

Why the Resistance to Statistical Innovations? Bridging the Communication Gap

Donald Sharpe
University of Regina

While quantitative methodologists advance statistical theory and refine statistical methods, substantive researchers resist adopting many of these statistical innovations. Traditional explanations for this resistance are reviewed, specifically a lack of awareness of statistical developments, the failure of journal editors to mandate change, publish or perish pressures, the unavailability of user friendly software, inadequate education in statistics, and psychological factors. Resistance is reconsidered in light of the complexity of modern statistical methods and a communication gap between substantive researchers and quantitative methodologists. The concept of a Maven is introduced as a means to bridge the communication gap. On the basis of this review and reconsideration, recommendations are made to improve communication of statistical innovations.

Keywords: quantitative methods, resistance, knowledge transfer

David Salsburg (2001) began the conclusion of his book, *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*, by formally acknowledging the statistical revolution and the dominance of statistics: “As we enter the twenty-first century, the statistical revolution in science stands triumphant. . . . It has become so widely used that its underlying assumptions have become part of the unspoken popular culture of the Western world” (p. 309). Psychology has wholeheartedly embraced the statistical revolution. Statistics saturate psychology journals, classes, and textbooks. Statistical methods in psychology have become increasingly sophisticated. Meta-analysis, structural equation modeling and hierarchical linear modeling are advanced statistical methods unknown 40 years ago but employed widely today. Statistical articles have disproportionate impact on the psychology literature. Citation counts are a widely accepted measure of impact. Sternberg (1992) reported seven of the 10 most-cited articles in *Psychological Bulletin* addressed methodological and statistical issues. Replicating his analysis for 2013 finds eight of the 10 most cited articles again focus on method.

Discontent

Quantitative methodologists are those psychologists whose graduate training is primarily in quantitative methods, who write articles for quantitative journals, and who focus their teaching and research on statistics. Given statistics’ saturation, sophistication, and influence in psychology, this should be a golden age for

quantitative methodologists. Instead, I argue there is much discontent.

A persistent irritation for some quantitative methodologists is substantive researchers’ overreliance on null hypothesis significance testing (NHST). NHST is the confusing matrix of null and alternative hypotheses, reject and fail to reject decision making, and ubiquitous p values. NHST is widely promoted in our introductory statistics classes and textbooks, and it is the foundation for many of our theories and practices. The popularity of NHST defies critical books, journal articles and dissertations, and the reasoned arguments of some of our most respected and influential thinkers on statistical matters. Few quantitative methodologists have called for an outright ban on NHST (e.g., Schmidt, 1996); few regard the status quo to be satisfactory (e.g., Chow, 1996). Members of a self-described statistical reform movement have proposed alternatives to NHST, specifically effect sizes as a measure of the magnitude of a finding and confidence intervals as an expression of the precision of a finding. Whether these *alternatives* are better thought of as replacements for or as supplements to NHST, adoption of effect sizes and confidence intervals has been exasperatingly slow.

Discontent among quantitative methodologists by no means is limited to NHST. Quantitative methodologists have long documented substantive researchers’ failures to employ best statistical practices. One failure at best practice by substantive researchers is their ignorance of statistical power. Fifty years ago Cohen (1962) recognized that detecting small and medium sized effects requires sufficient statistical power (i.e., a large number of research participants). Cohen found the psychological literature to be riddled with underpowered studies, a finding replicated a quarter century later by Sedlmeier and Gigerenzer (1989). Paradoxically, those same underpowered studies sometimes produce statistically significant results, albeit spurious results that fail to replicate (Button et al., 2013).

Another failure at best practice is ignorance of statistical assumptions. All statistical tests have assumptions (e.g., the data

This article was published Online First September 30, 2013.

I thank Robert Cribbie, Cathy Faye, Gitte Jensen, Richard Kleer, Daniela Rambaldini, and William Whelton for their helpful comments and suggestions.

Correspondence concerning this article should be addressed to Donald Sharpe, Department of Psychology, University of Regina, Regina, SK, Canada, S4S 0A2. E-mail: sharped@uregina.ca

should follow the normal or bell-shaped distribution) that when violated can lead substantive researchers to erroneous conclusions for statistical significance, effect size magnitude, and confidence intervals. Robust statistics unaffected by assumption violations have been promoted by some statisticians (e.g., Wilcox, 1998) as means to reach more accurate conclusions. However, robust statistics have been utilized by few substantive researchers.

There are other recurring examples of poor practices in data analysis by substantive researchers. A partial list of poor practices includes failing to address outliers (Osborne, 2010b), employing mean substitution to replace missing data (Schlomer, Bauman, & Card, 2010), conducting stepwise analysis in multiple regression (B. Thompson, 1995), splitting continuous variables at the median (MacCallum, Zhang, Preacher, & Rucker, 2002), failing to interpret correlation coefficients in relation to beta weights in multiple regression (Courville & Thompson, 2001), and following a multivariate analysis of variance (MANOVA) with univariate analyses of variance (ANOVAs; Huberty & Morris, 1989).

Resistance

When quantitative methodologists express their discontent with substantive researchers, the word they most often use is *resistance*. Schmidt and Hunter (1997) acknowledged “changing the beliefs and practices of a lifetime . . . naturally . . . provokes resistance” (p. 49). Cumming et al. (2007) surveyed journal editors, some of whom spoke of “authors’ resistance” (p. 231) to changing statistical practices. Fidler, Cumming, Burgman, and Thomason (2004) commented on the “remarkable resistance” (p. 624) by authors to journal editors’ requests to calculate confidence intervals. B. Thompson (1999) titled an article “Why ‘Encouraging’ Effect Size Reporting Is Not Working: The Etiology of Researcher Resistance to Changing Practices.”

Resistance by substantive researchers to statistical innovations is all the more puzzling because it is not universal. Some statistical innovations (e.g., meta-analysis, structural equation modeling) are adopted rapidly even over strong initial objections. Other statistical innovations (e.g., power analysis) are resisted for a long period of time but adopted eventually. And some statistical innovations (e.g., robust statistics) encounter resistance that appears intractable.

Why Resistance?

I will not rehash the NHST controversy nor catalog what quantitative methodologists regard as faulty statistical practices by substantive researchers. What I will do is examine proposed sources of resistance to statistical innovations and offer alternative explanations for that resistance. My examination follows calls from writers such as Finch, Cumming, and Thomason (2001), who stated “We leave to historians and sociologists of science the fascinating and important question of why psychology has persisted for so long with poor statistical practice” (p. 206).

Lack of Awareness

Resistance could stem from a lack of awareness of developments in statistical theory and methods. Understandably substantive researchers focus their attention on applied work and may fail

to stay informed of statistical advances (Mills, Abdulla, & Cribbie, 2010; Wilcox, 2002).

Lack of awareness of robust statistics has been offered as an explanation for their limited adoption (Wilcox, 1998). In his controversial article claiming evidence for ψ , Bem (2011) disavowed use of robust statistics because he concluded such statistics are unfamiliar to psychologists. Indeed, Erceg-Hurn and Mirosevic (2008) authored a recent publication in *American Psychologist* with the explicit goal of raising awareness of robust statistics. However, Wilcox (1998) wrote a similar article published in *American Psychologist* 10 years earlier with the same intent but with no resulting increase in their use.

Journal Editors

There is a widespread belief journal editors could serve as catalysts for change in statistical practices but have failed to do so (e.g., Finch et al., 2001; B. Thompson, 1999). Kirk (1996) suggested that if editors mandated change, their involvement “would cause a chain reaction: Statistics teachers would change their courses, textbook authors would revise their statistics books, and journal authors would modify their inference strategies” (p. 757).

