Why All Leadership Faculty Should Initiate A Reform of Quantitative Methods Courses

Stanley Pogrow
San Francisco State University (Emeritus, University of Arizona)

It is time to reform the quantitative methods courses in leadership programs – typically, these are statistics courses with arcane statistics textbooks. There is growing evidence that these "rigorous" scientific methods actually mislead practice because the vast majority of practices found to be "effective" or "evidence-based" using these methods do not work in schools – including those validated by the federal What Works Clearinghouse. Fortunately, there are other quantitative methods that are more intuitive, and more accessible to leaders, leadership students, and leadership faculty – methods that are more relevant for improving practice and identifying "evidence-based" practices that are likely to actually improve schools. However, those who control the content of quantitative methods courses tend to be those with the most technical expertise in the traditional methods and will oppose any methodological alternatives as being "non-rigorous." The newer quantitative methods will not be emphasized in EdD and Masters programs until all leadership faculty exert greater voice on the content of such courses. A first step is to stop calling quantitative methods courses "Statistics" courses. They should be "Applied Quantitative Methods" courses. This non-technical article (a) outlines the problems with traditional statistics, (b) highlights some of the newer and simpler quantitative methods that are more relevant for improving schools, and (c) describes an alternative textbook as a key resource for transforming the quantitative methods course.

Keywords*:* quantitative methods, qualitative methods, statistical methods, statistical significance

This non-technical article is intended for the vast majority of leadership faculty who have little to no background in statistics. It is intended for those whose expertise is in educational leadership and management – i.e., the vast majority of leadership faculty. This article is a call to arms for the majority of faculty to wrest control of leadership quantitative methods courses away from those steeped in traditional statistical analysis and thereby make the courses more relevant to actually improving schools and solving problems of practice.

Quantitative methods courses are usually listed as a statistics course and emphasize the esoteric traditional methods and analyses used by psychologists in laboratory research and medical researchers to test the effects of an individual medicine. Those who question whether these methods are relevant to school leaders are usually cowed when informed that these statistical methods represent rigorous science, and that questioning their use is an indication of an academically weak program. While these methods are indeed rigorous for the purposes for which they were intended – largely PhD forms of basic research – these methods are not valid or useful for informing decisions in complex organizations such as schools and hospitals. In fact, there is growing evidence that the results from the traditional statistical methods have been misdirecting educational practice – particularly results identifying evidence-based practices. Indeed, it will be shown that the more "rigorous" the statistical analyses, and the more prestigious the journals in which the evidence of an effective practice is published, the *less* likely the findings are valid – e.g., the less likely the practice will in fact improve practice. Reliance on traditional statistical methods is not rigorous science but a misuse of science.

The traditional statistics course is the last remaining vestige of PhD programs imposed on EdD programs. Other courses in leadership graduate programs have been updated to meet the needs of leaders seeking to improve their organizations and solve problems of practice. While quantitative methods courses are essential, such courses need to recognize that traditional statistical methodology is but a subset of quantitative methods – and that there are newer quantitative methods that are far more useful towards helping leaders improve their organizations, what should be the major focus of EdD quantitative methods courses.

However, those with PhDs and/or an extensive background in statistics tend to control the content of the quantitative methods courses in leadership programs – and they will not give up this power without a fight. They simply will not accept that their methods, which are appropriate for PhD programs, are generally not useful for solving real world problems in schools. It is time for the rest of us who are experienced in the actual practices of improving educational organizations to impose our will and (a) demand reform of the quantitative methods course, and (b) use our expertise to have a major voice in the course content to ensure that more relevant forms of quantitative methods are taught.

