Paper 8280-2016

Transforming Data to Information in Education: Stop with the "Point-and-Click!!!"

Sean W. Mulvenon, University of Arkansas

ABSTRACT

Educational systems at the district, state, and national levels all report possessing amazing student-level longitudinal data systems (LDS). Are the LDS systems improving educational outcomes for students? Are they guiding development of effective instructional practices? Are the standardized exams measuring student knowledge relative to the learning expectations? Many questions exist about the effective use of the LDS system and educational data, but data architecture and analytics (including the products developed by SAS® are not designed to answer any of these questions. However, the ability to develop more effective educational interfaces, improve use of data at the classroom level, and improve student outcomes might only be available through use of SAS. The purpose of this session is to demonstrate an integrated use of SAS tools to guide the transformation of data to analytics that improve educational outcomes for all students.

INTRODUCTION

A feature of this paper is to introduce the concepts and challenges in expanding academic expectations beyond the limited skills required to "point-and-click" when using statistical software packages. It is obvious that improving the interface options, such as "point-and-click" or Graphical User Interfaces (GUI), have improved the ease of use of these statistical packages. Further, the improved functionality has also dramatically improved the ease of use for trained analysts and statisticians. As a statistics professor, I have seen a tremendous increase in the complexity of the statistical models employed by graduate students and faculty due to these interface improvements. Concomitantly, I have seen more questioning of the internal and external validity of reported results and conclusions based on the use of these more complex analytic models. My singular goal in presenting this paper is to introduce a philosophy that ease of use of statistical packages doesn't automatically transfer to more effective or knowledgeable use of analytic methods. Are data being used more effectively? Or is just more data being analyzed? As an educator who is passionate about the importance of analytics, the use of statistical methods to improve educational outcomes for students, and a professor in educational statistics it is essential to our field to brainstorm how to balance ease of use with improved use of analytics.

A BACKGROUND PERPECTIVE OF THE CHALLENGES

As I shared in my opening paragraphs I am an advocate in the use of the advanced statistical procedures and the long overdue development of "point-and-click" and GUI features for statistical packages. However, in a society where we have deemed it necessary to put labels on hairdryers warning people not to use these items in the shower or bathtub, I wonder about the need for similar warning labels on statistical packages? Imagine a warning label for SAS® GLM output such as:

> "The means you are generating in Proc GLM will be different than those you have generated using the Proc Means step due to subjects having incomplete data"

It is not uncommon in social science research to have researchers report means from one procedure with the F-test and probability from use of a second procedure. This practice may result in a discussion of statistical results as significant, but with values provided for the means and standard deviations incongruent with the reported F-test and probability. I have witnessed this unfortunate event at numerous conferences and other situations such as faculty brown bag lunch presentations. Typically, the statistical procedure being employed is relatively simple, such as an analysis of variance, and the more statistical capable faculty recognize the incongruence in results and challenge the conclusions of the presenter(s). Are the statistics incorrect? Did the statistical package make a computational error? This type of error may seem minor, but it is systemic and growing as people are mass producing analytics (due to ease) without growing their knowledge and understanding of analytics.

An aspect I have always loathed, just personally loathed as a statistician, has been the constant rant against the use of SAS by graduate students and faculty members who use SPSS. The basis for these rants is typically due to a belief that SPSS is so much easier to use than SAS. Should SAS be easier to use than SPSS? Most of the people I have attempted to help, support, or direct in scholarship that use SPSS also have had limited understanding of statistics. But they could generate statistical output and other “stuff” more easily in SPSS. I have always considered “point-and-click” statistical platforms more accessible, but not necessarily more effective in completing effective analytics. Worse, easy-to-use platforms may actually be doing a disservice to both research and researchers using them by creating the illusion that effective analytics is based on ease-of-use of a software package. I developed my skills to be an effective SAS programmer by spending many hours in computer labs reading SAS manuals. I always valued SAS as a package for those who sought to do things well and accurately. In contrast, I have long referred to SPSS as the **S**tatistical **P**ackage for those who do **S**loppy **S**tatistics. My apologies if anyone may be offended by this characterization of SPSS users as clearly there are many competent users of SPSS. This characterization is a reflection of my experience with too many users of SPSS who have championed this software package as a solution (due to its ease-of-use) as a basis for dismissing the need for research methods and statistics courses.

