Educational Leadership and Administration

Teaching and Program Development

The Journal of the California Association of Professors of Educational Administration

Volume 32
May, 2020

Statistics Course Improvement for School Administrator Preparation

Jianjun Wang
California State University, Bakersfield

Accompanied by increasing demands on school administrator preparation and rapid development of computer technology, educational statistics courses are exposed to unprecedented pressures for changing both curriculum content and computing platforms. In this article, the intended curriculum is reviewed according to data analysis expectations from state and national guidelines. Past recommendations on statistics instruction are examined to justify the need for quantitative research skills in school administrator preparation. The curriculum implementation is further investigated to reflect a fundamental revision of statistics content by the American Statistical Association. The article ends with an overview of the cutting-edge software development in R that is likely to reshape the future data processing, text analytics, and graphical display for school administrators.

Keywords: statistics curriculum, R computing, data analysis, text analytics

School administrators typically start their careers in teaching before completing a master's or doctoral degree in educational administration. During the process, Polnick and Edmonson (2005) asserted, "statistics courses as they are usually taught in graduate schools of education are not designed for the school principal" (p. 40). Thus, an important question has been raised of the need for statistical training to prepare educational administrators. The debate on the value of statistical courses is further extended to the traditional rivalry of quantitative and qualitative studies because some stakeholders prefer to read success stories over numeric findings in administrative reports (Hiller, DeChurch, Murase, & Doty, 2011).

Nonetheless, it is important to recognize that no subject, including statistical knowledge, remains stagnant. Prior to entering the 21st century, researchers already projected that "public agencies will need administrators who have sound backgrounds in quantitative data analysis and in computer usage" (Hy, Waugh, & Nelson, 1987, p. 139).

The Intended Curriculum of Statistical Training

According to the International Bureau of Education (2019), intended curriculum is indicated by a set of "formal documents which specify what the relevant national education authorities and society expect" (p. 1). In California, the Commission on Teacher Credentialing has developed the California Administrator Performance Assessment (CalAPA) to measure students' mastery of California Administrator Performance Expectations. With the credential requirements, educational administration programs are designed, in part, to help students pass CalAPA. The course setting, under the CalAPA model, provides an overarching conceptual framework to progressively refine the administrative candidate's thinking and decision-making skills.

At the beginning step, Leadership Cycle 1 of CalAPA calls for "Analyzing Data to Inform School Improvement and Promote Equity" (California Commission on Teacher Credentialing, 2019, p. 1). While school improvement is often contrasted to a baseline from the past, promoting equity may span across different demographic and school variables. Hence, systematic training is needed to investigate education factors from multilevel data in both time and space dimensions.

Beyond the state level, the National Policy Board of Educational Administration (NBPEA) insists that principals should be taught processes for experimenting and learning from real world data to meet challenges of the work environment. More explicitly, Polnick and Edmonson (2005) observed that "An essential expectation elaborated in the NPBEA training guidelines was the need for practicing principals to develop basic statistics and data analysis skills that will assist them in their day-to-day operations of the school" (p. 40).

While CalAPA and NPBEA addressed the expectation of effective leadership programs at state and national levels, Bernerth (2018) attached more emphasis on data gathering and analysis at the local level because "organizational decisions that have significant personal and financial implications are often made as a result of empirical research" (p. 133). For instance, Los Angeles Unified School District decided to show data on student improvement each year on how individual schools are helping students progress academically (Burke, 2019). A thorough analysis of the learning outcome involves a proper control of confounding variables, such as (1) race, age, gender, and primary language identity for students, (2) subject competency and years of instruction for teachers, and (3) funding resources and average class sizes in various schools. Without basic statistical knowledge to disentangle these variables, administrative decisions could be misguided to fit an ambiguous situation that precluded examination of these key characteristics across student, teacher, and school levels.

