feeley, T. H., & Moon, S. (2010). Update on journal impact ratings in communication: 2006_2008. Communication Research Reports, 27, 355-364.

Firestein, S. (2012). Ignorance: How it Drives Science. NY: Oxford University Press.

Gigerenzer, G., Swijink, Z., Porter, T., Daston, L., Beatty, J. & Kruger, L. (1990). The Empire of Chance: How Probability Changed Science and Everyday Life. NY: Cambridge University Press.

Gigerenzer, G. & Murray, D. J. (1987). Cognition as Intuitive Statistics. Psychology Press.

Levine, T. R. (2010). Ranking and trends in citation patterns of communication journals. Communication Education, 59, 41-51.

Levine, T. R. (2011). Quantitative social science methods of inquiry. In M. Knapp & J. Daley (Eds). Handbook of Interpersonal Communication. Sage.

Levine, T. R. (2012). Funding in published communication research. The Electronic Journal of Communication, 22.

Levine, T., R., Bresnahan, M., Park, H. S., Lapinski, M. K., Wittenbaum, G., Shearman, S., Lee, S. Y., Chung, D. H., & Ohashi, R. (2003). Self-construals scales lack validity. Human Communication Research, 29, 210-252.

Levine, T. R., Kotowski, M. R., Beatty, M. J., & Van Kelegom, M. J. (2012). Trait-behavior correlations in argumentativeness and verbal aggression: A meta-analysis in reply to Infante, Rancer and Wigley. Journal of Language and Social Psychology, 31, 95-111.

Levine, T. R., Park, H. S., & Kim, R. K. (2007). Some conceptual and theoretical challenges for cross-cultural communication research in the 21st century. Journal of Intercultural Communication Research, 36, 205-221.

Levine, T. R., Weber, R., Hullett, C. R., Park, H. S., & Lindsey, L. (2008). A Critical Assessment of Null Hypothesis Significance Testing in Quantitative Communication Research. Human Communication Research, 34, 171-187.

Levine, T. R., Weber, R., Park, H. S., & Hullett, C. R. (2008). A Communication Researchers Guide to Null Hypothesis Significance Testing and Alternatives. Human Communication Research, 34, 188-209.

Pavitt, C. (2001). Philosophy of Science and Communication Research. NY: Nova Science.

Serota, K. B., Levine, T. R., and Boster, F. J. (2010). The prevalence of lying in America: Three studies of reported deception. Human Communication Research, 36, 1-24.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. American Psychologist, 54, 93-105.

Trivers, R. (2011). The Folly of Fools: The Logic of Deceit and Self-Deception in Human Life. NY: Basic Books.