SAS has become much easier to use in the last decade and is on a level similar to SPSS in ease-of-use, but should that be the standard? A Boeing 747 is such an amazing marvel of engineering that you may essentially enter the gate where the plane is parked, where you want to fly, and the gate where you wish to park upon landing and then push “go” and the plane will virtually do everything for you. However, the Federal Aviation Administration (FAA) will only allow a highly skilled and trained pilot fly a Boeing 747. Why? It is because there are 350 people on board the plane? Or because the FAA recognizes the complexities of flying a plane go beyond the equipment? Extend this thought of safety and complexity of piloting a Boeing 747 to analytics in education. Why is there a perception and active marketing strategies to create an illusion that “pointing-and-clicking” features transcend the need for essential training in analytics to effectively evaluate the educational progress of 350 students in a school system? Too often the rhetoric is “we don’t need to make teachers or administrators statisticians” while concurrently placing greater demands on educators to use data more effectively. To borrow and modify a term from mathematics, an easy to use statistical software platform is a necessary condition to improve use of data in education, but is it a sufficient condition? Is the desire to improve use of analytics and data in education due to issues of improper analytics or inappropriate interpretations and conclusions? In my experiences, the analysis of the data and procedures submitted in various software packages have been compiled correctly.

EXAMPLES OF THE GOOD, THE BAD, AND THE UGLY IN USE OF ANALYTICS IN EDUCATION

As a method of demonstrating my concerns, I have compiled a series of examples where I am confident all of the statistical analyses were completed correctly, but the interpretations presented have either suspect internal or external validity (or both). Before proceeding, I will mention my goal in sharing these examples/studies is to demonstrate the challenges associated with just being able to “point-and-click” in using statistical packages. Additionally, I have elected to omit the specific references associated with these examples, though it may be possible to identify the authors due to specifics I do provide. It is important to share examples representative of concerns I have expressed to this point, but I did not think it paramount to cite the source of the examples. If you require specifics, please contact me via my contact information provided with this paper. However, I was once asked if I had a source of “bad” examples on the use of statistics and I responded “yes, the library.”

School Performance Index. A study was completed in 2006 and advocated use of the Arkansas School Performance Index or SPI. The SPI evaluated approximately 1,100 schools in Arkansas then developed a rating system of their performance using a regression model. The results from the SPSS regression models were reported correctly based on the output. However, the interpretation of the analysis and internal validity of the study were very suspect. One school district was identified as the best performing in Arkansas, but this same school district had been identified by the Arkansas Department of Education (ADE) as in chronic academic distress with more than 75% of students performing in the “Below Basic” category in Reading. The single best predictor of performance in the SPI was the percentage of African-American students attending a school. The weighting of the SPI model was very problematic. For example, if you held constant the academic scores of one elementary school that was ranked 887, but changed the percentage of African-American students from 15% to 90% their ranking would improve to 23rd. The SPI was dismissed as an effective resource and never employed by the ADE, but is was disconcerting this method was developed by two academics using SPSS and “advanced statistics.”

Meaningfulness of results of an R^2 of .30 versus R^2 of .18. A graduate student was conducting a study on how to improve the prediction of Freshman GPA in college through the use of high school GPA and ACT scores. First, I googled “predicting college success using GPA and ACT” and received 141 million results, which suggests this research has been done. Second, in Chapter 2 or the Literature Review, the student identified several studies which had on average R^2 = .30 or 30% of the variance explained. In her study the reported R^2 was .18 or 18% of variance explained, which was also identified as statistically significant at the .01 level and therefore “proved” her use of high school GPA and ACT scores was a better method. Are the results from this study meaningful? The conclusion was this was a better method because it was "statistically significant" and "proved" her hypothesis. I pointed out current research explained 67% more variance than her model (i.e., 18% versus 30%) and inquired if there was anything unique to her study that made it more practical? Suffice it to say, this question was disregarded because the results were "statistically significant."

