Improving Statistical Methods for Better Quantitative Research and Evaluation

ECU Five-Year Research Achievement Award Seminar

Guili Zhang, PhD
Research and Evaluation Methods
Department of Curriculum and Instruction
College of Education
East Carolina University
zhanggg@ecu.edu
My Research Focuses
My Research Focuses

Advancing

statistical & measurement methods
My Research Focuses

Advancing

statistical & measurement methods

Applying
Outline
Outline

- Advancing statistical and measurement methods.
Outline

- Advancing statistical and measurement methods.
- Applying statistical and measurement methods in research and evaluation (7 example studies)
Outline

- Advancing statistical and measurement methods.
- Applying statistical and measurement methods in research and evaluation (7 example studies)
- Grants
Outline

- Advancing statistical and measurement methods.
- Applying statistical and measurement methods in research and evaluation (7 example studies)
- Grants
- National leadership services
Outline

- Advancing statistical and measurement methods.
- Applying statistical and measurement methods in research and evaluation (7 example studies)
- Grants
- National leadership services
- National awards and honors
Improving Statistical & Measurement Methods
Improving Statistical & Measurement Methods

- *t*-test

  What is wrong and how to improve it?
Improving Statistical & Measurement Methods

- \( t \)-test
  What is wrong and how to improve it?

- \( F \)-test
  What is wrong and how to improve it?
Improving
Statistical & Measurement Methods

- $t$-test
  What is wrong and how to improve it?

- $F$-test
  What is wrong and how to improve it?

- Kenward-Roger $F$ test
  Does it work well?
Improving Statistical & Measurement Methods

- *t*-test
  What is wrong and how to improve it?
- *F*-test
  What is wrong and how to improve it?
- Kenward-Roger *F* test
  Does it work well?
- Cronbach’s *α* for Internal Consistency
  Can we do better?
Improving Statistical & Measurement Methods

- **t-test**
  What is wrong and how to improve it?

- **F-test**
  What is wrong and how to improve it?

- **Kenward-Roger F test**
  Does it work well?

- **Cronbach’s α for Internal Consistency**
  Can we do better?

- **Formative and Summative Evaluation**
  Is it systematic enough?
What is *t*-test?
What is *t*-test?

- The *t*-test assesses whether the means of two groups are different from each other.

- **Introduced in 1908** in *Biometrika* by William Gosset, a chemist working for the Guinness Brewery in Dublin, Ireland.

"Student" was Gosset’s pen name, that’s where the term “Student’s *t* Distribution” came from.

- Gosset devised the *t*-test as a cheap way to monitor the quality of stout.
What is Wrong with t-test?

\[ t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \]
What is Wrong with \( t \)-test?

- **It depends** (on sample size)!

\[
t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\]
What is Wrong with $t$-test?

- **It depends** (on sample size)!

- **It doesn’t tell us what we want to know.**
  - What we want to know: Given these data, what is the probability that the null hypothesis is true?
  - What it tells us: Given the null hypothesis is true, what is the probability of these (or more extreme) data?
What is Wrong with t-test?

- It depends (on sample size)!

- It doesn’t tell us what we want to know.
  - What we want to know: Given these data, what is the probability that the null hypothesis is true?
  - What it tells us: Given the null hypothesis is true, what is the probability of these (or more extreme) data?

- It tells us the direction, not the amount.
  - P-value does not tell you the amount.
  - The ritual dichotomous reject-accept decision is not the way any science is done. Physical scientists have learned much by storing up the amounts, not just directions. Measuring things on a communicable scale let us stockpile information about amounts.
What is Wrong with $t$-test?

- **It depends** (on sample size)!

- **It doesn’t tell us what we want to know.**
  - What we want to know: Given these data, what is the probability that the null hypothesis is true?
  - What it tells us: Given the null hypothesis is true, what is the probability of these (or more extreme) data?

- **It tells us the direction, not the amount.**
  - P-value does not tell you the amount.
  - The ritual dichotomous reject-accept decision is not the way any science is done. Physical scientists have learned much by storing up the amounts, not just directions. Measuring things on a communicable scale let us stockpile information about amounts.

- **The null hypothesis is always false** (given big enough sample size)!
What is Wrong with \( t \)-test?

- **It depends** (on sample size)!

- **It doesn’t tell us what we want to know.**
  - What we want to know: Given these data, what is the probability that the null hypothesis is true?
  - What it tells us: Given the null hypothesis is true, what is the probability of these (or more extreme) data?

- **It tells us the direction, not the amount.**
  - P-value does not tell you the amount.
  - The ritual dichotomous reject-accept decision is not the way any science is done. Physical scientists have learned much by storing up the amounts, not just directions. Measuring things on a communicable scale let us stockpile information about amounts.

- **The null hypothesis is always false** (given big enough sample size)!

- So, essentially, it’s testing whether there are a lot of subjects!
How to Improve $t$-test?

$$d = \frac{\bar{Y}_2 - \bar{Y}_1}{S}$$

$$\delta = \frac{\mu_1 - \mu_1}{\sigma}$$
How to Improve $t$-test?

- Use Effect Size (ES) (one of the most commonly used is Cohen’s $d$):

\[ d = \frac{\bar{Y}_2 - \bar{Y}_1}{S} \text{ estimates} \quad \delta = \frac{\mu_1 - \mu_1}{\sigma} \]

Sizes of the effect:
- **Small**: 0.2-0.3
- **Medium**: around 0.5
- **Large**: 0.8 to infinity
How to Improve $t$-test?

- Use **Effect Size (ES)** (one of the most commonly used is Cohen’s $d$):

  $$
  d = \frac{\bar{Y}_2 - \bar{Y}_1}{S} \quad \text{estimates} \quad \delta = \frac{\mu_1 - \mu_1}{\sigma}
  $$

Sizes of the effect:
- **Small**: 0.2-0.3
- **Medium**: around 0.5
- **Large**: 0.8 to infinity

Use **Confidence Interval (CI)** for Effect Size
How to Improve $t$-test? cont’d
How to Improve $t$-test? cont’d

Why Confidence Interval (CI) for Effect Size?
How to Improve $t$-test? cont’d

Why Confidence Interval (CI) for Effect Size?

- A CI contains all the information found in the significance tests.
How to Improve $t$-test? cont’d

Why Confidence Interval (CI) for Effect Size?

- A CI contains all the information found in the significance tests.

- A CI contains vital information not provided by the significance tests: **magnitude of effects and precision of estimates.**
How to Improve t-test? cont’d

Why Confidence Interval (CI) for Effect Size?

- A CI contains all the information found in the significance tests.

- A CI contains vital information not provided by the significance tests: magnitude of effects and precision of estimates.

- A CI indicates the range of population ESs with which the data are consistent. A significance test merely indicates whether the data are consistent with a population ES of zero.
How to Improve $t$-test? cont’d

What happens when data “misbehave”?
How to Improve $t$-test? cont’d

What happens when data “misbehave”?

Effect Size can become an inadequate measure of group separation.
How to Improve t-test? cont’d

What happens when data “misbehave”? 

Effect Size can become an inadequate measure of group separation. 
Then what do we do?
How to Improve $t$-test? cont’d

- **Use Robust Effect Size** (Algina et al., 2005)

  \[
  d_R = .642 \left( \frac{Y_{t2} - Y_{t1}}{S_W} \right) \quad \text{estimates} \quad \delta_R = .642 \left( \frac{\mu_{t2} - \mu_{t1}}{\sigma_W} \right)
  \]

  *The multiplier .642 is the Winsorized standard deviation of a standard normal distribution, and is used to ensure that when both samples being compared are drawn from normal distributions with equal variances.*

- **Use Confidence Intervals for Robust Effect Size**
  - Noncentral $t$-distribution based CI
  - Percentile bootstrap CI
What is $F$-test? What is Wrong with Significance Testing?
What is F-test? What is Wrong with Significance Testing?

- The F-test is one of the most commonly used significance tests for comparing three or more groups’ means. Developed by Sir Ronald Fisher in the 1920s as the variance ratio.

- Significance testing – the mechanical dichotomous decision around a sacred .05 criterion, has been severely criticized for the last five decades.

- A great deal of mischief has been associated with the test of significance (Bakan, 1966).

- Significance testing has not only failed to support the advance of social science as a science, but also has seriously impeded it (Cohen, 1994).
How to Improve $F$-test? Root Mean Square Standardized Effect Size (RMSSE)?