This call to arms is not written by someone who is against quantitative research. Quite the contrary. I am a math major who has taught statistics in several universities. I have presented twice at international conferences put on by the American Statistical Association, and may be the first educator to have been published in its flagship journal, *The American Statistician*, read by statisticians across the disciplines (Pogrow, 2019a). The ideas contained herein are based on that recent article. These ideas provide a basis for those with a non-technical background to understand:

- How traditional statistical practices mislead practice,
- How to make the case for reform, and

- What the key characteristics of an alternative quantitative course would be that is geared to the realities of educational practice.

Finally, a resource is presented to support this grassroots reform. It is an alternative quantitative methods textbook that presents quantitative analysis from the perspective of leaders' needs. This resource can be used to inform the reform effort and as the primary text for an authentic quantitative methods course.

Evidence for the Need to Reform Quantitative Methods Courses

Quantitative methods have traditionally been viewed as the province of mathematically sophisticated methods and individuals. Educators have:

- Accepted that esoteric mathematical and procedural complexity is a necessary condition for conducting rigorous science;
- Accepted the published findings of effective practices based on such research as gospel;
- Allowed those with the most expertise in statistics and research methodology to determine the content of EdD quantitative methods courses; and
- Assumed that the quantitative methods course in an EdD and a Masters leadership program should be a course in statistics.

However, statistics courses tend to overwhelm students and convince them that quantitative analysis is beyond them. Statistics courses create resentment against the use of quantitative analysis and convince most EdD students to conduct a qualitative dissertation. This backlash is at odds with the reality that educational leadership practice is becoming ever more quantitative. Increasingly, leaders are expected to use quantitative research evidence as the basis for making decisions on how to best improve their schools.

Of course, simply because content is difficult to grasp is not in and of itself an argument for minimizing its emphasis. In this case, emphasizing traditional statistics is inappropriate because traditional statistical methodology does not generally produce useful or valid evidence for leadership decision-making. Quite the opposite! Current sophisticated statistical methods tend to produce highly misleading evidence and conclusions as to whether interventions are likely to improve schools under real-world conditions – especially research published in the top research journals. Traditional statistical methodology was largely used by psychologists to produce evidence in lab settings where it is possible to control the potential intervening variables.

To understand the problem in the current forms of statistical methodology promoted by statisticians and education psychologists, consider the following:

> There's a famous joke about a dairy farmer who, hoping to increase milk production, seeks the help of a theoretical physicist at the local university. After carefully studying the problem, the physicist tells the farmer, "I have a solution, but it only works if we assume that cows are spheres".

Cows are not spheres and schools are not static, stable environments where you can control anything for more than a few minutes. Much of the statistical and methodological complexity is an effort to simulate such control – but in the end the rigorous traditional methodology conducts analyses about a hypothetical mathematical world. Stated more simply, gold standard experimental methodology, typically referred to as Randomized Controlled Trials (RCT), with

all of its methodological and statistical wizardry cannot describe the reality of the myriad social and programmatic interactions within complex organizations such as schools or hospitals.

Even worse, the statistical criteria being used to interpret whether the numbers generated by statistical analysis such as the F-test, the t-test, regression, etc., indicate that an improvement or relationship is significant – i.e., statistical significance – grossly overestimates the importance of findings. Therefore findings of "significant benefits," or "significant positive relationships" that are used as the basis for concluding that a practice is evidence-based and should be adopted by leaders, actually mislead the field. Pogrow (2017; 2018; 2019a) summarizes this and other problems, such as self-serving statistical adjustments, in using traditional statistics to guide practice. This misdirection and overstatement of effects in the highest quality and most influential research is not only a problem in education but across the disciplines.

There is a growing body of scholarship that is now actively criticizing the use of traditional statistical methods for guiding clinical practice in a variety of fields – including psychology itself. Scholars recently started discovering that the vast majority of the most influential published laboratory research in a variety of fields cannot be replicated in subsequent experiments in the lab and/or were not reflecting what clinicians were seeing in practice. If evidence of "evidence-based practice" cannot be replicated in the lab, whatever benefits were identified in the research are not going to replicate in schools.