Over-parametized Factor Model. A graduate student used factor analysis to evaluate the Arkansas Writing Benchmark Framework and the reported results indicated only two constructs. However, until 2014 when the use of this exam was discontinued, professional development models in Arkansas consistently referred to the five constructs of the Arkansas Writing Benchmark Exams. The person who completed this factor analytic study was the Director of Assessment and Accountability in Arkansas for a number of years. I share this example, because this person was an intelligent and well -intended educator who to the best of my knowledge simply did not understand the significance of the results from her own dissertation. However, I am positive the analytics and use of the statistical package were correct. Table 1 provides a copy of the statistical output from this dissertation and the identification of two underlying factors, one associated with Content and Style with the second emphasizing Mechanics, Sentence Structure, and Usage. Two questions were asked for each student presented by the use of "1" and "2" to represent the Domain associated with each question. The reported results where there are five factors associated with the discrete Domains.

Table 1. *Factor Loadings for Factor Analysis of Writing Benchmark Scores*

	Loadings	
Domain	Factor 1	Factor 2
Content 1	**.81536**	.26799
Style 1	**.79690**	.29786
Sentence 1	.43000	**.64163**
Usage 1	.19233	**.70696**
Mechanics 1	.17654	**.72928**
Content 2	**.83819**	.23540
Style 2	**.82748**	.25873
Sentence 2	.45358	**.62554**
Usage 2	.22781	**.69663**
Mechanics 2	.21865	**.73750**

Note: A case for a unidimensional factor could also be made based on these results.

Suspect Statistics and Validity. A variation of a common saying that I deplore is "you can prove anything with statistics." This saying has a deleterious impact on the effective use of statistics and analytics because it creates an opportunity for people to randomly dismiss those statistical results where they have a disagreement. The model I am about to share is from an accountability system in Arkansas where again, the statistics and modeling are correct, but the interpretation and use of the results are very suspect. Figure 1 provides an overview of the results from analytics to assign accountability labels to high schools. The internal and external validity of the assigned accountability labels demonstrates an issue which is often overlooked in our use of analytics, which is the policy applications. It is clear in review of Figure 1 that Schools A and B are both performing at an academic level which exceeds the averages of the "High Achieving" High Schools. More specifically, "you can say anything with statistics" which is evident in educational policy when evaluating schools as part of "No Child Left Behind" and federal accountability programs. However, if you are astute and knowledgeable of statistics, despite the assigned labels in Figure 1 you may be able to interpret these results more objectively. Educational policy has a tremendous influence on interpretation of analytics and contributes to many challenges in effectively evaluating school performance (Mulvenon & Bowman, 2015).

High School(s) and Accountability Label	Literacy Proficient	Math Proficient	TAGG Literacy Proficient	TAGG Math Proficient	Graduation Rate	* ACT English	* ACT Math
School A (Needs Improvement School)	88.9%	92.8%	74.6	84.3%	85.2%	24.0	22.9
School B (Needs Improvement Focus School)	82.3%	83.5%	73.6%	73.6%	86.8%	24.3	23.4
Achieving High Schools (N = 15)	75.4%	72.0%	71.3%	69.3%	88.8%	20.6	19.7

Source: Arkansas Department of Education 2013-2014 School Data Reports. * Denotes 2012-2013 State Report Card Data (Note: weighted averages employed for ACT comparisons).

Figure 1. *Are High Schools A & B Really Less Effective than ALL the High Achieving Schools?*

A Regression Model Course for Educational Administrators? A common refrain from many educators or school administrators is the need to be able to predict student performance. Our Educational, Statistics, and Research Methods (ESRM) program has been asked on numerous occasions to provide a multiple regression course for the Curriculum Department and students completing advanced degrees in the various teacher or educational leadership programs. At one point, and you cannot make this up, ESRM was requested to provide a multiple regression that would be provided from 6 – 10 p.m. on Wednesday – Friday, then 9 a.m. – noon on Saturday mornings; and to repeat this process for three consecutive weeks. The group requesting this format claimed it would provide the students enrolled the 45 hours required for a 3-credit course in multiple regression. Natural follow-up questions by ESRM faculty included:

1) How would a student have the time to complete the necessary eight homework exercises that provide essential practice in our traditional 15 week semester course? **Answer.** Homework would not be necessary because students only need to know how to interpret results.
2) When would students complete the two midterm and final exams? **Answer.** You don't need to provide exams because students only need to know how to interpret results.
3) How would students have time to complete a research paper on an applied project that helps them demonstrate use and application of multiple regression? **Answer.** A research paper isn't required because students only need to know how to interpret results.