Besides the primary data analyses, school administrators may benefit from literature review to borrow the wheel from others. Although literature review typically relies on document reading, reflection and qualitative research, statistical training may play an indispensable role in reporting the result aggregation. For example, meta-analysis is a widespread statistical method for combining research outcomes from multiple studies. On December 20, 2019, an online search of the California State University libraries showed "gender difference" in the title of 29,909 books, articles, and reports. If a reviewer can use 10 words to summarize each item, the literature description may take over 600 pages, making the information nearly impossible to synopsize through qualitative inquires. However, metaanalysis simplifies the literature summary into an effect size across a massive number of studies (Kugler, Reif, Kaschner, &

Brodbeck, 2018). "When the treatment effect (or effect size) is consistent from one study to the next, meta-analysis can be used to identify this common effect" (Elsayir, 2015, p. 630). More recently, effect size computing is supported by online calculators (Lenhard & Lenhard, 2016).

Due to advancement in statistical computing, quantitative findings have become more readily available in the literature. The concept of data has become more inclusive, as Yoshikawa, Weisner, Kalil, and Way (2013) maintained that "The world is not inherently qualitative or quantitative; it is the act of human representation through numbers or non-numeric signifiers like words that make aspects of the scientific enterprise qualitative or quantitative" (p. 4). Hence, intentional depreciation of research methods, such as the ones from statistics courses, will inevitably result in the impediment of inquiry outcomes along the line of paradigm division.

In summary, the examination of intended curriculum for educational administrator preparation, per guidelines of CalAPA and NBPEA, does not support exclusion of statistical training in the graduate program. The need of qualitative and quantitative methods depends on the nature of research questions. Rather than tweaking a question to fit a convenient method, and thus, exclude statistics applications, school leaders should be equipped with well-rounded tools for choosing or creating pertinent methods to handle practical questions. By definition, "Statistics is the science of making decisions by collecting, analyzing and making inferences from data" (Stats, 2019, p. 1).

Reality of the Implemented Statistical Course Offerings

Despite the consensus in the intended curriculum to include statistical training for school administrator preparation, statistics is often taught by educational statisticians, rather than someone in educational administration. The intent is to prepare principals to effectively analyze and report their findings to various stakeholders (Creighton, 2001; Holcomb, 2004). To date, no attempt has been made in the curriculum setting to resolve the competition for more instructional time among different educator preparation programs. As Polnick and Edmonson (2005) reported, "While the NPBEA standards require principals to look at statistics and data analysis, very little training on how to gather and analyze data to make informed decisions is provided in the training manual or in many preparation programs" (p. 40). Even if a real dataset is included from a school, teacher candidates may place more interest in analyzing student performance at the class level while principal candidates show more interests in school variables. To reflect the specialty in educational administration, it has been suggested that "principals should be taught processes for experimenting and learning from real world data to meet the challenges of the work environment" (Polnick & Edmonson, 2005, p. 39).

However, data analysis skills seem inadequate because "too few school leaders have had the opportunity to acquire in their graduate work or have seen [data analysis] modeled in their own experiences" (Holcomb, 2004, p. 27). With a purpose of improving statistical course offerings, McNamara and Thompson (1996) proposed seven guidelines:

(1) Emphasizing data analysis (Statistics is a set of methods used to analyze real-world data, which allows practitioners to focus on producing accurate results to inform school improvement);

(2) Using real world data (Basic statistics courses should be taught as an integral part of the principal preparation program using real-world data that principals encounter in problem-solving and decision-making tasks in their job performance);

(3) Focusing on descriptive statistics (Principals typically use data on all students to solve pressing problems and to make decisions for their current academic year);

(4) Using accurate descriptions (The previous three properties needed to accurately describe a univariate distribution including the measure of center, measure of spread, and the shape of the distribution);

(5) Learning exploratory data (Viewing data using open-ended assumptions reveals truth about random fluctuations, error and other confusion often encountered in school data);

(6) Using graphic displays (This guideline emphasizes the importance of using data graphics in all aspects of real-world data analysis); and

(7) Reporting outliers (This guideline emphasizes why a principal should learn to analyze and report outliers).