In a balanced, one-way, between-subjects, fixed-effects design, Root Mean Square Standardized Effect Size (RMSSE), denoted by $f^*$, is defined by Steiger & Fouladi (1997) as follows:

$$f^* = \sqrt{\frac{\sum_{j=1}^{J} (\mu_j - \mu)^2}{(J-1) \sigma^2}}$$

where $\mu_j$ is the mean for the $j$ th level, $\mu$ is the grand mean, and $\sigma^2$ is the within-level variance, which is assumed to be constant across levels.
**Estimate RMSSE**

Based on expected mean squares in a balanced design, \( \hat{f}^* \) can be estimated by using

\[
\hat{f}^* = \sqrt{\frac{MS_B - MS_W}{nMS_w}} = \sqrt{\frac{1}{n}(F-1)} \tag{A}
\]

if \( F \geq 1 \) and by using \( \hat{f}^* = 0 \), otherwise. Alternatively, based on the expected value of \( F \) under normality \( f^* \) can be estimated by using

\[
\hat{f}^* = \sqrt{\frac{(N-J-2)}{n(N-J)}(F-1)} \tag{B}
\]

if \( F \geq 1 \) and \( \hat{f}^* = 0 \) otherwise. Both estimates are very similar but the estimate in Equation A was used in our study because it does not require the normality assumption.
Noncentral $F$ Distribution-Based CI for RMSSE

The CIs for $f^*$ can be constructed based on the noncentral $F$ distribution. In a one-way, between-subjects, fixed-effects ANOVA, the $F$ statistic with $J - 1$ and $N - J$ degrees of freedom has noncentrality parameter

$$
\lambda = \frac{\sum_{j=1}^{J} n_j (\mu_j - \mu)^2}{\sigma^2}.
$$

Clearly in a balanced design

$$
f^* = \frac{\sqrt{\lambda}}{\sqrt{n(J - 1)}}.
$$

To find a 95% CI for $f^*$, we first use the noncentral $F$ distribution to find a 95% CI for $\lambda$. We then transform the endpoints of the CI for $\lambda$ by dividing $\lambda$ by $(J - 1)n$ and then take the square root. The result is an exact CI for $f^*$. 

How well does the CI for RMSSE Perform? The Investigation

- The noncentral $F$ distribution-based and the percentile bootstrap CIs were implemented for all combinations of the following five factors:
  - five population distributions including normal distribution and 4 nonnormal distributions from the family of the $g$ and $h$ distributions;
  - two numbers of levels for treatment groups: $J = 3$ and $J = 6$;
  - three cell sample sizes in each treatment: 20, 35, and 50;
  - six values of population RMSSEs: 0, .1, .25, .40, .55, and .70;
  - two mean configurations: the equally spaced mean configuration and the one extreme mean configuration.

- The nominal confidence level for all intervals investigated was .95 and each condition was replicated 2500 times.
- The number of bootstrap replications in the bootstrap procedure was 1000.
Coverage performance of CIs for RMSSE under Normality

- When sampling from a normal distribution, the coverage probability of the noncentral $F$ distribution-based CI should be .975 when $f^* = 0$, and the results in are consistent with the theory.

- When $f^* > 0$, the coverage probability of the noncentral $F$ distribution-based CI is expected to be .95 under normality and the results are consistent with this expectation.

- The percentile bootstrap CI also work adequately under normality.
What Happens when Data “Misbehave”?

- Zhang and Algina (2011) investigated the coverage performance of the noncentral $F$ distribution-based CI and the percentile bootstrap CI for RMSSE in a one-way, fixed-effects, between-subjects ANOVA.

We found that both the noncentral $F$ distribution-based CI and the percentile bootstrap CI for RMSSE yielded inadequate coverage probabilities under data nonnormality.

So what do we do now?
Development of the Robust Root Mean Square Standardized Effect Size (RMSSE$_R$) and its CIs

To overcome the weaknesses in $f^*$, we developed a robust version of the generalized effect size, the Robust Root Mean Square Standardized Effect Size (RMSSE$_R$), denoted by $f^*_R$ in our study.

In a balanced one-way between-subjects ANOVA design, $f^*_R$ is defined as

$$f^*_R = .642 \sqrt{\frac{\sum_{j=1}^{J} (\mu_{ij} - \mu_T)^2}{(J-1)\sigma_W^2}},$$

where $\mu_{ij}$ is the trimmed mean for the $j$th level, $\mu_T$ is the grand mean based on the trimmed means, and $\sigma_W^2$ is the within-level Winsorized variance, which is assumed to be constant across levels. The quantity .642 is the square root of the population Winsorized variance for a standard normal distribution. Therefore, including .642 in the definition of the robust effect ensures that $f^*_R = f^*$ when the data are drawn from normal distributions with equal variances.
Estimate $\text{RMSSE}_R$

An estimate of $f_R^*$ can be attained from sample statistics by applying the following formula:

$$
\hat{f}_R^* = 0.642 \times \sqrt{\frac{\sum_{j=1}^{J} (\bar{Y}_{ij} - \bar{Y}_T)^2}{(J-1)S_{wp}^2}},
$$

where $\bar{Y}_{ij}$ is the trimmed sample mean for the $j$th level, $\bar{Y}_T$ is the sample grand trimmed mean, and $S_{wp}^2$ is the sample pooled within-level Winsorized variance.
Noncentral $F$ distribution-based CI for $\text{RMSSE}_R$

In a balanced one-factor between-subject design with equal $n$s, $f^*_R$ can be written as a function of $\lambda_R$:

$$f^*_R = \sqrt{\frac{0.4129 \times \left( \sum_{j=1}^{J} n_j - J \right)}{\sum_{j=1}^{J} n_j - J \times (J - 1) n}} \cdot \lambda_R. \quad (9)$$

To find a $(1 - \alpha)\%$ (95% in this study) CI for $f^*_R$, we first use the noncentral $F$ distribution to find a 95% CI for $\lambda_R$. Once the CI on $\lambda_R$ is found, we then apply Equation 9 to transform the endpoints of the CI for $\lambda_R$ to obtain the endpoints for the CI for $f^*_R$. 
Percentile Bootstrap CI for RMSE<sub>R</sub>

To apply the percentile bootstrap method, the following steps are completed 1000 times within each replication of a condition.

1. A sample of size \( n_j \) is randomly selected with replacement from the scores for the group \( J \), \( j = 1, \ldots, J \). These \( J \) samples are combined to form a bootstrap sample.

2. The parameter \( \hat{f}^{*2} \) is estimated by using

\[
\hat{f}^{*2} = \frac{.642}{n} \sum_{j=1}^{J} \frac{n_j - 3}{\sum_{j=1}^{J} h_j - 3} (F_{R} - 1).
\] (15)

3. The 1000 \( f^{*2}_R \) estimates are ranked from low to high. The lower limit of the CI for \( f^{*2}_R \) is determined by finding the 26\(^{th}\) estimate in the rank order [i.e., the \((.025 \times 1000)+1\)^{th} estimate]; and the 975\(^{th}\) estimate is the upper limit of the CI for \( f^{*2}_R \) (i.e., the \((975 \times 1000)\)^{th} estimate).

4. The lower limit of the CI for \( f^*_R \) is equal to the square root of the lower limit of the CI for \( f^{*2}_R \) if the latter lower limit is larger than zero and is zero otherwise. The upper limit of the CI for \( f^*_R \) is equal to the square root of the upper limit of the CI for \( f^{*2}_R \).
Investigating the Performance of $\text{RMSSE}_R$

- The noncentral $F$ distribution-based and the percentile bootstrap CIs were implemented for all combinations of the following five factors:
  - five population distributions including the normal distribution and four nonnormal distributions from the family of the $g$ and $h$ distributions;
  - two numbers of levels for treatment groups: $J = 3$ and $J = 6$;
  - three cell sample sizes in each treatment: 20, 35, and 50;
  - six values of population $\text{RMSSE}_R$: 0, .1, .25, .40, .55, and .70;
  - two mean configurations: the equally spaced mean configuration and the one extreme mean configuration.

- The nominal confidence level for all intervals investigated was .95 and each condition was replicated 2500 times.

- The number of bootstrap replications in the bootstrap procedure was 1000.
Conclusions on $\text{RMSSE}_R$

- The noncentral $F$ distribution-based CIs for $f_R^*$, which was proposed in the current study and was formulated with the robust parameters including the trimmed means and Winsorized variances, yielded fairly adequate coverage probabilities and better coverage probability than the percentile bootstrap CI.

- Accordingly, researchers who want to set a CI for $f_R^*$ can use the CI constructed by using the noncentral $F$ distribution. These researchers will enjoy the additional benefit of using a robust measure of effect size, that is, a measure that is not likely to be strongly affected by outlying data points.
Conclusions on \( t \)-test, \( F \)-test, Effect Size, and Robust Effect Size
Conclusions on $t$-test, $F$-test, Effect Size, and Robust Effect Size

- $t$-test and $F$-test depend on sample size, and fail to provide the magnitude of effects and precision of estimation.
Conclusions on \( t \)-test, \( F \)-test, Effect Size, and Robust Effect Size

- \( t \)-test and \( F \)-test depend on sample size, and fail to provide the magnitude of effects and precision of estimation.