A theme emerged in the responses which raised questions regarding the real motives of the request to provide a multiple regression course in using this three-week format. Was the purpose to provide a new format to present a 15-week semester course in multiple regression or create the illusion of a 15-week course in a 3-week format? Regardless of the motives for this specific request, similar experiences have occurred with other academic programs, departments, and our Dean who has been very outspoken proclaiming all graduate students need to know is how to interpret statistics. In short, a growing perception of many in academe is that graduate students should *only* be required to know how to interpret results from statistical output. Which raises the question is the ability to interpret statistical results a sufficient requirement to effectively use analytics?

QUESTIONAL USE OF ANALYTICS IN PUBLISHED RESEARCH: ISOLATED OR SYSTEMIC?

A common exercise I use in my multiple regression and experimental design courses is to provide instruction on use of Interactive Matrix Language (IML) and basic programming code to allow students to replicate published results. Students are assigned the task of identifying a published manuscript in their field that reports means, standard deviations, and correlation tables of key variables. Use of this information and simple coding techniques provide the ability to replicate the analyses from these studies (see Figure 2). Originally I used this exercise as a precursor to demonstrating to students a method for completing pilot studies using Monte Carlo simulation techniques (Mulvenon & Turner, 1999). However, it became immediately clear that many students were having problems replicating reported results from published studies. My immediate inclination was there must be coding errors in the programs of the students, but in working to resolve these issues it was almost immediately clear the programs were correct and it was the reported results that were problematic. Further review also identified that in many cases when the statistical results were correct the reported conclusions may be problematic.

```
data z1(type= corr);
infile cards missover;
input _type_ $ _name_ $ x1 x2 x3 x4 x5 x6 x7 x8 x9;

/* Source: Viers, D. & Blieszner, R. (2004). Career and Relationship
Satisfaction Among Female Faculty in Marriage and Family Therapy Programs.
Contemporary Family Therapy, 26(4), December 2004 pg. 481-501 */

label
 x1= "Career Satisfaction"
 x2= "Relationship Satisfaction"
 x3= "Friendship Intimacy"
 x4= "Psychosocial Mentoring"
 x5= "Career Mentoring"
 x6= "Ethnicity"
 x7= "Received Award for Research"
 x8= "Received Award for Service"
 x9= "Number of Hours of Service"
;

cards;
mean . 3.41 18.49 131.08 3.99 3.42 .92 .30 .32 5.42
std . 0.41 2.71 21.02 0.88 1.11 0.28 0.46 0.47 4.43
n . 37 37 37 37 37 37 37 37 37
corr x1  1.000
corr x2 -0.031  1.000
corr x3  0.036 -0.048  1.000
corr x4  0.209 -0.047  0.273  1.000
corr x5 -0.117 -0.030  0.200  0.533  1.000
corr x6  0.253  0.054 -0.137 -0.190 -0.214 1.000
corr x7  0.253  0.191  0.054 -0.343 -0.121 0.193  1.000
corr x8  0.244 -0.018  0.054 -0.184  0.025 0.206  0.055  1.000
corr x9  0.512  0.143 -0.123  0.004 -0.049 0.096  0.114  0.251  1.000
;

run;

options ls= 121 ps= 40 pageno= 1 nodate;
title "Replication of Data that Doesn't Match Reported Results";

proc reg data= z1;
* model x1= x2 - x9 / stb scorr1 scorr2 vif;
 model x1= x2 - x9 / selection= backward scorr1;
 model x1 = x4 x9 x5 x7  / stb scorr1 scorr2 vif;
 mtest x4, x9, x5, x7;
 model x1 = x2-x5 / stb scorr2 vif;
 model x1 = x4 x7 x9 / stb scorr2 vif;
run;
```

```
quit;
```

Figure 2. *Sample SAS Code to Replicate Published Multiple Regression Results.*

IMPROVING EFFECTIVE USE OF ANALYTICS IN EDUCATION

The preceding examples have a common theme, correct analytics and incorrect or misguided interpretations. Which also raises the question of how do you identify if there is a problem with the analytics? More specifically, if the statistical package generates output it must be correct. Right? To improve the use analytics in education (and other fields) it is essential more effort is placed on generating and understanding the statistics. For example, if there is an error in the generation of statistical output how would you know if you possess limited understanding of statistics? What if there is a mistake in the data? A mistake in the model? Are there outliers? Are they influential outliers? What's an outlier? Which of the two regressions lines in Figure 3 are impacted by an influential outlier?