In other words, traditional statistical methods and procedures make it easy to legitimately produce high quality research that claims to have found a new discovery where there is none – and that is infecting all of science. The replication crisis has gotten to be so pronounced that the American Statistical Association has just called for research to stop relying on statistical significance ($p<.05$) (Schirm, Lazar, & Wasserman, 2019). Pogrow (2019a) has also called for eliminating the use of small benefits (effect sizes) to indicate program effectiveness.

There is also a major replication crisis in education. In 2010 the U.S. Department of Education funded the dissemination of 67 interventions based on research findings of positive impacts on students' achievement – research findings that had been validated by the scientifically rigorous criteria employed by the federal What Works Clearinghouse (WWC). The 2018 final evaluation of the impact of these interventions (Bouley, et al., 2018) was that of the 67 interventions with rigorous statistical evidence of effectiveness:

- 58 of the 67 produced no statistically significant increase in student achievement. However, given the inadequacy of statistical significance, this author has developed an alternative statistical measure of practical benefit, which indicates whether any increases would be sufficiently large to be noticeable and of value (Pogrow, 2019a). Further analysis by this author indicated that…
- 63 of the 67 produced no noticeable benefit in student achievement.

Nor were these poor results the fault of practitioners. The evaluation concluded that these poor results occurred despite quality implementation by the schools.

This comprehensive evaluation means that (a) any leader that adopted one of these evidence-based interventions would have made a terrible mistake in the overwhelming majority of cases, and (b) the evidence provided by traditional statistical methodology generally does not replicate in practice. This failure of traditional forms of evidence is particularly problematic in an era where federal legislation such as ESAA are requiring schools to implement evidence-based practice.

Probably the most consequential misdirection of practice from rigorous research evidence was the widespread adoption nationally of the Success for All reading reform intervention for high-poverty schools over a 25-year period. The adoption of this intervention was based on research evidence in top research journals demonstrating its success. The What Works Clearinghouse characterized this research as demonstrating the strongest evidence of success, and the widespread use of Success for All was the research community's shining example of the value of research. However, research by Pogrow (2002) and Boulay, et al. (2018) found that this expensive program had never actually been effective in practice and that a surprising number of sites had quickly decided to drop it. Districts that adopted it wasted hundreds of millions of dollars and shortchanged the education of cohorts of their most needy students.

Furthermore, while traditionalists consider RCT research to be the "most rigorous" form of research, it only produces differences/improvements that are too tiny to be of practical benefit—and these benefits are dwarfed by the statistical errors used to produce the result. Lortie-Forgeus and Inglis (2019) found that among all the 141 large-scale RCTs aimed at improving educational outcomes in grades K-12 funded by the UK and the U.S. National Center for Educational Evaluation and Regional Assistance, the average effect size (ES) was a miniscule .06. Considering that an ES of .2 indicates a difference that is "difficult to detect" (Cohen, 1988), .06 means that RCT research, which is the most expensive, technical, and time consuming form of research, produces results that are less than one-third of "difficult to detect." Such findings are useless for practice in complex organizations such as schools and even in the medical field.

However, statisticians routinely consider such trivial results to be important. For example, the key research claiming to show that charter schools are better for Black students than traditional public schools did so on the basis of a difference that was a tenth of "difficult to detect." The What Works Clearinghouse also considers trivial effects to indicate that something is working – i.e., effective. In other words, researchers now routinely make claims of discoveries of effectiveness using statistical criteria that are too trivial to have any practical importance and therefore mislead practice. If practitioners and policy makers knew what the numbers actually represented, they would not, and should not, consider using, advocating, funding, or endorsing the vast majority of evidence-based practices that "rigorous" research has deemed to be effective.

Despite this growing body of evidence that "rigorous" sophisticated advanced traditional statistical methodology has largely failed to provide useful findings for improving practice and has misdirected policy and decision-making, those who generally teach and/or decide what the nature of quantitative methods courses defend these methods as the only form of rigorous science. However, these methods are *not* scientifically rigorous for applied research in improving complex organizations such as schools and should no longer monopolize leadership quantitative methods courses.