One editor who attempted reform of statistical practices was Geoffrey Loftus. During his term as editor of *Memory & Cognition*, Loftus required article authors to report confidence intervals. The number of *Memory & Cognition* articles reporting confidence intervals rose during Loftus’s editorship but fell soon after (Finch et al., 2004). A more recent editorial effort to change statistical practices was made by Annette La Greca, the editor of the *Journal of Consulting and Clinical Psychology (JCCP)* who mandated effect sizes and confidence intervals for those effect sizes. Odgaard and Fowler (2010) reviewed *JCCP* articles from 2003 to 2008 for compliance. Effect size reporting rose from 69% to 94% of articles but confidence intervals for effect sizes were less frequently reported (0% in 2003 to 38% in 2008). A third effort at editorial reform was by editors of *Psychological Science*, encouraging researchers to use p_{rep} in place of p values. Killen (2005) derived p_{rep} as an estimate of the probability an effect would replicate. While many authors submitting articles to *Psychological Science* adopted p_{rep} , the journal editors subsequently dropped their recommendation. In the case of p_{rep} , editorial reform failed not because subsequent editors lacked commitment to the reform or because substantive researchers’ resisted the reform but rather because quantitative methodologists found p_{rep} failed as a measure of effect replication (e.g., Maraun & Gabriel, 2010).

Journal editors can facilitate change but editors also can function to maintain the status quo and impede change (Sedlmeier & Gigerenzer, 1989). Hyde (2001), herself a past journal editor, acknowledges that many editors were trained in statistics in a different era and may not be familiar with statistical advances. Furthermore, even when journal editors are familiar with statistical advances and dictate their use, authors may pay lip service to those instructions. Cumming et al. (2007) found authors who reported confidence intervals in their results sections as required by editorial policies rarely interpreted those confidence intervals in their discussion sections. McMillan and Foley (2011) reached the same conclusion for effect sizes.

In their defense, journal editors depend on the availability of reviewers with statistical expertise. Statistical reviewers demon-

strably improve the quality of published articles (see Cobo et al., 2007). However, there are a limited number of qualified and willing quantitative reviewers. Ozonoff (2006) pointed out that “obtaining two reviewers with appropriate specialist expertise is difficult enough without requiring yet another reviewer to evaluate the use of statistics in a paper” (para. 5). And the pool of reviewers with quantitative skills is shrinking. According to the publisher of the American Psychological Association, Gary R. VandenBos, “APA has over 45 editors, and they all have in their Rolodexes the name and address of the same 24 people—almost all of whom are over the age of 60” (Clay, 2005, p. 27).

Publish or Perish

Publish or perish is a phrase attributed to Caplow and McGee (1958), who investigated the lives of academics in the mid-20th century. According to Caplow and McGee, tenure and promotion decisions were made solely on the basis of research activity; academics with insufficient research activity were fired. In the second decade of the 21st century, academics must be more than merely research active. Competition for academic jobs and research funding demands frequent publication in the top ranked journals, those journals with the highest rejection rates, the most challenging reviewers, the most selective editors, and the most compelling, hypothesis supporting findings (Fanelli, 2012).

Publish or perish pressures have implications for statistical practices. First, publish or perish thinking contributes to the misuse of statistics (Gardenier & Resnik, 2002). Top journals and competitive granting agencies demand complex statistical methods to answer complex research questions (Aiken, West, & Millsap, 2008). These complex statistical methods may exceed some substantive researchers’ abilities to properly conduct their analyses (Wasserman, 2013). Second, publish or perish demands sway some researchers to engage in statistical practices that border on (or cross the line into) the unethical. Researchers can *cook* results to make them statistically significant, *mine* data looking for statistically significant results, and *selectively publish* to support preexisting hypotheses (Fanelli, 2009). Third, publish and perish forces may contribute to outright data fraud. According to disgraced social psychologist Diederik Stapel profiled in the *New York Times Magazine*, “There are scarce resources, you need grants, you need money, there is competition” (Bhattacharjee, 2013, para. 25). Stapel spoke of sitting at his kitchen table for many hours over a number of days, generating statistically significant differences of believable effect size magnitudes.

Software

Software can play a positive or a negative role in statistical innovation. According to Aiken et al. (2008), accessible and available software facilitates adoption of statistical innovations. Conversely, Keselman et al. (1998) attribute the failure to adopt best statistical practices to software’s “inaccessibility and/or complexity” (p. 379).

A good news software story is LISREL’s role in the growth of structural equation modeling. Karl Jöreskog, the father of modern structural equation modeling, codeveloped LISREL. Without LISREL, structural equation modeling might not have become so popular so quickly (Bollen, 1989). The growth in popularity of

structural equation modeling begat other structural equation software such as EQS, Amos, and Mplus (see Byrne, 2012). Similarly, power analysis was facilitated by the presence of readily accessible software such as G*Power (Faul, Erdfelder, Lang, & Buchner, 2007).

A bad news software story is robust statistics. The absence of robust statistics in commercial software packages has been offered as an explanation by proponents of robust statistics (e.g., Erceg-Hurn & Mirosevich, 2008; Wilcox, 2002) for why robust statistics are not used more. Robust statistics are accessed primarily through R and, to a lesser extent, SAS, STATA, and SPSS. R is free, easily acquired, and extremely popular with statisticians. Recently Field wrote a version of his bestselling SPSS statistics textbook for R (Field, Miles, & Field, 2012). In a review of Field’s R textbook and two others, Shuker (2012) exclaimed “R is taking over the statistical world” (p. 1597). However, Shuker then cautioned “R has a very steep and slippery learning curve” and acknowledged R is “not a stats package. Rather, it is a programming language in which statistical programmes have been developed and collected together to form a de facto statistical computing environment” (p. 1597).

How much uptake of R is there by substantive researchers? In a survey of the popularity of data analysis software, Muenchen (2013) found R is in high demand with bloggers and data miners, but SPSS and SAS are dominant among scholars. I surveyed all issues of the *Journal of Consulting and Clinical Psychology* for 2012. Of 99 empirical articles, authors of 63 articles identified the statistical package(s) used to analyze their data. Twenty-one authors reported using SPSS, 18 authors SAS, 12 authors HLM, 12 authors MPlus, and nine authors some other statistical package. Only seven authors reported using R, and six of those seven authors used R in combination with another statistical package. Similar results were found for 2012 issues of the *Journal of Personality and Social Psychology* and *Developmental Psychology*. While undoubtedly R will grow in popularity among substantive researchers, the game changer for robust statistics may be a bootstrapping module in recent versions of SPSS. Bootstrapping is already familiar to substantive researchers as a means to test indirect effects and the SPSS bootstrapping module allows users to readily compare the results from traditional statistics to robust bootstrapped estimates.

An ugly new software story is that software contributes to the persistence of faulty statistical practices. Fabrigar, Wegener, MacCallum, and Strahan (1999) suggested commonly employed but inferior choices in factor analysis strategies (e.g., varimax rotation) can be attributed to those choices being the defaults in SPSS. More generally, Osborne (2010a) bemoans the *point and click mentality* encouraged by modern statistical software. An honor psychology student wrote on her course evaluation she did not want to learn statistical theory but merely to be told what buttons to click in SPSS.

Inadequate Education

Failure to adopt better approaches to data analysis has long been blamed on inadequate statistical education (Henson, Hull, & Williams, 2010). Muthén (1989) concluded faulty application of basic statistical techniques was an educational issue and recommended better training of students. Graduate student training is limited in

advanced statistical methods (Aiken et al., 2008; Keselman et al., 1998). Gorman and Primavera (2010) tattled on graduate students who do well in statistics classes but who cannot analyze their own thesis data.

Curriculum in statistics classes is one culprit in this purported inadequate education. K. Thompson and Edelstein (2004) suggest these classes emphasize the theoretical and the abstract over applied training. To the contrary, Shaver (1993) regards the statistical curriculum to be too applied, too focused on selecting the correct statistical test. The majority of faculty teaching statistics in psychology are not trained primarily in statistics (Rossen & Oakland, 2008). In a survey of 18 Canadian psychology departments, Golinski and Cribbie (2009) found most departments had none or only one quantitative faculty member. In education, the faculty who teach statistics tend to be outside of the core curriculum. Yet in education, like psychology, the statistical training targets traditional, basic methods rather than modern, advanced techniques (Henson et al., 2010). Even traditional methods such as regression are sometimes neglected in statistics classes. MacCallum et al. (2002) suggested that the practice of dichotomization of continuous variables reflects researchers' and graduate students' greater familiarity with ANOVA over regression.