- CI for RMSSE in ANOVA provides the magnitude of group difference and precision of estimation. However, it yields inadequate coverage probabilities under data non-normality.
Conclusions on \( t \)-test, \( F \)-test, Effect Size, and Robust Effect Size

- \( t \)-test and \( F \)-test depend on sample size, and fail to provide the magnitude of effects and precision of estimation.

- CI for RMSSE in ANOVA provides the magnitude of group difference and precision of estimation. However, it yields inadequate coverage probabilities under data non-normality.

- The \( \text{RMSSE}_R \) and Noncentral F-distribution based CI for \( \text{RMSSE}_R \) I developed perform more satisfactorily under various conditions. Reporting it will be more accurate than the commonly used RMSSE.
Conclusions on $t$-test, $F$-test, Effect Size, and Robust Effect Size

- $t$-test and $F$-test depend on sample size, and fail to provide the magnitude of effects and precision of estimation.

- CI for RMSSE in ANOVA provides the magnitude of group difference and precision of estimation. However, it yields inadequate coverage probabilities under data non-normality.

- The $\text{RMSSE}_R$ and Noncentral F-distribution based CI for $\text{RMSSE}_R$ I developed perform more satisfactorily under various conditions. Reporting it will be more accurate than the commonly used RMSSE.

***This string of research is published in:
- Journal of Modern Applied Statistical Methods,
- Middle Grades Research Journal,
- Research Supporting Middle Grades Research (book).
Performance of Kenward-Roger $F$-test
Performance of Kenward-Roger $F$-test

- We investigated the Type I error rate of the Kenward-Roger (KR) $F$-test, using SAS PROC MIXED, through a simulation study for a between-by-within-subjects factor split-plot design with non-normal ignorable missing data MCAR (missing completely at random) and MAR (missing at random).
Performance of Kenward-Roger $F$-test

- We investigated the Type I error rate of the Kenward-Roger (KR) $F$-test, using SAS PROC MIXED, through a simulation study for a between- by within-subjects factor split-plot design with non-normal ignorable missing data MCAR (missing completely at random) and MAR (missing at random).

- The KR $F$-test for the between-subjects factor is robust, in terms of controlling the Type I error, to non-normal ignorable missing data.
Performance of Kenward-Roger $F$-test

- We investigated the Type I error rate of the Kenward-Roger (KR) $F$-test, using SAS PROC MIXED, through a simulation study for a between-by-within-subjects factor split-plot design with non-normal ignorable missing data MCAR (missing completely at random) and MAR (missing at random).

- The KR $F$-test for the between-subjects factor is robust, in terms of controlling the Type I error, to non-normal ignorable missing data.

- The KR $F$-test for the within-subjects factor is robust to non-normal data as long as the missing data mechanism is MCAR.
Performance of Kenward-Roger $F$-test

- We investigated the Type I error rate of the Kenward-Roger (KR) $F$-test, using SAS PROC MIXED, through a simulation study for a between- by within-subjects factor split-plot design with non-normal ignorable missing data MCAR (missing completely at random) and MAR (missing at random).

- The KR $F$-test for the between-subjects factor is robust, in terms of controlling the Type I error, to non-normal ignorable missing data.

- The KR $F$-test for the within-subjects factor is robust to non-normal data as long as the missing data mechanism is MCAR.

- The KR $F$-test for the within- by between-subjects interaction performed fairly well under MCAR. Unfortunately, it is not robust under MAR.
Performance of Kenward-Roger $F$-test

- We investigated the Type I error rate of the Kenward-Roger (KR) $F$-test, using SAS PROC MIXED, through a simulation study for a between-by within-subjects factor split-plot design with non-normal ignorable missing data MCAR (missing completely at random) and MAR (missing at random).

- The KR $F$-test for the between-subjects factor is robust, in terms of controlling the Type I error, to non-normal ignorable missing data.

- The KR $F$-test for the within-subjects factor is robust to non-normal data as long as the missing data mechanism is MCAR.

- The KR $F$-test for the within-by between-subjects interaction performed fairly well under MCAR. Unfortunately, it is not robust under MAR.

***This research is supported in part by NIH grant: R01AR052658-03S1
In collaboration with Padilla & Min.

Published in Journal of Modern Applied Statistical Methods
Cronbach’s $\alpha$

- A typical way of measuring a latent construct is through a scale or questionnaire containing items indirectly measuring the construct.

- It’s important that the items be consistent or reliable so that the questionnaire itself is consistent or reliable.

- Cronbach’s $\alpha$, aka coefficient alpha, was introduced by Lee Cronbach in 1951. It’s the most commonly used method of measuring reliability of a scale. It will generally increase as the intercorrelations among test items increase, and is thus known as an internal consistency estimate of reliability of test scores.

$$\hat{\alpha}_c = \frac{p}{p-1} \left( 1 - \frac{\sum_{i=1}^{p} \hat{\sigma}_{ii}}{\sum_{i=1}^{p} \sum_{j=1}^{p} \hat{\sigma}_{ij}} \right)$$ estimates $\alpha_c = \frac{p}{p-1} \left( 1 - \frac{\sum_{i=1}^{p} \sigma_{ii}}{\sum_{i=1}^{p} \sum_{j=1}^{p} \sigma_{ij}} \right)$
## Cronbach’s α for Internal Consistency

### Levels of Internal Consistency

<table>
<thead>
<tr>
<th>Cronbach's alpha</th>
<th>Internal consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha \geq .9$</td>
<td>Excellent</td>
</tr>
<tr>
<td>$.9 &gt; \alpha \geq .8$</td>
<td>Good</td>
</tr>
<tr>
<td>$.8 &gt; \alpha \geq .7$</td>
<td>Acceptable</td>
</tr>
<tr>
<td>$.7 &gt; \alpha \geq .6$</td>
<td>Questionable</td>
</tr>
<tr>
<td>$.6 &gt; \alpha \geq .5$</td>
<td>Poor</td>
</tr>
<tr>
<td>$.5 &gt; \alpha$</td>
<td>Unacceptable</td>
</tr>
</tbody>
</table>
Cronbach’s α for Reliability: can we do better?
Cronbach’s α for Reliability: can we do better?

- Reporting only a point estimate of coefficient alpha is no longer sufficient. Confidence Intervals for coefficient alpha were first introduced in 1960s and further developed.
Cronbach’s $\alpha$ for Reliability: can we do better?

- Reporting only a point estimate of coefficient alpha is no longer sufficient. Confidence Intervals for coefficient alpha were first introduced in 1960s and further developed.

However, all current coefficient alpha CIs are frequentist-based and thus have the traditional, less desirable CI interpretation: we are 95% confident that the interval captures the true population parameter.
Cronbach’s α for Reliability:
can we do better?

- Reporting only a point estimate of coefficient alpha is no longer sufficient. Confidence Intervals for coefficient alpha were first introduced in 1960s and further developed.

However, all current coefficient alpha CIs are frequentist-based and thus have the traditional, less desirable CI interpretation: we are 95% confident that the interval captures the true population parameter.

- Current coefficient alpha CIs cannot use prior information to stabilize inferences or update information.
Can we do Better?
Bayesian Coefficient Alpha

- Traditional frequentist analyses are only composed of data, **Bayesian** analysis is composed of data and prior knowledge and/or beliefs. Through the combination of data and prior knowledge, more can be learned about the phenomenon under study and knowledge can be updated accordingly.

- We developed a **Bayesian coefficient alpha** \(\alpha_b = E(\alpha_c | y)\)

where

\[
\alpha_c^{(t)} = \frac{p}{p-1} \left( 1 - \frac{\sum_{i=1}^{p} \sigma_{ii}^{(t)}}{\sum_{i=1}^{p} \sum_{j=1}^{p} \sigma_{ij}^{(t)}} \right)
\]

- Bayesian credible intervals can then be obtained by the lower \(\alpha/2\) and upper \(1-\alpha/2\) percentiles of the sample.
Investigating Performance of Bayesian Coefficient Alpha

- A $4 \times 3 \times 6$ Monte Carlo simulation design was utilized to investigate the properties of Bayesian coefficient alpha.
  - First, the number of items was investigated: 5, 10, 15 and 20.
  - Second, the mean item correlation was investigated: 0.173, 0.223, and 0.314. These mean items correlations were investigated because they generate coefficient alphas that range from 0.50 to 0.90.
  - Third, sample size was also explored: 50, 100, 150, 200, 250 and 300.
  - Multivariate normal data were generated with mean vector zero and correlation matrix $R$ of dimensions defined by the number of items in the simulation.
  - For each condition of the simulation study 1,000 replications were obtained. In each replication, Bayesian Coefficient Alpha was computed along with the SE and 95% BCIs.