While these guidelines are well-intentioned to increase the practical value of statistics training, Points 1 and 2 are not mutually exclusive, particularly on the duplication of emphasis in learning opportunities for real-world data analyses. The outlier identification in Point 7 is also a byproduct from examining the measure of spread in Point 4. Likewise, using graphic displays in Point 6 happens quite frequently in portraying probability distributions, but it can also occur to qualitative data, such as word cloud plots in non-statistical contexts (e.g., Jayashankar & Sridaran, 2017), because of its relevancy to statistical result presentation.

With the focus of statistical training on descriptive statistics in Point 3, school administrators might have difficulty generalizing their findings beyond a local context, which downplays the importance of statistical inference for result dissemination. In this regard, one may borrow arguments from qualitative studies to claim case similarities for result relevancy in other schools. But random fluctuations, measurement errors, and confounding factors often impact education data and undermine the similarity assertion for Point 5. As Norman (2017) pointed out, "Qualitative researchers in our midst might well be feeling a bit justifiably smug at this point. After all, it is an axiom of their discipline that observations don't generalize; every observation is so influenced by contextual details that replication is bound to fail" (p. 1052). One may wonder why a particular school report should even be read if it has nothing to do with others in different schools.

Furthermore, Point 5 stresses description of a univariate distribution as if there is no need for analyzing relations among multiple variables. Point 3 also delimits the focus of decision-making in the current academic year, which hinders expansion of the result interpretation for visionary leaders in the time dimension. Altogether, the seven guidelines might help simplify content for statistics instruction, but they are unlikely to support preparation of professional school leadership for well-rounded decision making at various times and/or in different settings.

Given the shortcomings of these guidelines, content reduction should not be aimlessly implemented in statistics courses. More consideration should be given to resolve the persistent issue that school leaders can not conduct effective data analyses after completing their programs in educational administration (Holcomb, 2004). Bradshaw and Phillips (2002) proposed adjustment, instead of reduction, of the course content in statistics. Polnick and Edmonson

(2005) examined the current course structure, and complained that too little "time is devoted to survey methods, estimation techniques, exploratory data analysis, and statistical graphs for reporting the findings of practical inquiries, which are the essential statistics and data-analysis skills principals need to be successful on the job" (p. 41). These discussions primarily focus on supplying data analysis tools in statistics classes.

From a demand perspective, the need for real-world data analyses is grounded on the structure of an education system in which classes are nested in schools and schools are nested in school districts. Consequently, quantitative results from the hierarchical system involves disaggregation of school data at multiple levels. In particular, Bernhardt (2013) describes four layers of data disaggregation:

Layer 1. How many students are there? Male vs. female/Limited English Proficiency (LEP) vs. non-LEP ethnicities/Lunch codes.

Layer 2. How have the demographics changed over time? Increases vs. decreases in categorical variables.

Layer 3. What percentage of students are gifted, and are they equally distributed among genders and ethnicities?

Layer 4. How has the enrollment of LEP students entering the building changed over the years? Do students with higher attendance get better grades?

While data from Layer 1 can be subjected to contingency table analyses of discrete variables, Layer 2 involves continuous variables, such as time, in statistical reporting. Hence, so-called "real-world needs" vary according to school administrators' responsibility at a particular layer. Similarly, education leaders at Layers 3 and 4 should be trained at a more advanced level because they are required to analyze data distributions in multiple dimensions (Layer 3) and/or model the stochastic process for multiple variables (i.e., LEP, attendance, and grades in Layer 4).

In addition, although gender, ethnicity, giftedness and LEP status can be classified as categorical variables on a nominal scale, lunch codes relate to family socioeconomic status that is typically represented on a ordinal scale. Student performance, as indicated by a grade-point average, could be on an interval scale. To project changes over future years and estimate relations among different variables, data analyses involve both parametric and non-parametric statistical methods. With proper approaches in the data gathering, school administrators are not only needed to describe the data features, but also required to estimate the variation of empirical findings in statistical inference.

Because not all schools are of the same size, the tasks of data analysis also depend on the environmental settings. In a small school, descriptive statistics could be used more often when the data are gathered across the entire population. In large school districts, school administrators might choose to draw a random sample to represent the entire population, and thus, inferential statistics should be used instead. Since school leaders may experience job transitions, learning different statistical methods is an effective way to strengthen their wellrounded leadership capacity to handle change. Alternatively, partial endorsement of narrowlyfocused statistics content might result in insufficient school administrator preparation for the real job market.