Textbooks supplement the curriculum in many statistics classes and accessible textbooks can play a role in advancing statistical practice. Rucci and Tweney (1980) attributed the growth in the use of ANOVA to then popular textbooks. However, textbook reviews from a decade ago (e.g., Pituch, 2004) showed coverage of recent statistical advances to be hit or miss. These reviews are dated and perhaps the situation has improved. The latest edition of the popular introductory statistics textbook by Gravetter and Wallnau (2013), for example, provides an 11-page discussion of effect sizes and statistical power in their hypothesis testing chapter and presents instructions for calculating effect sizes and confidence intervals for a number of basic statistics.

Mindset

B. Thompson (1999) attributed resistance to changing statistical practices to psychological factors in the minds of substantive researchers. Thompson labeled one psychological factor as *confusion or desperation* akin to what Schmidt (1996) called *false beliefs* (i.e., the level of statistical significance [$p < .05$ vs. $p < .00001$] speaks to the size of the relationship). Another psychological factor identified by Thompson is *atavism* or a fear of deviating from normative practices. Substantive researchers fear their work will not be published if they fail to follow established methods.

Resistance Reconsidered

Let us reconsider what is meant by *resistance*. While philosophers of social science have written on resistance to paradigm change (e.g., Kuhn, 1962), organizational management theorists have examined resistance to change *within* established paradigms. Three lessons can be drawn from the organizational management literature. First, apply the resistance label cautiously. Labeling behavior or attitudes as resistance ignores legitimate concerns (Piderit, 2000). Better to regard resistance as "a useful red flag—a signal that something is going wrong" (Lawrence, 1969, p. 56).

Second, resistance to change is often rational. Resistance is tolerated and even encouraged in some circumstances (Ford, Ford, & D'Amerlio, 2008). Third, it is not just *resisters* who need to change their behavior.

What principles can we take from these three lessons to better understand resistance to statistical innovations? First, resistance does not constitute irrational behavior to be remedied, for example by editorial policies. Second, resistance says less about substantive researchers, their awareness, education, pressures or mindset, and more about the properties of a statistical innovation such as its *complexity*. Third, resistance arises out of failed relationships, specifically the *communication* between substantive researchers and quantitative methodologists.

Complexity

In an e-mail, I chided Erceg-Hurn and Mirosevich (2008) for their choice of the word "easy" in their description of robust methods as "An Easy Way to Maximize the Accuracy and Power of Your Research." Traditional statistics, advanced statistics, robust statistics—few statistics are easy. Tukey (1986) wrote "statistics is intrinsically complex, at least so far as any of us can see today" (p. 591).

Statistical complexity is seen in the controversy around NHST. Members of the statistical reform movement regard the alternatives to NHST to be *new statistics* (Cumming, 2012) and the discipline to be moving *beyond significance testing* (Kline, 2013). However, both reformers and moderates acknowledge a continuing role for NHST. Harris (1991) recognized that "[NHST] provide[s] a useful check that the results of our contrasts are not due merely to random sampling fluctuation. However, that is all they do" (p. 378); Kline (2013) conceded "statistical significance provides even in the best case nothing more than low-level support for the existence of an effect, relation, or difference" (p. 114). But there is a need for this support if substantive researchers seek to determine the presence or absence of a phenomenon. Abelson (1997) offered the examples of whether infants understand addition and subtraction and whether extrasensory perception exists: "The moral of the story is that if we give up significance tests, we give up categorical claims asserting that something surprising and important has occurred" (p. 14).

One widely promoted alternative to NHST is estimating an effect size. What is one to make of the size of an effect? Cohen's (1988) scheme of .2 as a small effect, .5 as a medium effect and .8 as a large effect strikes even statistical reformists (e.g., B. Thompson, 2001) as no less arbitrary than the .05 cutoff for a statistically significant result. Indeed, Cohen (1988) recognized that the breadth of the behavioral sciences defies easy classification of effect sizes. Furthermore, size is not everything. A small effect can be important when the independent variable is manipulated minimally or the dependent variable is difficult to influence (Prentice & Miller, 1992). Conversely, not all large effect sizes are meaningful. Conducting a meta-analysis in education, occasionally I would calculate double-digit effect sizes when children with and without intellectual disabilities were compared on measures of academic achievement. Effect sizes need to be interpreted in the context of the research domain (Durlak, 2009), and effect size magnitude is impacted by the research design, the reliability of measures, the distributional assumptions of the data, and the effect size statistic

selected (Grissom & Kim, 2012). Indeed, there is not *one* effect size statistic, but more than 40 different *types* of effect size statistics (Kirk, 1996). Confidence intervals are another widely promoted alternative to NHST. Like effect sizes, confidence intervals have their own complexities, such as misinterpretations of their definition and meaning (Fritz, Scherndl, & Kuhberger, 2013).

Few quantitative methodologists endorse *naked p* values, the practice of reporting *p* values without any supporting information (Anderson, Link, Johnson, & Burnham, 2001). Indeed, there is no contradiction in reporting an NHST based *p* value together with an effect size and a confidence interval for that effect size. Authors of a well-publicized recent study claimed hand clenching led to improved memory. In a blog posting, the lead author offered large effect sizes (e.g., $d = 0.85$ for one important comparison, statistically significant at $p < .05$) as support for robust findings in spite of small sample sizes. However, a critic calculated the 95% confidence interval for the $d = 0.85$ effect size to be $-.06$ to $+1.78$, suggesting this *large* effect was far from impressive.

Statistical complexity is a broader concern than NHST and its alternatives. Altman, Goodman, and Schroter (2002) quoted Luykx: "It is now almost inconceivable that a study of any dimensions, in medical science, can be planned without the advice of a statistician" (p. 2817). Luykx made that statement in 1949. While the distance was not great in 1949 between statistical methods taught in introductory statistics classes and statistical methods used in research, the distance is now substantial (Wilcox, 2002). A 2004 PhD psychology graduate wrote "It feels like every time you open any journal in quantitative psychology, there's an article about a new breakthrough in the field. . . . What's scary is that the number of people who can really understand and use those developments is dwindling" (Novotney, 2008, para. 4).

Even frequently used statistical methods such as exploratory factor analysis and meta-analysis have high *researcher degrees of freedom* (Simmons, Nelson, & Simonsohn, 2011), the numerous, often unstated decisions made in the process of performing statistical analyses. It is when making these decisions that substantive researchers may seek the advice of quantitative methodologists. Lance and Vandenberg (2009) collected statistical myths, long-standing beliefs about how a statistical test functions or how the results of a statistical test should be interpreted. For many of these statistical myths, there was a *kernel of truth* that could be traced to misinterpretation of statements from the statistical literature. These misinterpretations arise because data analysis is rarely simple and issues that appear to be settled frequently are not. DeCoster, Iselin, and Gallucci (2009) tested justifications offered by substantive researchers who had committed the statistical sin of dividing continuous variables at some arbitrary value (e.g., the median). Alas, DeCoster et al. reported "the picture is a bit more complicated than that . . . there are some circumstances in which the dichotomized indicator performed just as well or even slightly better than the continuous indicator" (p. 363).

Complex statistical techniques can suggest new ways to conduct research (Aiken et al., 2008). However, many of the most influential psychologists of the last half century (e.g., Milgram, Skinner) employed basic statistics or no statistics at all (Cohen, 1990). MacCallum (1998) warned "We sometimes forget that important research is not necessarily characterized by the application of sophisticated or complex statistical techniques" (p. 3). Furthermore, Zimiles (2009) observed complex statistical techniques

serve to divide researchers into those who can and cannot use such techniques (or have access to consultants who can). This division is exemplified when journal reviewers ask article authors to replace a traditional, *minimally sufficient* analysis (Wilkinson & the Task Force on Statistical Inference, 1999), with a modern, more complex statistical approach.