- The results from our Monte Carlo investigations indicate that the Bayesian Coefficient Alpha was relatively unbiased under all investigated conditions.
Advantages of Bayesian Coefficient Alpha

- Bayesian Coefficient Alpha has the advantage of **having the credible intervals** (BCIs), which have the interpretation researchers really want to make with CIs: we are 95% confident that the true population parameter lies between the bands of the credible interval, a simpler and more powerful statement.
Advantages of Bayesian Coefficient Alpha

- Bayesian Coefficient Alpha has the advantage of **having the credible intervals** (BCIs), which have the interpretation researchers really want to make with CIs: we are 95% confident that the true population parameter lies between the bands of the credible interval, a simpler and more powerful statement.

- Bayesian Coefficient Alpha can **make use of prior information to stabilize inferences**. Through the combination of data and prior knowledge, more can be learned about the phenomenon under study and knowledge about the phenomenon can be updated accordingly.
Advantages of Bayesian Coefficient Alpha

- Bayesian Coefficient Alpha has the advantage of **having the credible intervals** (BCIs), which have the interpretation researchers really want to make with CIs: we are 95% confident that the true population parameter lies between the bands of the credible interval, a simpler and more powerful statement.

- Bayesian Coefficient Alpha can **make use of prior information to stabilize inferences**. Through the combination of data and prior knowledge, more can be learned about the phenomenon under study and knowledge about the phenomenon can be updated accordingly.

***Presented at the American Educational Research Association annual meeting
Published in Journal of Modern Applied Statistical Methods***
Formative & Summative Evaluation: Is it systematic enough?

- Michael Scriven coined the terms in 1967.

- Formative evaluation provides information needed for planning a treatment, while summative evaluation provides summary type of judgment regarding the effectiveness of the treatment.
The CIPP Model

- Created in the 1960s by Daniel Stufflebeam, commonly known as the founding father of the evaluation profession.

- The CIPP model includes Context, Input, Process, and Product evaluation.
The CIPP Model
as a Comprehensive Guiding Framework for Program Planning, Implementation and Evaluation
The CIPP Model
as a Comprehensive Guiding Framework for Program Planning, Implementation and Evaluation

- **Context Evaluation** is often referred to as needs assessment. It asks the question "What needs to be done?"
The CIPP Model
as a Comprehensive Guiding Framework for Program Planning, Implementation and Evaluation

- **Context Evaluation** is often referred to as needs assessment. It asks the question "What needs to be done?"

- **Input Evaluation** helps prescribe a project to address the identified needs. It asks the question "How should it be done?"
The CIPP Model
as a Comprehensive Guiding Framework for Program Planning, Implementation and Evaluation

- **Context Evaluation** is often referred to as needs assessment. It asks the question "What needs to be done?"

- **Input Evaluation** helps prescribe a project to address the identified needs. It asks the question "How should it be done?"

**Process Evaluation** monitors the project implementation process. It asks the question "Is it being done?"
The CIPP Model
as a Comprehensive Guiding Framework for Program Planning, Implementation and Evaluation

- **Context Evaluation** is often referred to as needs assessment. It asks the question "What needs to be done?"

- **Input Evaluation** helps prescribe a project to address the identified needs. It asks the question "How should it be done?"

- **Process Evaluation** monitors the project implementation process. It asks the question "Is it being done?"

- **Product Evaluation** identifies and assesses project outcomes. It asks the question "Did the project succeed?", and is similar to outcome evaluation.
The CIPP Model
as a Comprehensive Guiding Framework for Program Planning, Implementation and Evaluation

- **Context Evaluation** is often referred to as needs assessment. It asks the question "What needs to be done?"

- **Input Evaluation** helps prescribe a project to address the identified needs. It asks the question "How should it be done?"

- **Process Evaluation** monitors the project implementation process. It asks the question "Is it being done?"

- **Product Evaluation** identifies and assesses project outcomes. It asks the question "Did the project succeed?", and is similar to outcome evaluation.

***I am currently completing a book that provides a 21st century update of the CIPP Model with Dr. Stufflebeam.***
Applied Research Study Example #1
Identifying Impact Factors
Applied Research Study Example #1
Identifying Impact Factors

- Identifying factors influencing engineering student graduation and retention: A longitudinal and cross-institutional study.
  Published in *Journal of Engineering Education*.

Supported by a 10-year NSF grant (Grant No. EEC-9727411). NSF developed the Engineering Education Coalition (EEC) program to stimulate bold, innovative, and comprehensive models for systemic reform of undergraduate engineering education.

- A total of six EECs were established and funded for 10 years, involving 45 engineering institutions nationwide. The Southeastern University and College Coalition for Engineering Education (SUCCEED) is one of the six engineering Coalitions.
Applied Research Study Example #1
Identifying Impact Factors

As the statistical analyst & program evaluator of SUCCEED, I created the SUCCEED longitudinal database (LDB), the largest engineering database of its kind in the U.S. It contains all undergraduate students from 9 universities: Clemson, Florida A&M, FSU, Georgia Tech, NCAT, NCSU, UFL, UNCC, and Virginia Tech.
Applied Research Study Example #1
Identifying Impact Factors
Applied Research Study Example #1
Identifying Impact Factors

Characteristics of SUCCEED Schools

- Enroll over 28,000 engineering undergraduates
Applied Research Study Example #1
Identifying Impact Factors

Characteristics of SUCCEED Schools

• Enroll over 28,000 engineering undergraduates
• Award $\frac{1}{12}$ of all U.S. engineering degrees
Applied Research Study Example #1
Identifying Impact Factors

Characteristics of SUCCEED Schools

- Enroll over **28,000** engineering undergraduates
- Award **1/12** of all U.S. engineering degrees
- Award **1/4** of all U.S. engineering degrees awarded to Black students
Applied Research Study Example #1
Identifying Impact Factors

Characteristics of SUCCEED Schools

- Enroll over 28,000 engineering undergraduates
- Award 1/12 of all U.S. engineering degrees
- Award 1/4 of all U.S. engineering degrees awarded to Black students
- Award 1/10 of all U.S. engineering degrees awarded to women
Applied Research Study Example #1
Identifying Impact Factors

- The study evaluated pre-existing factors’ impacts on engineering student success.
- The data were all engineering students at 9 universities spanning 15 years (87,167 engineering students) extracted from the LDB.
- A multiple logistic regression model with step-wise selection procedure is fitted to each institution’s data to explore the relationship between graduation and the pre-existing factors. Multiple logistical regression techniques allow us to examine the effect of each predictor while controlling for the other variables.

Test Statistics used:
- Chi-square statistic
- Wald test and Wald confidence interval
- Odds ratio.
- Cochran-Mantel-Haenszel test for homogeneity of odds ratio.
- Coefficient of Determination
Applied Research Study Example #1
Identifying Impact Factors

Percentage of First Time in College (FTIC) Engineering Students Graduated by Years since Matriculation

Note: error bars show one standard deviation.
# Applied Research Study Example #1
## Identifying Impact Factors

Type III analysis of effects: Wald $\chi^2$ and p-value.

<table>
<thead>
<tr>
<th>GENDER</th>
<th>HSGPA</th>
<th>SATQ</th>
<th>SATV</th>
<th>ETHNIC</th>
<th>CITIZEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$</td>
<td>$\chi^2$</td>
<td>$\chi^2$</td>
<td>$\chi^2$</td>
<td>$\chi^2$</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(p-value)</td>
<td>(p-value)</td>
<td>(p-value)</td>
<td>(p-value)</td>
<td>(p-value)</td>
</tr>
<tr>
<td>A</td>
<td>4.287 (0.038)</td>
<td>Not tested*</td>
<td>Not tested*</td>
<td>Not tested*</td>
<td>13.884 (0.031)</td>
</tr>
<tr>
<td>B</td>
<td>---</td>
<td>66.445 (0.0001)</td>
<td>22.517 (0.0001)</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>C</td>
<td>---</td>
<td>41.097 (0.0001)</td>
<td>15.705 (0.0001)</td>
<td>4.041 (0.044)</td>
<td>16.659 (0.005)</td>
</tr>
<tr>
<td>D</td>
<td>---</td>
<td>600.93 (0.0001)</td>
<td>234.50 (0.0001)</td>
<td>39.664 (0.0001)</td>
<td>24.525 (0.0001)</td>
</tr>
<tr>
<td>E</td>
<td>16.591 (0.0001)</td>
<td>88.924 (0.0001)</td>
<td>17.052 (0.0001)</td>
<td>---</td>
<td>18.262 (0.0001)</td>
</tr>
<tr>
<td>F</td>
<td>11.365 (0.0007)</td>
<td>168.806 (0.0001)</td>
<td>74.329 (0.0001)</td>
<td>39.985 (0.0001)</td>
<td>17.278 (0.002)</td>
</tr>
<tr>
<td>G</td>
<td>26.090 (0.0001)</td>
<td>195.343 (0.0001)</td>
<td>68.543 (0.0001)</td>
<td>12.035 (0.0005)</td>
<td>11.944 (0.036)</td>
</tr>
<tr>
<td>H</td>
<td>---</td>
<td>Not tested*</td>
<td>19.635 (0.0001)</td>
<td>6.183 (0.0129)</td>
<td>---</td>
</tr>
<tr>
<td>I</td>
<td>0.0057 (0.940)</td>
<td>Not tested*</td>
<td>240.997 (0.0001)</td>
<td>2.207 (0.0073)</td>
<td>43.947 (0.0001)</td>
</tr>
</tbody>
</table>
**Applied Research Study Example #1**