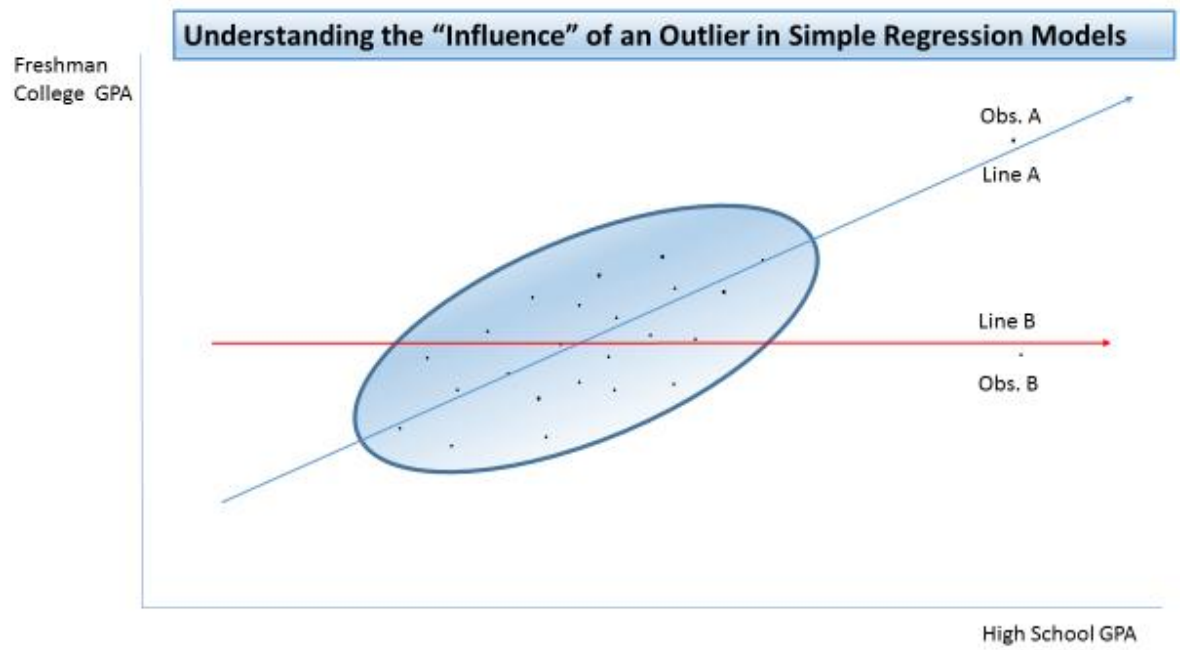

Figure 3. *Understanding Outliers in Regression.*

TOP OF THE PYRAMID

A common statement I share with graduate students and colleagues is the reliance on the "top of the pyramid" in referring to analytics. More specifically, the "top of the pyramid" refers to the F-tests and probabilities often treated as the *de facto* standard for evaluating the results from various analytic models. In reality, there are many other elements that must be considered and provide the ability to make much more accurate assessments or evaluations of statistical results. Figure 4 provides a graphical representation of this concept with numerous examples of other analytics required to effectively analyze data.