When complexity abounds, simple, compelling ideas can be underappreciated. Kenny (2008), reflecting on Baron and Kenny (1986), told the story of when they first submitted their article to *American Psychologist*, the article was rejected "because reviewers felt what we were saying was known to everyone" (p. 354). The article was then submitted to the *Journal of Personality and Social Psychology*, whose editor rejected it for the same reason. Baron and Kenny resubmitted the article to a new editor who overruled some of the reviewers in publishing the article. Baron and Kenny (1986) became the most cited article published in the *Journal of Personality and Social Psychology* and introduced meditational analysis to a generation of researchers. In the same vein, Cohen (1992) said that he was initially rebuked by a number of methodologists he sent the Cohen (1968) regression article. Cohen (1992) wrote "Given my poor background in math and math statistics, I didn't understand much of their criticism, and in any case, my own intuitive nonmathematical approach was the only way I could write. I had no choice but to write as I thought and as I thought most psychologists thought" (p. 409). The revised article was accepted by *Psychological Bulletin* and became one of Sternberg's (1992) 10 most cited articles.

Communication

Borsboom (2006) speaks of disconnect between quantitative methodologists and substantive researchers attributable to their failure to communicate. Wilcox (2002) put it bluntly: "the lines of communication have virtually broken down between statistics and psychology, partly because each is absorbed in its own disciplinary enterprise, and partly because each views the other as isolationist" (para. 15).

Journals. Historically, journal articles were an important conduit for practical statistical advice. Crutchfield and Tolman's (1940) article coincided with a substantial increase in the use of ANOVA by psychologists (Lovie, 1979). Crutchfield and Tolman's presentation of ANOVA was not mathematical and used examples drawn from psychology. More recently, Busk (1993) surveyed statistical consultants and their clients and found the consultants relied upon and referred clients to journal articles.

Unfortunately, there is no single source for statistical journal articles accessible to substantive researchers. *Psychological Methods* was spun off from *Psychological Bulletin* to serve that function. Appelbaum and Sandler (1996) wrote in the first sentence of the opening editorial "We begin publication of the *Psychological Methods* with great hopes and expectations that this journal will provide a mechanism for effective communication between those individuals who concern themselves with methodological issues and those whose substantive research depends on excellence in methodology" (p. 3). Some feel *Psychological Methods* has failed this mandate. Roger Kirk, interviewed by Fidler (2005), stated "I was a little disappointed, because I wanted [*Psychological Methods*] to be more tutorial, and speak to the non-specialist as well as the specialist. That is not the direction it took" (p. 112).

Some quantitative journals have sections devoted to introducing statistical techniques but accessibility of the articles in those sections varies. *The Teacher's Corner* in a recent issue of *Structural Equation Modeling* offered for consideration an article titled "Advanced Nonlinear Latent Variable Modeling: Distribution Analytic LMS and QML Estimators of Interaction and Quadratic Effects." Some substantive journals do publish introductory articles on new and established statistical techniques. On one hand, the authors of these introductory articles have reached out to substantive researchers by publishing in the journals substantive researchers read. On the other hand, one must scour the literature for the appearance of these articles. A clear description of how to calculate contrasts following a factorial ANOVA is found in an appendix of an article in the *Journal of Clinical Child and Adolescent Psychology* (Jaccard & Guilamo-Ramos, 2002); Osborne (2010b) provided a state of the art tutorial on data cleaning in *Newborn and Infant Nursing Reviews*; Durlak (2009) explained effect sizes, and Finch and Cumming (2009) confidence intervals in their respective articles in the *Journal of Pediatric Psychology*. If one is not a developmental researcher, one might have missed these articles.

Teaching. Classroom teaching is the primary means by which quantitative methodologists communicate with the next generation of substantive researchers. In a provocative article in the *American Statistician*, Meng (2009) concluded there is a need for statisticians to develop what he colorfully described as "more appetizing happy courses," courses "that would truly inspire students to learn—and learn happily—statistics as a way of scientific thinking" (p. 205).

There is no denying statistics courses have a poor reputation among students (and some faculty). Student anxiety around statistics is widely acknowledged (see Onwuegbuzie & Wilson, 2003). Gorman and Primavera (2010) observed buttons at psychology conventions: "I survived statistics"; "Roses are Red; Violets are Blue; I Hate Stats"; and "The Surgeon General warns that Statistics may be harmful to your health" (p. 21). Popular statistical textbooks reflect and reinforce these negative perceptions with titles such as *Statistics for the Terrified* and *Statistics for People Who (Think They) Hate Statistics* or, as Gorman and Primavera offered as a generic title for all such books, "The Loathsome Study of Statistics for Those Who Are Utterly Confused and Incompetent" (p. 21).

Negative views of statistics have been described as a barrier to students expressing an interest in pursuing advanced quantitative training (Landes, 2009). Graduate students take the minimum number of statistics classes (Aiken et al., 2008) and the number of quantitative doctoral programs is in decline (*American Psychological Association*, nd). A good statistics teacher can reverse students' negative views of statistics. However, knowledge of statistics is not sufficient to be a good teacher. Gorman and Primavera (2010) warn "It's dangerous to assume that someone who knows statistics well can also be a good statistics teacher. We've probably all heard statements like 'He's brilliant but I don't understand anything he's saying'" (p. 22).

Consulting. Statistical consulting is another way for quantitative methodologists to reach substantive researchers. Boen and Zahn (1982) list a number of characteristics of the ideal statistical consultant such as knowing statistics and the statistical literature, being an effective problem solver, possessing good communication skills, and producing high quality work in a timely manner. But according to Boen and Zahn, the most important characteristic

is the consultant wants to *educate* their clients about statistical matters. Henson et al. (2010) wrote, "We hope to see the day when education researchers no longer rely on a *wizard* methodologist who retires to a back room with a computer, conjures the spirits of Spearman, Fisher, Cohen, and Cattell, and emerges with a Results section for the next publication" (p. 237).

Mavens

Seeking to bridge the communication gap and navigate the complexities of advanced statistics is a small number of individuals. These individuals have not been formally recognized nor their contributions widely appreciated. Yet many psychology departments have someone whom researchers, faculty, and students seek out when their data need analyzing. I refer to these individuals as Mavens. A Maven is a trusted expert in a field who passes on knowledge to others (Feick & Price, 1987). In commercials for the Vita seafood company, a voiceover spoke, "Get Vita at your favorite supermarket, grocery or delicatessen. Tell them the beloved Maven sent you. It won't save you any money, but you'll get the best herring" (Denker, 2007, p. 87). Mavens featured prominently in Malcolm Gladwell's (2002) bestseller, *The Tipping Point*. Gladwell described Mavens as functioning to gather knowledge about consumer products and services, but Gladwell recognized that Mavens' importance is not merely the knowledge they gather but more their desire, willingness and aptitude for communication. According to Gladwell, "The critical thing about Mavens, though, is that they aren't passive collectors of information. It isn't just that they are obsessed with how to get the best deal on a can of coffee. What sets them apart is that once they figure out how to get that deal, they want to tell you about it too" (p. 62).

Someone akin to a Maven has been mentioned previously in the quantitative psychology literature. Muthén (1989) talked of the pressure on a small number of *bridgers* between statisticians and the users of statistics. While not regarded by Muthén to be highly trained in statistics, these bridgers serve by teaching and consulting users of statistics. Similar to Muthén's bridgers, Fiske's (1981) *methodologists* were graduate students trained in a substantive area but who also studied quantitative methods, psychometrics, and research design. In the same vein, *twofers* (Aiken et al., 2008) are faculty who contribute to a substantive research area but also have acquired some training and expertise in statistics.