**Identifying Impact Factors**

Odds Ratio Estimates and 95% Wald Confidence Intervals: Universities A to E

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GENDER (Female vs. Male)</strong></td>
<td>0.872 [0.766, 0.993]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1.507 [1.236, 1.837]</td>
</tr>
<tr>
<td><strong>SATQ</strong></td>
<td>Not tested*</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1.006 [1.004, 1.009]</td>
</tr>
<tr>
<td><strong>SATV</strong></td>
<td>Not tested*</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.998 [0.995, 1.000]</td>
</tr>
<tr>
<td><strong>ETHNIC (Overall)</strong></td>
<td>Not available**</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Not available**</td>
</tr>
<tr>
<td><strong>ETHNIC (Hispanic vs. Other)</strong></td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>2.744 [1.311, 5.743]</td>
<td></td>
</tr>
<tr>
<td><strong>CITIZEN (Citizen vs. NRAllen)</strong></td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td><strong>CITIZEN (NRAllen vs. ResAllen)</strong></td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1.827 [1.328, 2.513]</td>
<td></td>
</tr>
</tbody>
</table>

*Not tested because of missing field in database.*

**Significant variable overall, but odds ratio can only be computed between specific variable levels.**

--- Indicates variable was not found to be significant.
# Applied Research Study Example #1

## Identifying Impact Factors

### Odds Ratio Estimates and 95% Wald Confidence Intervals: Universities F to I

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GENDER</strong></td>
<td>0.835</td>
<td>0.554</td>
<td>0.554</td>
<td>0.554</td>
</tr>
<tr>
<td>(Female vs. Male)</td>
<td>[0.752, 0.927]</td>
<td>[0.442, 0.695]</td>
<td>0.554</td>
<td>0.554</td>
</tr>
<tr>
<td><strong>JHSGPA</strong></td>
<td>2.118</td>
<td>4.103</td>
<td>4.103</td>
<td>4.103</td>
</tr>
<tr>
<td></td>
<td>[1.892, 2.372]</td>
<td>[3.366, 5.001]</td>
<td>Not tested*</td>
<td>Not tested*</td>
</tr>
<tr>
<td><strong>SATQ</strong></td>
<td>1.003</td>
<td>1.006</td>
<td>1.003</td>
<td>1.003</td>
</tr>
<tr>
<td></td>
<td>[1.003, 1.004]</td>
<td>[1.005, 1.007]</td>
<td>[1.004, 1.011]</td>
<td>[1.006, 1.007]</td>
</tr>
<tr>
<td><strong>SATV</strong></td>
<td>0.998</td>
<td>0.998</td>
<td>0.997</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>[0.997, 0.998]</td>
<td>[0.996, 0.999]</td>
<td>[0.994, 0.999]</td>
<td>[0.998, 1.000]</td>
</tr>
<tr>
<td><strong>ETHNIC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Overall)</td>
<td>Not available**</td>
<td>Not available**</td>
<td>0.542</td>
<td>0.542</td>
</tr>
<tr>
<td><strong>ETHNIC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Black vs. Hispanic)</td>
<td>0.542</td>
<td>0.542</td>
<td>0.542</td>
<td>0.542</td>
</tr>
<tr>
<td></td>
<td>[0.344, 0.853]</td>
<td>[0.344, 0.853]</td>
<td>0.355</td>
<td>0.355</td>
</tr>
<tr>
<td></td>
<td>[0.228, 0.553]</td>
<td>[0.228, 0.553]</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CITIZEN</strong></td>
<td>1.975</td>
<td>1.975</td>
<td>1.975</td>
<td>1.975</td>
</tr>
<tr>
<td>(Citizen vs. NRAlien)</td>
<td>1.975</td>
<td>1.975</td>
<td>1.975</td>
<td>1.975</td>
</tr>
<tr>
<td></td>
<td>[1.124, 3.472]</td>
<td>[1.124, 3.472]</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CITIZEN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(NRAlien vs. ResAlien)</td>
<td></td>
<td></td>
<td>0.355</td>
<td>0.355</td>
</tr>
</tbody>
</table>

*Not tested because of missing field in database.

**Significant variable overall, but odds ratio can only be computed between specific variable levels.

--- Indicates variable was not found to be significant.
Applied Research Study Example #1
Identifying Impact Factors

Findings

- High school GPA and SAT Math scores were positively correlated with graduation rates for all universities.
Findings

- **High school GPA** and **SAT Math** scores were positively correlated with graduation rates for all universities.

- **SAT Verbal** scores correlated negatively with odds of graduation—for 7 out of 8 universities, making this observation even more notable.
Applied Research Study Example #1
Identifying Impact Factors

Findings

- **High school GPA** and **SAT Math** scores were positively correlated with graduation rates for all universities.

- **SAT Verbal** scores correlated negatively with odds of graduation—for 7 out of 8 universities, making this observation even more notable.

- While gender, ethnicity and citizenship also showed significant effects, these were not consistently positive or negative. In 3 universities, the graduation rate for males was higher than females, while in 1, it was higher for females.
Applied Research Study Example #1
Identifying Impact Factors

Findings

- **High school GPA** and **SAT Math** scores were positively correlated with graduation rates for all universities.

- **SAT Verbal** scores correlated negatively with odds of graduation—for 7 out of 8 universities, making this observation even more notable.

- While gender, ethnicity and citizenship also showed significant effects, these were **not** consistently positive or negative. In 3 universities, the graduation rate for males was higher than females, while in 1, it was higher for females.

- Ethnicity was significant in 7 universities.
Findings

- **High school GPA** and **SAT Math** scores were positively correlated with graduation rates for all universities.

- **SAT Verbal** scores correlated negatively with odds of graduation—for 7 out of 8 universities, making this observation even more notable.

- While gender, ethnicity and citizenship also showed significant effects, these were **not** consistently positive or negative. In 3 universities, the graduation rate for males was higher than females, while in 1, it was higher for females.

- Ethnicity was significant in 7 universities.

- In 2 universities, citizenship significantly affected graduation.
Applied Research Study Example #2
Students Leaving Engineering

Grade-Point Average, changes of major, and majors selected by students leaving engineering.

Received *Frontiers in Education (FIE)* **Benjamin J. Dasher Best Paper Award.**

A significant majority of students leave engineering with passing GPAs, and that the distribution is actually weighted toward higher GPAs.
Applied Research Study Example #2
Students Leaving Engineering
Where do students go when they leave engineering?

Destination major for students leaving engineering as related to their GPA when they change majors

- Business
- Physical Sciences
- Education / Social Sciences

Percent of students in GPA range changing into each major grouping

Cumulative GPA when leaving engineering
Applied Research Study Example #2
Students Leaving Engineering

Retention in first non-engineering major

% of students changing into major

Cumulative GPA when leaving engineering

- All majors
- Business
- Physical Sciences
- Education / Social Sciences
Applied Research Study Example #2
Students Leaving Engineering

Destination Major for Students leaving engineering for the Science

Destination major for students leaving engineering for the sciences as related to their GPA when they change majors.

- Natural Resources
- Computer Information Systems
- Biology
- Mathematics
- Physics

Percent of students in GPA range changing into each major

Cumulative GPA when leaving engineering

0 - 0.1
1 - 1.75
2 - 2.25
2.25 - 2.5
2.5 - 3
3 - 3.5
3.5 - 4
Applied Research Study Example #2
Students Leaving Engineering

What happens to students who leave engineering?