Where Do We Go From Here?

Quantitative methods are essential for improving practice. It is now clear that quantitative methods courses in leadership programs need more than a tweak. They need a reconceptualization that builds upon the traditions of management and decision-making theory and perspectives. Fortunately, there are new methods for generating and using quantitative

evidence that are more appropriate for improving practice in complex organizations such as schools and even in medicine. A good model is hospitals' efforts to improve the delivery of health care. Like schools, hospitals are also complex organizations that seek to improve the outcomes for their clients. These newer methods are generally referred to as improvement science.

One major success of improvement science in the medical field is obstetrics' reduction of infant mortality. Obstetrics has made more progress in saving lives over the past 50 years than any other branch of medicine (Gawande, 2007). However obstetrics never used controlled experiments or sophisticated statistics. It used checklists and pragmatic rapid prototyping of alternative approaches, as well as quick analysis of results using simple descriptive statistics and rapid sharing of results across networks of obstetrics departments.

Such success validates the use of the more modern quantitative methods of improvement science for improving complex organizations, such as schools, using the most basic statistics and very pragmatic, intuitive analyses. The goal is to quickly discover innovative approaches that produce improvements that are so clearly noticeable and consistent with goals that there is little reason to conduct any statistical analysis beyond basic descriptive statistics, such as averages and standard deviations. Human common sense and leadership judgment can determine what "clearly noticeable" means much better than a finding of $p<.01^{**}$. Indeed, it can be argued that if a leader needs a statistician to discern whether quantitative evaluation results indicate that an intervention is effective — it probably isn't.

These newer methods can be considered to be a new epistemology of applied quantitative research. They are more scientifically rigorous because when the findings are replicated they provide more valid and useful observations about the real world that lead to better decisions by leaders on how to improve their schools. These newer quantitative methods have an impressive track record of improving clinical practice in complex organizations and in time-sensitive complex processes in a variety of disciplines — even in medicine.

Pogrow (2018, 2019a,b) has incorporated these newer ideas for quantitative research and has developed the statistical criterion of practical benefit to replace statistical significance and practical significance. The criterion of practical benefit includes a number of simple measures for leaders to use to determine whether the findings of specific published research are likely to benefit their schools in noticeable ways. He also shows how leaders can apply these newer quantitative research methods to their schools' data to design and test improvement strategies, and how to use those methods to develop useful EdD dissertations (and Masters theses).

Reforming Leadership Quantitative Methods Courses

Leadership quantitative methods courses need to be very different from what currently exists. Such courses need to shift from complex traditional statistical methodology to the more intuitive methods of improvement science and quantitative reasoning that have proven to be more valuable in improving complex organizations.

Unfortunately, the perspectives and needs of statisticians/researchers seeking to preserve their traditions and beliefs, and those of leaders seeking to produce noticeable improvement, are now at odds. Several such examples of the latter divergence previously discussed are leaders' need for noticeable benefits and for evidence of replications in organizations like theirs, instead of single studies seeking to establish causation that exaggerate

the actual benefits found in the research. As a result, reforming quantitative methods courses should *not* start from the perspective of statistics or educational psychology, but from the perspective of how leaders – those successful in using data to make major improvements in their organizations – employ quantitative methods.