A historical example of a Maven is George Snedecor and his relationship to Sir Ronald Fisher's (1925) classic book *Statistical Methods for Research Workers*. Fisher's book is the one of the most influential statistics book of the 20th century. Ironically, Joan Fisher Box (1978) wrote in her father's biography that his book did not receive positive reviews. Gossett (Student, 1926), the developer of the *t* test and Fisher's friend, did provide a positive review of the book albeit with a caveat: "Dr. Fisher's book will doubtless be found in the laboratories of those who realise the necessity for statistical treatment of experimental results, but it should not be expected that full, perhaps even in extreme cases any, use can be made of such a book without contact, either personal or by correspondence with someone familiar with its subject matter" (p. 150).

Someone familiar with the book's subject matter was Snedecor. The success of Fisher's *Statistical Methods for Research Workers* was the result of textbook writers such as Snedecor contributing to its acceptance. Fisher Box (1978) referred to Snedecor as the

“midwife in delivering the new statistics in the United States” (p. 313). Lush (1972) described Snedecor’s *Statistical Methods* textbook as “a ‘how to’ book containing also a little ‘why,’ it nicely complements Fisher’s *Statistical Methods for Research Workers*” (p. 225). Lush attributed to a European researcher the quote, “When you see Snedecor again, tell him that over here we say, ‘Thank God for Snedecor; now we can understand Fisher!’ ” (p. 225).

Snedecor’s claim to Mavenhood does not rest solely with his textbook. Gertrude Cox was supervised by Snedecor and received the first degree in statistics awarded by Iowa State. In her reminiscence of Snedecor, Cox commented Snedecor “took two hours out of his busy schedule to explain the field of statistics to an 18-year-old student who was only curious” (Cox & Homeyer, 1975, p. 266), remarked Snedecor showed “his appreciation for the student confronting statistical methods for the first time” (p. 266), and revealed Snedecor “would work with non-mathematical faculty on their applied mathematical problems, and he took the lead in demonstrating the uses of statistics to other research investigators” (p. 272). Salsburg (2001) conceded that while Snedecor did not make many original statistical contributions, he was exceptional at presenting and mentoring the work of others. In other words, George Snedecor was a Maven.

One can think of modern examples of Mavens (e.g., Field, 2009). Mavens explain the paradox of why some statistical innovations are adopted quickly while other statistical innovations languish. While not highly persuasive, Mavens are disproportionately influential. Mavens would be aware of innovations in quantitative methods and theory. Innovations reviewed positively by Mavens would be recommended to substantive researchers and those innovations would become popular.

Mavens explain another paradox. Most quantitative articles have no discernible impact on the substantive research literature. A review by Mills et al. (2010) of citations from six major substantive journals and four major quantitative journals determined most substantive authors cite no quantitative articles and most quantitative articles are not cited by any substantive author. Indeed, most citations to quantitative articles are by the authors of other quantitative articles. Yet many of the citation classics in psychology are quantitative articles. *Psychological Methods* is by far the most influential quantitative psychology journal. In reviewing the more than 500 articles published in *Psychological Methods* to date, I identified 16 articles with more than 400 citations. These 16 articles account for half of all citations to *Psychological Methods*. What accounts for the success of these 16 articles? First, the 16 articles cover topics relevant to substantive researchers in psychology. Six of the 16 articles address issues relating to factor analysis and structural equation modeling; another seven articles address mediation, meta-analysis, handling missing data, or dichotomizing continuous variables. Only one of the 16 articles addresses NHST. Second, all but two of the 16 articles review data analysis methods rather than introduce a new statistical method. Third, the 16 articles are concrete in providing examples and illustrations of statistical analyses using actual or simulated data. Fourth, the 16 articles are prescriptive in making specific recommendations to substantive researchers. And fifth, Mavens are implicated in the authorship of these 16 articles. Most telling, four of the 16 *Psychological Methods* most cited articles were coauthored by Robert

MacCallum; together, his four articles have generated more than 3,500 citations.

To Communicate Better

What can be done to improve the communication of statistical innovations? Change is never easy but in the first chapter of his book on statistical innovations, Kline (2013) wrote “Maybe I am a naive optimist, but I believe there is enough talent and commitment to improving research practices among too many behavioral scientists to worry about unheeded calls for reform. But such changes do not happen overnight” (p. 25). Accepting the complexity of statistical methods, acknowledging the communication gap between quantitative methodologists and substantive researchers, and recognizing the role Mavens play in bridging the gap results in the following recommendations.

Highlight Solutions to Statistical Problems

Quantitative methodologists who propose a statistical innovation should highlight how the innovation solves a statistical problem or meets a statistical need for substantive researchers. Why? Statistical innovations for which there is demand are more likely to be adopted. ANOVA and structural equation modeling are two statistical innovation success stories from two different eras that succeeded because of the synchrony between availability and demand.

Use Real World Examples

Authors of statistical articles and journal editors should prioritize real-world examples of analyses and data sets. To establish the importance of evaluating correlations between predictors and outcome variables in multiple regression, Courville and Thompson (2001) showed that authors of published articles reached erroneous conclusions when they fail to do so. To demonstrate the impact of robust statistics on hypothesis testing, Wilcox and Keselman (2012) showed that analysis of published and unpublished data sets resulted in different conclusions for traditional versus robust statistical methods.

Show How to Do the Innovation

Some statistical innovations fail because quantitative methodologists do not show in concrete terms how to do the innovation. In introducing bootstrapping to readers of the British Psychological Association’s magazine *The Psychologist*, Wright and Field (2009) presented the R code and resulting output in sufficient detail to allow readers unfamiliar with R to replicate their analysis of a robust *t* statistic. In contrast, in introducing robust regression to readers of the *European Journal of Personality*, Wilcox and Keselman (2012) offered no specifics for conducting the analysis because “no single [robust] method is always optimal” (p. 173) and directed readers to Wilcox’s textbooks.

Create an Introductory Psychology Quantitative Journal

There is a genuine need for an introductory quantitative methods journal in psychology. The American Psychological Association

publishes *Psychological Methods*, but few of its articles are pitched at a level appropriate for most substantive researchers. The Association for Psychological Science publishes no quantitative methods journal. Their journal *Perspectives on Psychological Science* (PPS) sends a contradictory message about methods to prospective authors. In an interview in *ScienceWatch* (2012), PPS's editor, Barbara Spellman, attributed the journal's high citation rate in part to methodological articles. However, in another venue she stated "PPS gets many submissions about scientific methodology. Because it is not a 'methods journal' per se, most are politely rejected" (Spellman, 2012, p. 58).

One anonymous reviewer nominated *Understanding Statistics*; the journal folded in 2005 because the publisher was dissatisfied with the number of subscriptions (B. S. Everitt, personal communication, May 7, 2013). *Tutorials in Quantitative Methods* publishes an eclectic mix of introductory and specialized articles in two or three small issues a year. About one quarter of the articles are published in French. The journal has a low profile; the median number of citations to English language articles is three. *Behavior Research Methods* focuses primarily on computer technology and instrumentation. Two online journals—*Frontiers in Quantitative Psychology and Measurement* and *Practical Assessment, Research and Evaluation*—come closest to being introductory quantitative method journals aimed at substantive researchers in psychology. *Frontiers* charges fees for publishing some types of articles but both online journals can be freely accessed by anyone with a computer. The tremendous success of Field's (2009) introductory statistics textbook, robust citation counts for introductory statistics articles, and millions of viewers of *StatNotes* (according to the website of G. David Garson, n.d., para. 1) and other such introductory statistics websites suggest there is a market for a high profile, introductory statistical methods journal.