Graduation success of students changing majors out of engineering

- Graduated in first non-ENGR major
- Graduated after changing majors again
- Didn't graduate

% of all students leaving engineering

Cumulative GPA when leaving engineering

0 - 1
1 - 1.75
1.75 - 2
2 - 2.25
2.25 - 2.5
2.5 - 3
3 - 3.5
3.5 - 4
Applied Research Study Example #2
Students Leaving Engineering
Applied Research Study Example #2
Students Leaving Engineering

Findings and Discussion

- Students do not just leave engineering because they “flunk out” – in fact, many are doing quite well academically.
  - Why do they leave?
  - What can we do to help students choose the right major?
Applied Research Study Example #2
Students Leaving Engineering

Findings and Discussion

- Students do not just leave engineering because they “flunk out” – in fact, many are doing quite well academically.
  - Why do they leave?
  - What can we do to help students choose the right major?

- Students with lower GPAs were more likely to change into business majors; those with higher GPAs were more likely to change into science majors.
  - Does this indicate some stereotyped perceptions about different majors?
  - What can we do to make sure students get the facts, not the stereotypes?
Applied Research Study Example #2
Students Leaving Engineering

Findings and Discussion

- Students do not just leave engineering because they “flunk out” – in fact, many are doing quite well academically.
  - Why do they leave?
  - What can we do to help students choose the right major?
- Students with lower GPAs were more likely to change into business majors; those with higher GPAs were more likely to change into science majors.
  - Does this indicate some stereotyped perceptions about different majors?
  - What can we do to make sure students get the facts, not the stereotypes?
- The fraction of students changing into education and social science majors was 10-15% across all levels of GPA.
  - This may indicate that students change into these majors because of a “calling.”

Can we make sure students see the social relevance of engineering?
Applied Research Study Example #3
How do chemical engineering students differ

How do chemical engineering students differ from others? Received *American Society for Engineering Education Best Paper Award.*

Using the SUCCEED longitudinal database (LDB), undergraduate chemical engineering students were compared with other engineering and non-engineering students on demographics and academic performance measures.
Applied Research Study Example #3
How do chemical engineering students differ

Difference in gender among Chemical Engineering, Other Engineering, Science, and Non-science undergraduates

<table>
<thead>
<tr>
<th>GENDER</th>
<th>CHE (%)</th>
<th>OENG (%)</th>
<th>SCI (%)</th>
<th>NSCI (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>36.25</td>
<td>20.49</td>
<td>44.47</td>
<td>56.97</td>
</tr>
<tr>
<td>Male</td>
<td>63.75</td>
<td>79.51</td>
<td>55.53</td>
<td>43.03</td>
</tr>
</tbody>
</table>
# Applied Research Study Example #3

How do chemical engineering students differ

## Percentage of Student Flow between Engineering Subfields (n=22,610)

<table>
<thead>
<tr>
<th>BEGIN</th>
<th>Chemical</th>
<th>Civil</th>
<th>Computer</th>
<th>Electrical</th>
<th>Industrial</th>
<th>Mechanical/Aerospace</th>
<th>Other Engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical</td>
<td>59.8%</td>
<td>0.9%</td>
<td>0.2%</td>
<td>0.9%</td>
<td>0.5%</td>
<td>1.0%</td>
<td>1.5%</td>
</tr>
<tr>
<td>Civil</td>
<td>2.8%</td>
<td>72.2%</td>
<td>1.1%</td>
<td>2.6%</td>
<td>2.5%</td>
<td>4.1%</td>
<td>3.7%</td>
</tr>
<tr>
<td>Computer</td>
<td>0.3%</td>
<td>0.2%</td>
<td>70.1%</td>
<td>1.9%</td>
<td>0.2%</td>
<td>0.5%</td>
<td>0.6%</td>
</tr>
<tr>
<td>Electrical</td>
<td>1.5%</td>
<td>1.3%</td>
<td>10.3%</td>
<td>66.1%</td>
<td>1.1%</td>
<td>2.7%</td>
<td>1.7%</td>
</tr>
<tr>
<td>Industrial</td>
<td>4.8%</td>
<td>2.5%</td>
<td>2.9%</td>
<td>4.4%</td>
<td>77.7%</td>
<td>4.3%</td>
<td>2.9%</td>
</tr>
<tr>
<td>Mechanical/Aerospace</td>
<td>2.6%</td>
<td>2.9%</td>
<td>1.1%</td>
<td>3.7%</td>
<td>2.6%</td>
<td>63.5%</td>
<td>4.5%</td>
</tr>
<tr>
<td>Other</td>
<td>3.6%</td>
<td>1.9%</td>
<td>0.3%</td>
<td>2.1%</td>
<td>1.1%</td>
<td>3.3%</td>
<td>63.3%</td>
</tr>
<tr>
<td>Engineering</td>
<td>Total</td>
<td>75.4%</td>
<td>82.0%</td>
<td>86.1%</td>
<td>81.7%</td>
<td>85.6%</td>
<td>79.4%</td>
</tr>
<tr>
<td>Total</td>
<td>78.2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Applied Research Study Example #3
How do chemical engineering students differ

**Percentage of Engineering Student Flow to Non-engineering fields**

\( n=22,610 \)

<table>
<thead>
<tr>
<th>BEGIN GRAD</th>
<th>Chemical</th>
<th>Civil</th>
<th>Computer</th>
<th>Electrical</th>
<th>Industrial</th>
<th>Mechanical/Aerospace</th>
<th>Other Engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biology</td>
<td>3.8</td>
<td>0.8</td>
<td>0.7</td>
<td>1.1</td>
<td>0.6</td>
<td>1.1</td>
<td>1.6</td>
</tr>
<tr>
<td>Business</td>
<td>3.9</td>
<td>4.9</td>
<td>3.0</td>
<td>4.5</td>
<td>7.7</td>
<td>5.5</td>
<td>4.7</td>
</tr>
<tr>
<td>CIS</td>
<td>0.8</td>
<td>0.4</td>
<td>5.5</td>
<td>2.6</td>
<td>0.6</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>Physical Science</td>
<td>7.4</td>
<td>0.9</td>
<td>0.6</td>
<td>0.6</td>
<td>0.5</td>
<td>1.3</td>
<td>2.0</td>
</tr>
<tr>
<td>Social Science</td>
<td>1.6</td>
<td>1.9</td>
<td>1.5</td>
<td>1.8</td>
<td>1.1</td>
<td>2.7</td>
<td>1.7</td>
</tr>
<tr>
<td>Other Non-Engineering</td>
<td>7.1</td>
<td>9.1</td>
<td>2.6</td>
<td>7.6</td>
<td>4.0</td>
<td>8.7</td>
<td>10.5</td>
</tr>
<tr>
<td>Total</td>
<td>24.6</td>
<td>18.0</td>
<td>13.9</td>
<td>18.3</td>
<td>14.4</td>
<td>20.6</td>
<td>21.8</td>
</tr>
</tbody>
</table>
Applied Research Study Example #3
How do chemical engineering students differ
Applied Research Study Example #3
How do chemical engineering students differ

Analyses on the nine variables
Applied Research Study Example #3
How do chemical engineering students differ

Analyses on the nine variables

- **Multivariate Omnibus Test** was done first to determine whether there is an overall group effect. The null hypothesis is that the groups do not differ on any of the 9 variables. The test was done using SAS PROC GLM through a MANOVA with "group" as the between-subjects factor and the 9 academic characteristics as dependent variables. The test showed that the groups differ on at least one variable.
Applied Research Study Example #3
How do chemical engineering students differ

Analyses on the nine variables

- **Multivariate Omnibus Test** was done first to determine whether there is an overall group effect. The null hypothesis is that the groups do not differ on any of the 9 variables. The test was done using SAS PROC GLM through a MANOVA with “group” as the between-subjects factor and the 9 academic characteristics as dependent variables. The test showed that the groups differ on at least one variable.

- **ANOVA on Individual Variables.** We then wanted to find out which of the variables is significant. This was done with an ANOVA on the variable. For a given variable, the null hypothesis is that the groups do not differ on that particular variable.
Applied Research Study Example #3
How do chemical engineering students differ

Analyses on the nine variables

• **Multivariate Omnibus Test** was done first to determine whether there is an overall group effect. The null hypothesis is that the groups do not differ on any of the 9 variables. The test was done using SAS PROC GLM through a MANOVA with “group” as the between-subjects factor and the 9 academic characteristics as dependent variables. The test showed that the groups differ on at least one variable.

• **ANOVA on Individual Variables.** We then wanted to find out which of the variables is significant. This was done with an ANOVA on the variable. For a given variable, the null hypothesis is that the groups do not differ on that particular variable.