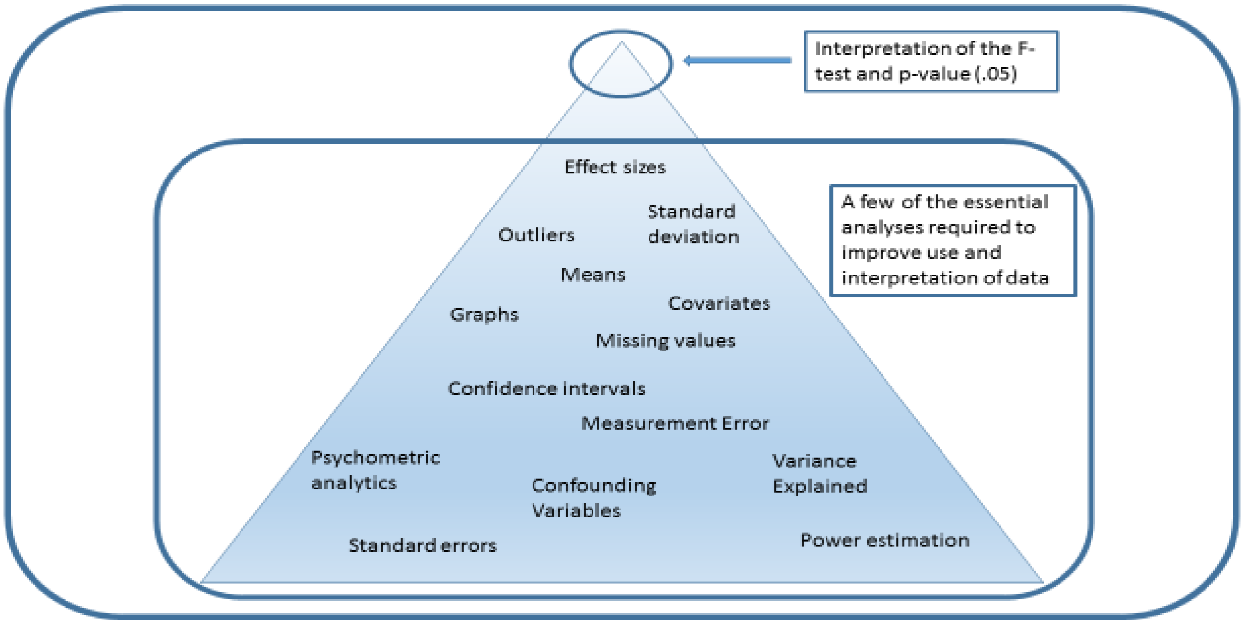

Figure 4. *Top of the Pyramid when interpreting statistics.*

Too often the "top of the pyramid" is all that is emphasized and use of data is reduced to being able to interpret these simplistic values. In reality, the use and interpretation of statistical data is based on additional methods listed in the pyramid. The additionally methods listed in Figure 4 represent just a few of the possible methods, but are shared as part of the pyramid to demonstrate that "effectively using data" requires more than the reporting of an F-test or P-value.

IMPROVED USE OF DATA IN EDUCATION AND SAS EDUCATORS CERTIFICATION

An element of my career, both as a professor and during a period as a senior advisor in Washington, D.C. is it is important to understand the challenges, but more importantly it is essential to have an idea on how to address and improve outcomes. A misunderstanding has emerged with the evolution of "point-and-click" and GUI interfaces for statistical platforms which is a belief that those features are of themselves all that is required to be an effective user of statistics. Too many times I have seen the presentations where "all the world's problems are solved" by using a statistic package that can disaggregate data, provides drop-down menus for analytics, can provide longitudinal charts of progress or is linked to the attendance system. A specific claim that may also be made is the statistical platform and analytics are "aligned with school curriculum."

However, a common belief among many educators is that even minimal expectations in research methodology or statistics are really a sinister and misguided effort to make them statisticians. I am always dismayed when I hear this belief expressed and typically I have two reactions: (1) this comment is a "safety term" which allows someone to disregard an academic expectation due to a personal proclamation (i.e., I'm a "visual learner" would be a similar type of term), and (2) the limited respect for the power of effective analytics while reducing all statistics to an F-test and probability value.

Warning labels and analytics. Imagine a world where if a teacher using SAS software was sued for inappropriate analysis of student achievement data? What if someone analyzing the effectiveness of a curriculum program erroneously concluded it was effective resulting in significant financial loss to the district? What are the consequences? Why do over the counter medicine products (e.g., Advil, Sudafed, etc.) have a recommendation on the label to "consult with your physician" prior to consuming the medication? Why are there labels on hairdryers warning against their use in the bathtub? It is simple, there is accountability and a direct cost for the results of ineffective use or application of medicine or hairdryers. The advertising of the "ease of use" of analytics products has no accountability. The examples using medicine and hairdryers may be extreme, but I have already shared numerous examples of so-called experts misusing statistic and analytics in education. The claims of "ease-of-use" and "aligned with your curriculum" should be balanced with an effort to encourage responsible and effective use of analytics in education.