In summary, no piecemeal approach should be taken to fragmentize statistical training in education. Built on the axiom that the whole could be larger than the sum of its parts, it is more desirable to include both descriptive or inferential statistics across the parametric and nonparametric domains. The shared statistical training for teacher and principal preparation may offer additional opportunities to facilitate data triangulation from different perspectives. In a book "Real World Research," Robson (2002) considers data triangulation as an important strategy for strengthening report validity, credibility, and reliability.

Adaptations to Changes of Inferential Statistics from the Latest Subject Reform

Since the 1990s, school administration has been influenced by several federal education initiatives, such as Goals 2000, No Child Left Behind, and Race to the Top (Klein, 2018). In contrast, the content of statistics course has been relatively stable. Heston and King (2017) noted that "The meaning and use of statistical significance as originally defined by RA Fisher, Jerzy Neyman and Egon Pearson has undergone little change in the almost 100 years since originally proposed" (p. 113). For the findings of rejecting a null hypothesis, statistical significance is usually based on a p value less than 0.05, which traps the date-based reasoning into binary thinking. Kennerly (2016) complained, "It's true that researchers typically use statistical formulas to calculate a '95% confidence interval' — or, as they say in the jargon of statistics, '$p < 0.05$' — but this isn't really a scientifically-derived standard" (p. 1). In most textbooks, statistical inferences are still based on p values for probabilistic inference.

This stagnancy is about to change. On behalf of The American Statistician journal, Wasserstein and Lazar (2016) acknowledged that "Underpinning many published scientific conclusions is the concept of 'statistical significance,' typically assessed with an index called the p-value. While the p-value can be a useful statistical measure, it is commonly misused and misinterpreted" (p. 131). Consequently, Wasserstein, Schirm, and Lazar (2019) cautioned that the phrase "statistically significant" has become all but "meaningless." Their recommendation is to abandon its use entirely! Similar arguments and proposals have been made in the journal *Nature* to "retire statistical significance" (Amrhein, Greenland, & McShane, 2019). Furthermore, McShane, Gal, Gelman, Robert, and Tacket (2019) urged to drop the null hypothesis testing paradigm "for research, publication, and discovery in biomedical and social sciences" (p. 235).

These proposed changes may lead to exclusion of confidence interval estimation in statistics courses. If a value specified by the null hypothesis is outside a 95% confidence interval, then the null hypothesis is automatically rejected at a=0.05. In *The American Statistician*, Amrhein, Trafimow, and Greenland (2019) argued that words like "significance," "p-values" and "confidence" with interval estimates may mislead users into overconfident claims. Hence, the criticism has been extended to both point and interval estimates of inferential statistics.

With the desertion of statistical significance, Johnson (2019) worried that "abandoning evidence-driven standards for these judgments will make it even more difficult to design experiments, much less assess their outcomes" (p. 2). Ahuja (2019) also reported that John Ioannidis of Stanford University expressed reservation against abolishing statistical significance, and defended it as a "convenient obstacle to unfounded claims." Without it, he warned, "Irrefutable nonsense would rule" (Ahuja, 2019, p. 1). Deborah Mayo, a philosopher of science at Virginia Tech, further suggested that, “Nature ought to invite somebody to bring out the weakness and dangers of some of these recommendations” (Harris, 2019, p. 3). She cautioned that "Banning the word 'significance' may well free researchers from being held accountable when they downplay negative results" (Harris, 2019, p. 3) and “We should be very wary of giving up on something that allows us to hold researchers accountable” (see Harris, 2019, p. 4).

Goodman (2019) addressed the question “Why Is Getting Rid of P-Values So Hard?” based on the need of considerable social change in academic institutions to diminish the impact of statistical significance on journal publication, grant funding and faculty promotion. To smooth the process, some statisticians believe that p values should be allowed in quantitative research reports (McShane, Gal, Gelman, Robert, & Tackett, 2019). With the prominent voices in statistics rejecting the call to discontinue the term “statistical significance,” the ASA (2016) recommends that statistics should support “understanding of the phenomenon under study, interpretation of results in context [and] ... No single index should substitute for scientific reasoning” (p. 132).