Make Better Use of Mavens

Promoting statistical innovations is Mavens' *raison d'être*. Mavens already serve as journal reviewers and authors of introductory tutorials in substantive journals. Acknowledged Mavens could contribute further to the editorial process by assisting quantitative methodologists in tailoring their articles for substantive researchers. Aspiring authors of quantitative articles could also learn from Mavens by modeling their highly cited articles. For example, Shadish, Phillips, and Clark (2003) conducted a case study of Campbell and Stanley's (1963) influential article on quasi-experimental designs. Another function Mavens could serve would be to monitor a website maintained by one of the psychological associations listing accessible quantitative articles by topic, something akin to socialpsychology.org. Inclusion of articles could be by nomination or by number of citations from substantive authors. Mavens could be invited to download introductory lectures, workshops and conference presentations to this website.

Mind the Gap

Recently, on a visit to a campus bookstore, I was surprised to see stacks of the venerable statistics textbook by Hays (1994). Many graduate students considered Hays to be challenging, while many quantitative faculty regard Hays fondly. What astonished me about seeing Hays in that campus bookstore was not that an

instructor had assigned a nearly 20-year-old textbook but that the textbook had been assigned to an *undergraduate* psychology statistics class! Perhaps the students in that class were exceptional, or perhaps the instructor was a particularly gifted communicator. Or perhaps this class instructor had inadvertently broadened the gap between quantitative methodologists as educators and budding substantive researchers as students by choosing this high-level textbook for an undergraduate class.

Different Audiences

A revealing exchange between quantitative methodologists recently appeared in the *Journal of Physiology*. Concerned over the quality of statistical analyses in articles submitted for editorial consideration, the senior statistics editor announced a series of introductory articles on statistical methods to be authored by the editor and medical statisticians. After the first three articles in the series appeared, Hopkins, Batterham, Impellizzeri, Pyne, and Rowlands (2011) wrote a scathing review that concluded "the [authors'] view of inference is flawed and outdated, the examples are inappropriate, there are serious errors and omissions, and the use of terms is imprecise" (p. 5329). Rather than mounting a point-by-point defense of their articles, Drummond and Tom (2011) reflected on how challenging it is to "express these statistical concepts clearly and simply"; cited a quotation of Mark Twain, "My books are water; those of the great geniuses are wine—everybody drinks water"; and offered a "final defense . . . 'horses for courses'". We are aiming at a different readership. We would not wish to mislead, and intend to correct any overt errors, but we are taking a gradual approach" (p. 5331). There should be an audience for journal articles that probe the complexity of statistical concepts. But there is an audience who wants clear and uncomplicated introductions to basic statistical methods. These are two different audiences.

Conclusion

Salsburg (2001) concluded his historical examination of statistics in the 20th century by prophesying new paradigms would challenge the dominance of quantitative methods: "[The statistical revolution in science] stands triumphant on feet of clay. Somewhere, in the hidden corners of the future, another scientific revolution is waiting to overthrow it" (p. 309). Some feel quantitative methods in psychology are already in decline (e.g., Shadish, 2007); others believe contenders such as qualitative approaches have earned the opportunity to challenge (Kidd, 2002). Recent developments including controversial finding of psi (Bem, 2011), admissions of data fraud and data fudging in social psychology (Ferguson, 2012), and perplexing failures to replicate established research findings (Yong, 2012) serve to highlight the vulnerability of and our dependence on statistical methods in psychology. Yet I believe this is a golden age for quantitative methods and methodologists. The challenge for this age is to decode messages of discontent and resistance to close the communication gap between quantitative methodologists and substantive researchers.

References

- Abelson, R. P. (1997). On the surprising longevity of flogged horses: Why there is a case for the significance test. *Psychological Science*, 8, 12–15. doi:10.1111/j.1467-9280.1997.tb00536.x

- Aiken, L. S., West, S. G., & Millsap, R. E. (2008). Doctoral training in statistics, measurement, and methodology in psychology: Replication and extension of Aiken, West, Sechrest, and Reno's (1990) survey of PhD programs in North America. *American Psychologist*, *63*, 32–50. doi:10.1037/0003-066X.63.1.32
- Altman, D. G., Goodman, S. N., & Schroter, S. (2002). How statistical expertise is used in medical research. *Journal of the American Medical Association*, *287*, 2817–2820. doi:10.1001/jama.287.21.2817
- American Psychological Association. (nd). *Report of the Task Force for Increasing the Number of Quantitative Psychologists*. Washington, DC: Author. Retrieved from <http://www.apa.org/research/tools/quantitative/quant-task-force-report.pdf>
- Anderson, D. R., Link, W. A., Johnson, D. H., & Burnham, K. P. (2001). Suggestions for presenting the results of data analyses. *Journal of Wildlife Management*, *65*, 373–378. doi:10.2307/3803088
- Appelbaum, M., & Sandler, H. (1996). Editorial. *Psychological Methods*, *1*, 3. doi:10.1037/h0092732
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182. doi:10.1037/0022-3514.51.6.1173
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*, 407–425. doi:10.1037/a0021524
- Bhattacharjee, Y. (2013, April 26). The mind of a con man. *New York Times Magazine*. Retrieved from <http://www.nytimes.com/>
- Boen, J. R., & Zahn, D. A. (1982). *The human side of statistical consulting*. Belmont, CA: Lifetime Learning.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*, 425–440. doi:10.1007/s11336-006-1447-6
- Busk, P. L. (1993). Improving statistical literacy among consultants and clients. *ASA Proceedings of the Section of Statistical Education*. Retrieved from <http://www.statlit.org/pdf/1993BuskASA.pdf>
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 365–376. doi:10.1038/nrn3475
- Byrne, B. M. (2012). Choosing structural equation modeling computer software: Snapshots of LISREL, EQS, Amos, and Mplus. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 307–324). New York, NY: Guilford Press.
- Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171–246). Chicago, IL: Rand McNally.
- Caplow, T., & McGee, R. J. (1958). *The academic marketplace*. New York, NY: Basic Books.
- Chow, S. L. (1996). *Statistical significance: Rationale, validity and utility*. Thousand Oaks, CA: Sage.
- Clay, R. (2005). Too few in quantitative psychology. *Monitor*, *36*, 26–27. Retrieved from <http://www.apa.org/monitor/>
- Cobo, E., Selva-O'Callaghan, A., Ribera, J.-M., Cardellach, F., Dominguez, R., & Vilardell, M. (2007). Statistical reviewers improve reporting in biomedical articles: A randomized trial. *PLoS ONE*, *3*, E332. doi:10.1371/journal.pone.0000332
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, *65*, 145–153. doi:10.1037/h0045186
- Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, *70*, 426–443. doi:10.1037/h0026714
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304–1312. doi:10.1037/0003-066X.45.12.1304
- Cohen, J. (1992). Fuzzy methodology. *Psychological Bulletin*, *112*, 409–410. doi:10.1037/0033-2909.112.3.409
- Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: β is not enough. *Educational and Psychological Measurement*, *61*, 229–248. doi:10.1177/0013164401612006.
- Cox, G. M., & Homeyer, P. G. (1975). Professional and personal glimpses of George W. Snedecor. *Biometrics*, *31*, 265–301. doi:10.2307/2529421
- Crutchfield, R. S., & Tolman, E. C. (1940). Multiple-variable design for experiments involving interaction of behavior. *Psychological Review*, *47*, 38–42. doi:10.1037/h0056488
- Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals and meta-analysis*. New York, NY: Routledge.
- Cumming, G., Fidler, F., Leonard, M., Kalinowski, P., Christiansen, A., Kleining, A., . . . Wilson, S. (2007). Statistical reform in psychology: Is anything changing? *Psychological Science*, *18*, 230–232. doi:10.1111/j.1467-9280.2007.01881.x
- DeCoster, J., Iselin, A.-M. R., & Gallucci, M. (2009). A conceptual and empirical examination of justifications for dichotomization. *Psychological Methods*, *14*, 349–366. doi:10.1037/a0016956
- Denker, J. (2007). *The world on a plate: A tour through the history of America's ethnic cuisine*. Lincoln, NE: University of Nebraska Press.
- Drummond, G. B., & Tom, B. D. M. (2011). Reply from Gordon B. Drummond and Brian D. M. Tom. *Journal of Physiology*, *589*, 5331–5332. doi:10.1113/jphysiol.2011.219063
- Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, *34*, 917–928. doi:10.1093/jpepsy/jsp004
- Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist*, *63*, 591–601. doi:10.1037/0003-066X.63.7.591
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272–299. doi:10.1037/1082-989X.4.3.272
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One*, *4*, e5738. doi:10.1371/journal.pone.0005738
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Sociometrics*, *90*, 891–904. doi:10.1007/s11192-011-0494-7
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. doi:10.3758/BF03193146
- Feick, L. F., & Price, L. L. (1987). The market maven: A diffuser of marketplace information. *Journal of Marketing*, *51*, 83–97. doi:10.2307/1251146
- Ferguson, C. J. (2012, July 17). Can we trust psychological research? *Time*. Retrieved from <http://ideas.time.com/>
- Fidler, F. (2005). *From statistical significance to effect estimation: Statistical reform in psychology, medicine and ecology* (Doctoral thesis, University of Melbourne, Melbourne, Australia). Retrieved from http://www.botany.unimelb.edu.au/envisci/docs/fidler/fidlerphd_aug06.pdf
- Fidler, F., Cumming, G., Burgman, M., & Thomason, N. (2004). Statistical reform in medicine, psychology and ecology. *Journal of Socio-Economics*, *33*, 615–630. doi:10.1016/j.socrec.2004.09.035
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, England: Sage.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London, England: Sage.