• **Pair-wise Comparison among Groups.** To determine the extent of the variation among the groups, each variable was subsequently tested using pair-wise comparison. The Shaffer-Holm procedure was used to control the family-wise error rate.
# Applied Research Study Example #3

How do chemical engineering students differ

Comparisons among Chemical Engineering, Other Engineering, Science and Non-science Majors

<table>
<thead>
<tr>
<th>Variable</th>
<th>Statistic</th>
<th>CHE</th>
<th>OENG</th>
<th>SCI</th>
<th>NSCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>*SAT Math Score</td>
<td>M</td>
<td>645.92</td>
<td>635.88</td>
<td>609.11</td>
<td>553.16</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>82.29</td>
<td>83.83</td>
<td>96.20</td>
<td>92.20</td>
</tr>
<tr>
<td>*SAT Verbal Score</td>
<td>M</td>
<td>533.96</td>
<td>517.34</td>
<td>519.66</td>
<td>490.06</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>86.73</td>
<td>87.62</td>
<td>95.66</td>
<td>87.45</td>
</tr>
<tr>
<td>**High School GPA</td>
<td>M</td>
<td>3.72</td>
<td>3.56</td>
<td>3.57</td>
<td>3.31</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.36</td>
<td>0.41</td>
<td>0.47</td>
<td>0.52</td>
</tr>
<tr>
<td>Time to Graduation</td>
<td>M</td>
<td>53.88</td>
<td>55.90</td>
<td>51.77</td>
<td>51.82</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>8.88</td>
<td>10.24</td>
<td>11.22</td>
<td>10.88</td>
</tr>
<tr>
<td>Cumulative GPA</td>
<td>M</td>
<td>3.17</td>
<td>2.98</td>
<td>3.04</td>
<td>2.97</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.49</td>
<td>.55</td>
<td>.68</td>
<td>.54</td>
</tr>
<tr>
<td>Number of Major Changes</td>
<td>M</td>
<td>.63</td>
<td>.95</td>
<td>.89</td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.76</td>
<td>.83</td>
<td>.92</td>
<td>1.03</td>
</tr>
<tr>
<td>Semesters to Graduation</td>
<td>M</td>
<td>13.04</td>
<td>13.87</td>
<td>11.69</td>
<td>10.94</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>4.07</td>
<td>4.35</td>
<td>3.62</td>
<td>2.62</td>
</tr>
<tr>
<td>Cumulative Semester</td>
<td>M</td>
<td>163.35</td>
<td>168.87</td>
<td>146.05</td>
<td>136.11</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>38.42</td>
<td>42.12</td>
<td>36.29</td>
<td>26.10</td>
</tr>
<tr>
<td>Average Hours per Semester</td>
<td>M</td>
<td>13.01</td>
<td>12.60</td>
<td>12.84</td>
<td>12.68</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>2.34</td>
<td>2.22</td>
<td>1.90</td>
<td>1.75</td>
</tr>
</tbody>
</table>
Chemical engineering students (CHE) majors had significantly better SAT math scores than other engineering majors (OENG), science majors (SCI) and non-science majors (NSCI).
Applied Research Study Example #3

How do chemical engineering students differ

Findings

- Chemical engineering students (CHE) majors had significantly better SAT math scores than other engineering majors (OENG), science majors (SCI) and non-science majors (NSCI).
Applied Research Study Example #3
How do chemical engineering students differ

Findings

- Chemical engineering students (CHE) majors had significantly **better SAT math** scores than other engineering majors (OENG), science majors (SCI) and non-science majors (NSCI).

CHE majors had much **better SAT verbal** scores than NSCI majors, as well as OENG and SCI majors, although the scores were closer.
Chemical engineering students (CHE) majors had significantly better SAT math scores than other engineering majors (OENG), science majors (SCI) and non-science majors (NSCI).

CHE majors had much better SAT verbal scores than NSCI majors, as well as OENG and SCI majors, although the scores were closer.

CHE majors came in with better high school GPAs than all other groups.
Applied Research Study Example #3
How do chemical engineering students differ

Findings

- Chemical engineering students (CHE) majors had significantly better SAT math scores than other engineering majors (OENG), science majors (SCI) and non-science majors (NSCI).

CHE majors had much better SAT verbal scores than NSCI majors, as well as OENG and SCI majors, although the scores were closer.

- CHE majors came in with better high school GPAs than all other groups.

- CHE students took significantly more time to graduate than both SCI and NSCI groups, but required less time to graduate than OENG students.
Chemical engineering students (CHE) majors had significantly **better SAT math** scores than other engineering majors (OENG), science majors (SCI) and non-science majors (NSCI).

CHE majors had much **better SAT verbal** scores than NSCI majors, as well as OENG and SCI majors, although the scores were closer.

- CHE majors came in with **better high school GPAs** than all other groups.
- CHE students took significantly more **time to graduate** than both SCI and NSCI groups, but required less time to graduate than OENG students.
- CHE majors have significantly **higher cumulative GPAs** than others.
Applied Research Study Example #3
How do chemical engineering students differ

- CHE majors changed major significantly fewer times than other majors.
Applied Research Study Example #3
How do chemical engineering students differ

- CHE majors changed major significantly fewer times than other majors.
- CHE students took an average of two more semesters to graduate than their SCI and NSCI counterparts, but required about a semester less time to graduate than OENG students.
Applied Research Study Example #3
How do chemical engineering students differ

- CHE majors changed major significantly fewer times than other majors.

- CHE students took an average of two more semesters to graduate than their SCI and NSCI counterparts, but required about a semester less time to graduate than OENG students.

- CHE students took significantly more semester hours to graduate than both SCI and NSCI students, but required less semester hours to graduate than OENG students.
CHE majors changed major significantly fewer times than other majors.

CHE students took an average of two more semesters to graduate than their SCI and NSCI counterparts, but required about a semester less time to graduate than OENG students.

CHE students took significantly more semester hours to graduate than both SCI and NSCI students, but required less semester hours to graduate than OENG students.

CHE students took slightly fewer hours each semester than OSCI, SCI and NSCI students.
Applied Research Study Example #4
Survival Analysis of Engineering Attrition
Applied Research Study Example #4
Survival Analysis of Engineering Attrition

Nonparametric Survival Analysis of the loss rate of undergraduate engineering students.
Supported by NSF’s MIDFIELD grant (Multiple-Institution Database For Investigating Engineering Longitudinal Development).
Published in Journal of Engineering Education.
Applied Research Study Example #4
Survival Analysis of Engineering Attrition

Nonparametric Survival Analysis of the loss rate of undergraduate engineering students.
Supported by NSF’s MIDFIELD grant (Multiple-Institution Database For Investigating Engineering Longitudinal Development).
Published in Journal of Engineering Education.

Background:
Student dropout from colleges and universities has long been a concern and focus for educators.
Nonparametric Survival Analysis of the loss rate of undergraduate engineering students.

Supported by NSF’s MIDFIELD grant (Multiple-Institution Database For Investigating Engineering Longitudinal Development).

Published in *Journal of Engineering Education*.

**Background:**

Student dropout from colleges and universities has long been a concern and focus for educators.

- The snapshot approach taken by the existing studies does not offer the means to fully understand the time-dependent dropout issue.
Nonparametric Survival Analysis of the loss rate of undergraduate engineering students.
Supported by NSF’s MIDFIELD grant (Multiple-Institution Database For Investigating Engineering Longitudinal Development).
Published in Journal of Engineering Education.

Background:
Student dropout from colleges and universities has long been a concern and focus for educators.

- The snapshot approach taken by the existing studies does not offer the means to fully understand the time-dependent dropout issue.
- In this study, using a longitudinal database of 100,179 engineering students from 9 universities and spanning 19 years, undergraduate student dropout by cohort group, gender, ethnicity, and SAT verbal and math scores were investigated by using nonparametric survival analysis.
Applied Research Study Example #4
Survival Analysis of Engineering Attrition
Applied Research Study Example #4
Survival Analysis of Engineering Attrition

Key Research Questions

- Does the risk of leaving engineering differ among groups with different background (i.e., cohort, gender, ethnicity, SAT math score, and SAT verbal score)?

  When are students most likely to leave engineering?

- Is SAT score a good predictor of the risk of dropping out?
Applied Research Study Example #4
Survival Analysis of Engineering Attrition

• We introduced survival analysis to engineering student dropout research. Survival analysis allows:
  ✓ modeling patterns of occurrence,
  ✓ comparing these patterns among groups, and
  ✓ building statistical models of the risk of occurrence over time.

• SAS software’s PROC LIFETEST function was used to perform survival analysis.