This ASA-pushed change offers two historic opportunities for strengthening statistics courses in education programs. First, ASA (2016) agrees that “a p-value without context or other evidence provides limited information” (p. 132), and thus, revision of statistics courses can be proposed to include more real world examples for justification of results at the school level. Secondly, ASA urges thoughtful research that “looks ahead to prospective outcomes in the context of theory and previous research” (Wasserstein, Schirm & Lazar, 2019, p. 4), which allows collaboration between statistics and qualitative research courses to expand student competence in analyzing big data that embrace the mixture of both numeric and text information.

In summary, the wisdom of changing inferential statistics is still delimited to incorporating uncertainty, quantifying it, and discussing it in a research context. Prior to a complete settlement of the dust, the ASA editors provide a bullet point list of their five don’ts (Wasserstein, Schirm, & Lazar, 2019) that can be relevant to renovating statistical training in school administrator preparation:

- Don’t base your conclusions solely on whether an association or effect was found to be “statistically significant.”
- Don’t believe that an association or effect exists just because it was statistically significant.
- Don’t believe that an association or effect is absent just because it was not statistically significant.
- Don’t believe that your p-value gives the probability that chance alone produced the observed association or effect or the probability that your test hypothesis is true.
- Don’t conclude anything about scientific or practical importance based on statistical significance (or lack thereof).

It should be noted that none for the five don’ts were mentioned by Holcomb (2004) or McNamara and Thompson (1996) for improving statistical training for school administrator preparation in the past. Norman (2017) asserted that “As anyone who has engaged in the culture wars between qualitative and quantitative researchers will attest, the debate between the two groups are unlikely to resolve anytime soon" (p. 1053). However, the demand for description of theoretical context and previous research, as advocated by ASA (see Wasserstein, Schirm, & Lazar, 2019), has built a bridge for articulating qualitative studies.

Albert Einstein once wrote on a blackboard, "Not everything that counts can be counted, and not everything that can be counted counts" (see Baker & Doyle, 2010, p. 5). Thus, statistics courses ought to support both data analyses and text analytics.

In retrospect, dividing quantitative and qualitative research in education, has been evolving for decades (Datta, 1992), and so has been the need for school administrator preparation in statistics (see Holcomb, 2004; McNamara & Thompson, 1996).

References

American Statistical Association (2016). ASA statement on statistical significance and p values. *The American Statistician*, *70*, 131-133.

Amrhein, V., Trafimow, D., & Greenland, S. (2019). Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication. *The American Statistician*, *73*, 262-270.

Baker, J., & Doyle, C. (2010). Not everything that counts can be counted. Retrieved from https://quoteinvestigator.com/2010/05/26/everything-counts-einstein/

Barczak, G. (2015). Publishing qualitative versus quantitative research. *Journal of Product Innovation Management, 32*(5), 658–658. DOI:10.1111/jpim.12277

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*. *3*(30), 774. https://doi.org/10.21105/joss.00774

Bernhardt, V.L. (2013). *Data analysis for continuous school improvement (3rd ed.)*. New York, NY: Routledge.

Bradshaw, L. K., & Phillips, J. C. (2002). *Learning what they need to know: School leaders' lessons of experience*. In G. Perreault & F. Lunenburg (Eds.), The changing world of school administration (pp.182-198). Lanham, MD: Scarecrow Press.

Burke, M. (2019). *Growth data, used by several California districts, measures a student's improvement year over year*. Retrieved from https://edsource.org/2019/how-california -measures-academic-success-is-changing-at-some-of-the-states-largest-districts/62

Commission on Teacher Credentialing (2019). CalAPA: Overview of leadership cycles and rubrics. Retrieved from http://www.ctcexams.nesinc.com/Content/Docs/CalAPA_Assessment_Overview.pdf

Creighton, T. (2001). Data analysis in administrators' hands: An oxymoron. *School Administrator*, *58*(4), 52-57.