If EdD programs are to prepare leaders who are better able to use quantitative evidence to improve their schools and solve problems of practice, it is critical to stop viewing the quantitative methods courses as the province of just those with highly technical statistical skills. All faculty need to work together to develop a more intuitive link between the newer, less technical, quantitative methods and the improvement of practice as a scientific endeavor. The techniques need to be ones that focus on schools as they actually are, not as spherical cows. The quantitative skills that need to be taught include how to:

- Use basic descriptive stats to mine datasets at the local, state, and federal levels to discover and precisely define inequalities and gaps, and progress;
- Use design thinking/improvement science to develop innovative solutions to try to address real problems of practice, and set up procedures to continuously review the performance of such solutions and to modify as needed;
- Develop measures and dashboards to track implementation and performance of innovative practice, and develop ways to share and disseminate such data in real time across the organization;
- Critique published research evidence in terms of its practical benefit; and
- Compare a given school's/district's performance to ones that are similar demographically.

In other words, it is now more important to know how to define, create, manage, and use data across one's organization than to know advanced statistics and statistical methodology, or how to use SPSS.

The above skills should not just be relegated to the quantitative methods course. Any course that discusses evidence-based practices should incorporate some of these ideas. Indeed, all faculty should reexamine their conclusions as to whether the practices they consider to be evidence-based actually are. This would make quantitative methods an integral part of an EdD program's ongoing focus on solving problems of practice.

Rethinking quantitative methods in such a fashion requires collaboration among all faculty in an EdD program. It requires the non-statisticians on the faculty to use their instincts from what they know about practice to take a strong role in re-designing the quantitative methods course and integrating the key ideas of improvement science, innovation design, and continuous improvement.

A starting point for reforming quantitative methods courses would be to cease calling them statistics courses and stop using statistics textbooks. An alternative quantitative methods textbook that is designed around the perspective and needs of leaders has been developed—Authentic Quantitative Analysis for Leadership Decision-Making and EdD Dissertations, (Pogrow, 2018). This textbook is published by the International Council of Professors of Educational Leadership (ICPEL). (To order go to lulu.com, and search "Pogrow".) A Masters level version is under development.

As this is written, approximately 12 EdD programs around the country have reformed their quantitative methods course around the use of this text. In each case it has been a political struggle to be allowed to do so — and such efforts do not always succeed. However, it is a battle worth fighting. For ammunition to make the case for reform, all leadership faculty can

download the following two of my open-source recent articles in top journals that summarize the problems with traditional PhD methods. Faculty can then discuss the ideas among themselves and share the articles with key college and university leaders: https://doi.org/10.14507/epaa.25.2517, and https://doi.org/10.1080/00031305.2018.1549101 The articles are written in an easy-to-understand language. Faculty can also contact me for support at *spogrow@sfsu.edu*

References

Boulay, B., et al. (2018). *The Investing in innovation fund: Summary of 67 evaluations*. Final Report. Institute of Education Sciences: U.S. Department of Education.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Gawande, A. (2007). *Better: A surgeon's notes on performance*. New York, NY: Metropolitan Books.

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned?. *Educational Researcher, 48*(3), 158-166.

Pogrow, S. (2002). Success for All is a failure. *Phi Delta Kappan, 83*(6), 463-468.

Pogrow, S. (2017). The failure of the U.S. education research establishment to identify effective practices: Beware of effective practice policies. *Education Policy Analysis Archives, 25*(5), Jan. 2017. http://dx.doi.org/10.14507/epaa.25.2517

Pogrow, S. (2018). *Authentic quantitative analysis for education leadership decision making and Ed.D. dissertations: A practical, intuitive, and intelligible approach. Edition 2,2*. ICPEL Press.

Pogrow, S. (2019a). How practical significance (Effect Size) misleads clinical practice: The case for switching to practical benefit to assess applied research findings, in Moving to a world beyond P<0.05. Schirm, A., Lazar, N., & Wasser4man, R.L. (eds). *The American Statistician, 73*, 1-19. DOI: 10.1080/00031305.2019.1583913

Pogrow, S. (2019b). *Reforming the design, application, and teaching of quantitative methods for improving practice and EdD programs: A new epistemology of applied quantitative methods, in Who Controls the Preparation of Education Administrators?* Black, W. and Danzig, A. (eds). IAP: Charlotte, NC.