- Finch, S., & Cumming, G. (2009). Putting research in context: Understanding confidence intervals from one or more studies. *Journal of Pediatric Psychology, 34*, 903–916. doi:10.1093/jpepsy/jsn118
- Finch, S., Cumming, G., & Thomason, N. (2001). Reporting of statistical inference in the *Journal of Applied Psychology*: Little evidence of reform. *Educational and Psychological Measurement, 61*, 181–210. doi:10.1177/0013164401612001.
- Finch, S., Cumming, G., Williams, J., Palmer, L., Griffith, E., Alders, C., . . . Goodman, O. (2004). Reform of statistical inference in psychology: The case of *Memory & Cognition*. *Behavior Research Methods, Instruments & Computers, 36*, 312–324. doi:10.3758/BF03195577
- Fisher, R. A. (1925). *Statistical methods for research workers* (1st ed.). Edinburgh, Scotland: Oliver & Boyd.
- Fisher Box, J. (1978). *R. A. Fisher: The life of a scientist*. New York, NY: Wiley.
- Fiske, D. W. (1981). Methodologist: A growing career specialty. *American Psychologist, 36*, 318–319. doi:10.1037/0003-066X.36.3.318
- Ford, J. D., Ford, L. W., & D'Amelio, A. (2008). Resistance to change: The rest of the story. *Academy of Management Review, 33*, 362–377. doi:10.5465/AMR.2008.31193235
- Fritz, A., Scherndl, T., & Kuhberger, A. (2013). A comprehensive review of reporting practices in psychological journals: Are effect sizes really enough? *Theory & Psychology, 23*, 98–122. doi:10.1177/0959354312436870
- Gardenier, J. S., & Resnik, D. B. (2002). The misuse of statistics: Concepts, tools, and a research agenda. *Accountability in Research, 9*, 65–74. doi:10.1080/08989620212968
- Garson, G. D. (n.d.). *Statnotes: Topics in multivariate analysis*. Retrieved from <http://faculty.chass.ncsu.edu/garson/PA765/statnote.htm>
- Gladwell, M. (2002). *The tipping point*. New York, NY: Hatchette.
- Golinski, C., & Cribbie, R. A. (2009). The expanding role of quantitative methodologists in advancing psychology. *Canadian Psychology, 50*, 83–90. doi:10.1037/a0015180
- Gorman, B. S., & Primavera, L. (2010). The crisis in statistical education of psychologists. *General Psychologist, 45*, 21–27.
- Gravetter, F. J., & Wallnau, L. B. (2013). *Statistics for the behavioral sciences* (9th ed.). Belmont, CA: Wadsworth.
- Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate applications* (2nd ed.). New York, NY: Routledge.
- Harris, M. J. (1991). Significance tests are not enough: The role of effect-size estimation in theory corroboration. *Theory & Psychology, 1*, 375–382. doi:10.1177/0959354391013007
- Hays, W. L. (1994). *Statistics* (5th ed.). Belmont, CA: Wadsworth
- Henson, R. K., Hull, D. M., & Williams, C. S. (2010). Methodology in our education research culture: Toward a stronger collective quantitative proficiency. *Educational Researcher, 39*, 229–240. doi:10.3102/0013189X10365102
- Hopkins, W. G., Batterham, A. M., Impellizzeri, F. M., Pyne, D. B., & Rowlands, D. S. (2011). Statistical perspectives: All together not. *Journal of Physiology, 589*, 5327–5329. doi:10.1113/jphysiol. 2011.218529
- Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin, 105*, 302–308. doi: 10.1037/0033-2909.105.2.302
- Hyde, J. S. (2001). Reporting effect sizes: The roles of editors, textbook authors, and publication manuals. *Educational and Psychological Measurement, 61*, 225–228. doi:10.1177/0013164401612005
- Jaccard, J., & Guilamo-Ramos, V. (2002). Analysis of variance frameworks in clinical child and adolescent psychology: Issues and recommendations. *Journal of Clinical Child and Adolescent Psychology, 31*, 130–146. doi:10.1207/S15374424JCCP3101_15
- Kenny, D. A. (2008). Reflections on mediation. *Organizational Research Methods, 11*, 353–358. doi:10.1177/1094428107308978
- Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., . . . Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research, 68*, 350–386. doi:10.3102/00346543068003350
- Kidd, S. A. (2002). The role of qualitative research in psychological journals. *Psychological Methods, 7*, 126–138. doi:10.1037/1082-989X.7.1.126
- Killeen, P. R. (2005). An alternative to null-hypothesis significance tests. *Psychological Science, 16*, 345–353. doi:10.1111/j.0956-7976.2005.01538.x
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*, 746–759. doi: 10.1177/0013164496056005002
- Kline, R. B. (2013). *Beyond significance testing: Statistics reform in the behavioral sciences* (2nd ed.). Washington, DC: American Psychological Association. doi:10.1037/14136-000
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.
- Lance, C. E., & Vandenberg, R. J. (2009). *Statistical and methodological myths and urban legends*. New York, NY: Routledge.
- Landes, R. D. (2009). Passing on the passion for the profession. *American Statistician, 63*, 163–172. doi:10.1198/tast.2009.0031
- Lawrence, P. R. (1969). How to deal with resistance to change. *Harvard Business Review, 47*, 49–57.
- Lovie, A. D. (1979). The analysis of variance in experimental psychology: 1934–1945. *British Journal of Mathematical and Statistical Psychology, 32*, 151–178. doi:10.1111/j.2044-8317.1979.tb00591.x
- Lush, J. L. (1972). Early statistics at Iowa State University. In T. A. Bancroft (Ed.), *Statistical papers in honor of George W. Snedecor* (pp. 211–226). Ames, IA: The Iowa State University Press.
- MacCallum, R. (1998). Commentary on quantitative methods in I/O research. *The Industrial-Organizational Psychologist, 35*, 19–30.
- MacCallum, R. C., Zhang, S. B., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19–40. doi:10.1037/1082-989X.7.1.19
- Maraun, M., & Gabriel, S. (2010). Killeen's (2005). p_{rep} coefficient: Logical and mathematical problems. *Psychological Methods, 15*, 182–191. doi:10.1037/a0016955
- McMillan, J. H., & Foley, S. (2011). Reporting and discussing effect size: Still the road less traveled? *Practical Assessment, Research & Evaluation, 16*(14). Retrieved from <http://pareonline.net/>
- Meng, X.-L. (2009). Desired and feared: What do we do now and over the next 50 years? *American Statistician, 63*, 202–210. doi:10.1198/tast.2009.09045
- Mills, L., Abdulla, E., & Cribbie, R. A. (2010). Quantitative methodology research: Is it on psychologists' reading lists? *Tutorials in Quantitative Methods for Psychology, 6*, 52–60.
- Muenchen, R. A. (2013). *The popularity of data analysis software*. Retrieved from <http://r4stats.com/articles/popularity/>
- Muthén, B. (1989). Teaching students of educational psychology new sophisticated statistical techniques. In M. C. Wittrock & F. Farley (Eds.), *The future of educational psychology* (pp. 181–189). Hillsdale, NJ: Erlbaum.
- Novotney, A. (2008). *Postgrad growth area: Quantitative psychology*. Retrieved from <http://www.apa.org/gradpsych/2008/09/quantitative-psychology.aspx>
- Odgaard, E. C., & Fowler, R. L. (2010). Confidence intervals for effect sizes: Compliance and clinical significance in the *Journal of Consulting and Clinical Psychology*. *Journal of Consulting and Clinical Psychology, 78*, 287–297. doi:10.1037/a0019294
- Onwuegbuzie, A. J., & Wilson, V. A. (2003). Statistics anxiety: Nature, etiology, antecedents, effects, and treatments: A comprehensive review of the literature. *Teaching in Higher Education, 8*, 195–209. doi: 10.1080/1356251032000052447