Datta, L. (1992, November). *Paradigm wars: Can we give peace a chance?* Paper presented at the annual meeting of the American Evaluation Association, Seattle, Washington.

Elsayir, H. (2015). Significance test in meta-analysis approach: A theoretical review. *American Journal of Theoretical and Applied Statistics***,** *4*(6), 630-639.

Gelo, O., Braakmann, D., & Benetka, G. (2008). Quantitative and qualitative research: Beyond the debate. *Integrative Psychological & Behavioral Science*, *42*(3), 266–290.

Gharehchopogh, F., & Khalifelu, Z. (2011). *Analysis and evaluation of unstructured data: text mining versus natural language processing*. Retrieved from https://ieeexplore.ieee.org/document/6111017

Heston, T., & King, J. (2017). Predictive power of statistical significance. *World Journal of Methodology*, *7*(4), 112-116.

Hiller, N. J., DeChurch, L. A., Murase, T., & Doty, D. 2011. Searching for outcomes of leadership: A 25-year review. *Journal of Management*, *37*, 1137-1177.

Holcomb, E. L. (2004). Getting excited about data: Combining people, passion, and proof to maximize student achievement. Thousand Oaks, CA: Corwin Press.

Hy, R., Waugh, W., & Nelson, P. (1987). The future public administrator and quantitative skills. *Public Administration Quarterly*, *11*(2), 134-149.

International Bureau of Education (2019). Intended curriculum. Retrieved from http://www.ibe.unesco.org/en/glossary-curriculum-terminology/i/intended-curriculum

Jayashankar, S., & Sridaran, R. (2017). Superlative model using word cloud for short answers evaluation in eLearning. *Education and Information Technologies*, *22*, 2383-2402.

Klein, A. (2018). Year One Check-In: What's Trump Done on K-12 Compared to Other Presidents? Retrieved from http://blogs.edweek.org/edweek/campaign-k-12/2018/01/how_much_has_trump_accomplished.html

Kugler, K., Reif, J., Kaschner, T., & Brodbeck, F. (2018). Gender differences in the initiation of negotiations: A meta-analysis. *Psychological Bulletin*, *144*(2), 198-222.

Lenhard, W. & Lenhard, A. (2016). Calculation of effect sizes. Retrieved from: https://www.psychometrica.de/effect_size.html. Dettelbach (Germany*): Psychometrica*. DOI: 10.13140/RG.2.2.17823.92329

Loecher, M., & Ropkins, K. (2012). R Google Maps and loa: Unleashing R graphics power on map titles. *Journal of Statistical Software*, *63*, 1–18. doi:10.18637/jss.v063.i04.

McNamara, J. F., & Thompson, D. P. (1996). Teaching statistics in principal preparation: Part one. *International Journal of Educational Reform*, *5*(3), 381-389.

NAEP Data Training Workshop (2019). Retrieved from http://naep-research.airprojects.org/Training/Index

Norman, G. (2017). Generalization and the qualitative–quantitative debate. *Advances in Health Sciences Education*, *22*(5), 1051–1055. DOI: 10.1007/s10459-017-9799-5

Polnick, B., & Edmonson, S. L. (2005). Expanding data analysis skills in educational leaders: Implications for preparation programs. *Scholar-Practitioner Quarterly*, **3***(*1), 39-52.

Robson, C. (2002). *Real world research : A resource for social scientists and practitioner researchers*. Oxford: Blackwell.

Stats (2019). *Statistics*. Retrieved from https://www.k-state.edu/stats/

Wang, J. (2020, April). *An examination of quality assurance impact on student satisfaction in online teaching*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA,

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA’s statement on p-values: Context, process, and purpose. *The American Statistician, 70*, 129-133. https://doi.org/10.1080/00031305.2016.1154108

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “p<0.05.*The American Statistician, 73*(Suppl.1), 1–19. https://doi.org/10.1080/00031305.2019.158391

Yoshikawa, H., Weisner, T., Kalil, A., & Way, N. (2013). Mixing qualitative and quantitative research in developmental science: Uses and methodological choices. *Qualitative Psychology*, *1*(S), 3–18.