SAS CERTIFICATION AND EDUCATORS

It is essential SAS provide leadership in the innovation and creation of programs to improve the effective use of analytics in education. Additionally, SAS should take a leadership role in advocating greater academic requirements and expectations for those engaged in use of analytics when evaluating student achievement. It is envisioned both elements could be accomplished with the creation of a SAS Certification for Educators.

SAS and Education. SAS has an amazing series of free resources available for educators, including SAS in the Cloud, Curriculum Pathways and On-line Support. The availability of these resources and SAS expertise in data architecture, point-and-click analytics and GUI interfaces provides the foundation for leading a renaissance in education and the use of analytics.

A SAS Certification for Educators would be consistent with other SAS certifications, but focused on essential elements for educators working in the Cloud and aligned with Curriculum Pathways resources, including:

a) Data architecture
b) "Bottom of the Pyramid" analytics to evaluate data structures (see Figure 2)
c) Analytics associated with education
    a. Growth models
    b. Hierarchical linear modeling
    c. Prediction models
    d. Student achievement and improvement models
d) Integration of educational policy and accountability models

A focused SAS Certification in Education designed to address components (a) – (d) would help direct and emphasize the need for expertise in analytics which transcends the current belief that ease-of-use is a necessary and sufficient condition.

A second element SAS could begin advocating is the role of analytics in K-12 preparation and graduate degree programs. The goal of research methods and statistics courses should be more than three credits associated with completion of a degree. The academic expectations in undergraduate and graduate level methodology and statistics courses should be more than a misguided notion that the ability to point-and-click demonstrates competency or the ability to effectively interpret results. More specifically, a "warning label" or concerted SAS marketing effort to advocate for academic expectations and rigor in programs that involve analyzing student level achievement data. Why should analytics be exempt from "warning labels" when so much is at risk in the evaluation of student achievement?

CONCLUSION

Are all teachers "effective" in the classroom? Does everyone have the ability to play golf as well as Tiger Woods? Does the ability to "point-and-click" make all people equally effective in the use analytics? For example, to be a more effective counselor it is required that student's complete internships to obtain essential experience. Pre-service teachers are expected to complete internships in the classroom with experienced teachers. Certified accountants are required to complete and pass an incredibly difficult exam to become "certified" in their field.

The representation "anyone" can be effective in their use of analytics because they can "point-and-click" may actually be marginalizing the dedicated efforts and skills sets of those who have invested the time to be "effective" in use of analytics. As a group, educators do not need to become "statisticians," but there should be more emphasis and respect for the dedication and investment of time required to be effective in use of analytics; and SAS could be leader in this effort.

The title "Transforming data to information in education" represents a relatively parsimonious goal. I am hopeful this brief narrative has helped share some of the many challenges associated with achieving this goal in education. It is essential people transition from a belief that possessing either (1) an ease-of-use interface or (2) the ability to interpret results as sufficient conditions to be effective in the use of analytics. It is essential that as professionals in statistics that we emphasize that effectively using data to improve educational outcomes of students requires a foundational understanding of basic research methods and statistics. Finally, that as a profession we value (and require) the investment of time by educators to become proficient in use of data by supporting their efforts with a "warning label" and programs (i.e., SAS Certification) in recognition of their efforts.

REFERENCES

Arkansas Department of Education Parent Report. (2014). http://www.arkansased.gov/im-looking-for/parents.

Mulvenon., S. & Bowman, S. (2015). An Evaluation of how the “Politics of K-12 Testing” Impact the Effectiveness of Global Testing Programs. Smith, W. (Editor) *The Global Testing Culture: Shaping Education Policy, Perceptions and Practice.* Oxford Press: Oxford, UK.

Mulvenon, S., & Turner, R. (1999). Using SAS® software to conduct pilot studies: An instructor’s guide. Manuscript published in the 24th Annual *SAS User Group International Conference Proceeding Guide,* pp. 1561-1566.

Viers, D. & Blieszner, R. (2004). Career and Relationship Satisfaction Among Female Faculty in Marriage and Family Therapy Programs. Contemporary Family Therapy, 26(4), December 2004 pg. 481-501.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Sean W. Mulvenon, Ph.D.
University of Arkansas
479-575- 8727
seanm@uark.edu