- Osborne, J. W. (2010a). Challenges for quantitative psychology and measurement in the 21st century. *Frontiers in Psychology, 1*, 1. doi:10.3389/fpsyg.2010.00001
- Osborne, J. W. (2010b). Data cleaning basics: Best practices in dealing with extreme scores. *Newborn and Infant Nursing Reviews, 10*, 37–43. doi:10.1053/j.nainr.2009.12.009
- Ozonoff, D. (2006). Quality and value: Statistics in peer review. *Nature, 441*. doi:10.1038/nature04989
- Piderit, S. K. (2000). Rethinking resistance and recognizing ambivalence: A multidimensional view of attitudes toward an organizational change. *Academy of Management Review, 25*, 783–794. doi:10.5465/AMR.2000.3707722
- Pituch, K. A. (2004). Textbook presentations on supplemental hypothesis testing activities, nonnormality, and the concept of mediation. *Understanding Statistics, 3*, 135–150. doi:10.1207/s15328031us0303_1
- Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin, 112*, 160–164. doi:10.1037/0033-2909.112.1.160
- Rossen, E., & Oakland, T. (2008). Graduate preparation in research methods: The current status of APA-accredited professional programs in psychology. *Training and Education in Professional Psychology, 2*, 42–49. doi:10.1037/1931-3918.2.1.42
- Rucci, A. J., & Tweney, R. D. (1980). Analysis of variance and the “second discipline” of scientific psychology: A historical account. *Psychological Bulletin, 87*, 166–184. doi:10.1037/0033-2909.87.1.166
- Salsburg, D. (2001). *The lady tasting tea: How statistics revolutionized science in the twentieth century*. New York, NY: William Freedman.
- Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology, 57*, 1–10. doi:10.1037/a0018082
- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods, 1*, 115–129. doi:10.1037/1082-989X.1.2.115
- Schmidt, F. L., & Hunter, J. E. (1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. In L. L. Harlow, S. A. Mulaik & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 37–64). Mahwah, NJ: Erlbaum.
- ScienceWatch. (2012, June). *Barbara A. Spellman on the impact of Perspectives on Psychological Science*. Retrieved from <http://www.sciencewatch.com/inter/jou/2012/12junPerspectPsychSci/>
- Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105*, 309–316. doi:10.1037/0033-2909.105.2.309
- Shadish, W. R. (2007, August). Social influences in science: The decline of quantitative methods in psychology. In W. T. Hoyt (Chair), *Improving research validity—Lessons from the deterioration effects controversy*. Paper presented at the annual meeting of the American Psychological Association, San Francisco, CA. Retrieved from <http://www.scu.edu/SCU/Projects/Hospice/APA%202007%20TIDE%20SYMPOSIUM.pdf>
- Shadish, W. R., Phillips, G., & Clark, M. H. (2003). Content and context: The impact of Campbell and Stanley. In R. J. Sternberg (Ed.), *The anatomy of impact: What makes the great works of psychology great* (pp. 161–176). Washington, DC: American Psychological Association. doi:10.1037/10563-009
- Shaver, J. P. (1993). What statistical significance testing is, and what it is not. *Journal of Experimental Education, 61*, 293–316.
- Shuker, D. M. (2012). Book Review: A. P. Beckerman, O. L. Petchey, *Getting Started With R: An Introduction for Biologists*; R. Thomas, I. Vaughan, J. Lello, *Data Analysis With R Statistical Software. A Guidebook for Scientists*; A. Field, J. Miles, Z. Field, *Discovering Statistics Using R. Animal Behaviour, 84*, 1597–1600. doi:10.1016/j.anbehav.2012.09.019
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359–1366. doi:10.1177/0956797611417632
- Spellman, B. A. (2012). Introduction to the special section: Data, data, everywhere . . . especially in my file drawer. *Perspectives on Psychological Science, 7*, 58–59. doi:10.1177/1745691611432124
- Sternberg, R. J. (1992). Psychological Bulletin’s top 10 “hit parade.” *Psychological Bulletin, 112*, 387–388. doi:10.1037/0033-2909.112.3.387
- Student. (1926). Statistical methods for research workers. *Eugenics Review, 18*, 148–150.
- Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement, 55*, 525–534. doi:10.1177/0013164495055004001
- Thompson, B. (1999). Why “encouraging” effect size reporting is not working: The etiology of researcher resistance to changing practices. *Journal of Psychology: Interdisciplinary and Applied, 133*, 133–140. doi:10.1080/00223989909599728
- Thompson, B. (2001). Significance, effect sizes, stepwise methods, and other issues: Strong arguments move the field. *Journal of Experimental Education, 70*, 80–93. doi:10.1080/00220970109599499
- Thompson, K., & Edelstein, D. (2004, Summer/Fall). A reference model for providing statistical consulting services in an academic library setting. *IASSIST Quarterly, 35*–38. Retrieved from <http://www.iassistdata.org/>
- Tukey, J. W. (1986). What have statisticians been forgetting? In L. V. Jones (Ed.), *The collected works of John W. Tukey: Philosophy and principles of data analysis 1965–1986* (Vol. IV, pp. 587–599). Belmont, CA: Wadsworth.
- Wasserman, R. (2013). Ethical issues and guidelines for conducting data analysis in psychology research. *Ethics & Behavior, 23*, 3–15. doi:10.1080/10508422.2012.728472
- Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist, 53*, 300–314. doi:10.1037/0003-066X.53.3.300
- Wilcox, R. R. (2002, April). Can the weak link in psychological research be fixed? *APS Observer*. Retrieved from <http://www.psychologicalscience.org/observer/getArticle.cfm?id=931>
- Wilcox, R. R., & Keselman, H. J. (2012). Modern regression methods that can substantially increase power and provide a more accurate understanding of associations. *European Journal of Personality, 26*, 165–174. doi:10.1002/per.860
- Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594–604. doi:10.1037/0003-066X.54.8.594
- Wright, D. B., & Field, A. P. (2009). Giving your data the bootstrap. *Psychologist, 22*, 412–413.
- Yong, E. (2012, May 17). Replication studies: Bad copy. *Nature, 485*, 298–300. Retrieved from <http://www.nature.com/>
- Zimiles, H. (2009). Ramifications of increased training in quantitative methodology. *American Psychologist, 64*, 51. doi:10.1037/a0013717

Received October 18, 2012

Revision received June 4, 2013

Accepted July 12, 2013 ■