• The data for this study were extracted from the MIDFIELD database that we developed.
## Applied Research Study Example #4
Survival Analysis of Engineering Attrition

### Table 1. The frequency of gender by cohort groups

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>4,096</td>
<td>4,697</td>
<td>4,813</td>
<td>5,309</td>
<td>1,867</td>
<td>20,782</td>
</tr>
<tr>
<td>Male</td>
<td>16,756</td>
<td>17,541</td>
<td>16,937</td>
<td>20,088</td>
<td>8,075</td>
<td>79,397</td>
</tr>
<tr>
<td>Total</td>
<td>20,852</td>
<td>22,238</td>
<td>21,750</td>
<td>25,397</td>
<td>9,942</td>
<td>100,179</td>
</tr>
</tbody>
</table>

### Table 2. The frequency of gender by ethnicity

<table>
<thead>
<tr>
<th>MIDFIELD</th>
<th>Ethnicity</th>
<th>White</th>
<th>Asian</th>
<th>Minority</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>14,144</td>
<td>1,478</td>
<td>4,796</td>
<td>364</td>
<td>20,782</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td>61,435</td>
<td>5,404</td>
<td>10,445</td>
<td>2,113</td>
<td>79,397</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>75,579</td>
<td>6,882</td>
<td>15,241</td>
<td>2,477</td>
<td>100,179</td>
</tr>
</tbody>
</table>
Applied Research Study Example #4
Survival Analysis of Engineering Attrition

Table 3. The frequency of SAT math score groups

<table>
<thead>
<tr>
<th>SAT math score groups</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>SATM &lt; 500</td>
<td>5,430</td>
<td>5.42</td>
<td>5,430</td>
<td>5.42</td>
</tr>
<tr>
<td>500 ≤ SATM &lt; 550</td>
<td>8,854</td>
<td>8.84</td>
<td>14,284</td>
<td>14.26</td>
</tr>
<tr>
<td>550 ≤ SATM &lt; 600</td>
<td>15,826</td>
<td>15.80</td>
<td>30,110</td>
<td>30.06</td>
</tr>
<tr>
<td>600 ≤ SATM &lt; 650</td>
<td>22,987</td>
<td>22.95</td>
<td>53,097</td>
<td>53.00</td>
</tr>
<tr>
<td>650 ≤ SATM &lt; 700</td>
<td>24,036</td>
<td>23.99</td>
<td>77,133</td>
<td>77.00</td>
</tr>
<tr>
<td>700 ≤ SATM &lt; 750</td>
<td>15,510</td>
<td>15.48</td>
<td>92,643</td>
<td>92.48</td>
</tr>
<tr>
<td>750 ≤ SATM ≤ 800</td>
<td>7,536</td>
<td>7.52</td>
<td>100,179</td>
<td>100.00</td>
</tr>
</tbody>
</table>
Results

Figure 1. Survival function for first-time-in-college students matriculating in engineering.

Life-Table Survival Function Estimate

<table>
<thead>
<tr>
<th>semester</th>
<th>Survival Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2.5</td>
<td>0.8</td>
</tr>
<tr>
<td>5</td>
<td>0.6</td>
</tr>
<tr>
<td>7.5</td>
<td>0.4</td>
</tr>
<tr>
<td>10</td>
<td>0.2</td>
</tr>
<tr>
<td>12.5</td>
<td>0.0</td>
</tr>
</tbody>
</table>

No. of Subjects | Event | Censored
----------------|-------|-----------
100179          | 22% (21943) | 78% (78236)
Results

*Figure 2.* Hazard function for first-time-in-college students matriculating in engineering.
Results

Figure 3. Survival functions by cohort groups
Results

*Figure 2.* Hazard function for first-time-in-college students matriculating in engineering.
Results

Figure 3. Survival functions by cohort groups
Results

Figure 4. Hazard functions by cohort groups
Results

*Figure 5. Survival functions by gender*

![Life-Table Survival Function Estimates](image)

Logrank $p < 0.0001$

<table>
<thead>
<tr>
<th>Semester</th>
<th>Female</th>
<th>Event</th>
<th>Censored</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20782</td>
<td>26%</td>
<td>74%</td>
</tr>
<tr>
<td></td>
<td>79397</td>
<td>21%</td>
<td>79%</td>
</tr>
</tbody>
</table>
Results

*Figure 6. Hazard functions by gender*
Results

Figure 7. Survival functions by ethnicity
Results

Figure 8. Hazard functions by ethnicity
Results

Figure 9. Hazard functions by SAT math score groups
Results

**Figure 9.** Hazard functions by SAT Verbal score groups
Results

Figure 9. Hazard functions by SAT math score groups

![Hazard Function Estimates](image)

- 200 <= SATM < 500
- 500 <= SATM < 550
- 550 <= SATM < 600
- 600 <= SATM < 650
- 650 <= SATM < 700
- 700 <= SATM < 750
- 750 <= SATM <= 800

Hazard Rate vs. semester (0 to 12 semesters)
Results

Figure 9. Hazard functions by SAT Verbal score groups
Summary of Key Results
Summary of Key Results

• There are no significant differences among cohort subgroups for long survival times, but there are significant differences between cohort subgroups for early survival times, and for gender, ethnicity group, and SAT math and verbal scores subgroups.
Summary of Key Results

- There are no significant differences among cohort subgroups for long survival times, but there are significant differences between cohort subgroups for early survival times, and for gender, ethnicity group, and SAT math and verbal scores subgroups.

- Females are significantly more likely to be at risk of leaving engineering in semesters 3 to 5 than males, while the risks are similar during other semesters.
Summary of Key Results

- There are no significant differences among cohort subgroups for long survival times, but there are significant differences between cohort subgroups for early survival times, and for gender, ethnicity group, and SAT math and verbal scores subgroups.

- Females are significantly more likely to be at risk of leaving engineering in semesters 3 to 5 than males, while the risks are similar during other semesters.

- White students tend to leave engineering slightly more than minority students, who leave engineering more than Asians, and Asian students leave engineering more than other students.
Summary of Key Results

• Generally, engineering college students leave engineering during the 3rd semester rather than other semesters, which may be due to probationary periods offered in earlier periods. Students with an SAT math score < 550, however, tend to leave engineering during the 2nd semester.
Summary of Key Results

• Generally, engineering college students leave engineering during the 3rd semester rather than other semesters, which may be due to probationary periods offered in earlier periods. Students with an SAT math score < 550, however, tend to leave engineering during the 2nd semester.

• SAT math score better predicts the risk of failure than SAT verbal score, and, not surprisingly, the lower a student’s SAT math score, the more likely that student is to leave engineering.
Summary of Key Results

• Generally, engineering college students leave engineering during the 3rd semester rather than other semesters, which may be due to probationary periods offered in earlier periods. Students with an SAT math score < 550, however, tend to leave engineering during the 2nd semester.

• SAT math score better predicts the risk of failure than SAT verbal score, and, not surprisingly, the lower a student’s SAT math score, the more likely that student is to leave engineering.

• Engineering college students with SAT verbal score < 500 tend to survive more than the students whose SAT verbal is between 500 and 600.
Significance of this Research

- Survival analysis is superior in analyzing time-to-event or longitudinal data because it exceptionally affords important information on variation in dropout risk at various times, and permits the effects of predictors to fluctuate over time.

- This research can stimulate more advanced research on undergraduate student dropout and provide important knowledge as basis for better dropout prevention.
Applied Research Study Example #5
Using CIPP Model to Guide Service Learning Project

  (in collaboration with Zeller, Griffith, Metcalf, Williams, Shea, & Misulis)
Published in Journal of Higher Education Outreach and Engagement.

What is Service Learning?
Service-learning involves the integration of community service into the academic curriculum. Students learn while providing services that meet a community’s needs.
Successful service-learning requires researchers to:

1. **Identify** the learning needs of service providers and the needs of community partners,

2. **Design** a program that can effectively address both needs, and

3. Successfully **implement** the programs to **generate** desired outcomes to meet both needs.
Using the CIPP Model to Guide Service-learning Program

• In spring 2008, a service-learning tutoring program was initiated to address the identified learning needs of the preservice teachers in the Elementary Education program at ECU and those of at-risk readers in the local school system.

• 26 preservice teachers taking a course in Diagnostic/Prescriptive Teaching of Reading completed a service-learning component by tutoring 26 Response-to-Intervention students (RTI at-risk readers) in kindergarten, first, and second grades at a local elementary school.

• The CIPP model was used to systematically guide the conception, design, implementation, and assessment of this tutoring program and led to its successful completion and attainment of desired outcomes. The four CIPP components were methodically conducted using a variety of assessment techniques.
Use the CIPP Model to Guide Service-learning Project
Use the CIPP Model to Guide Service-learning Project

- First, Context Evaluation identified the needs of preservice teachers to gain firsthand experience of working with students from diverse background and the elementary school RTI at-risk readers’ needs for individualized assistance on reading.
Use the CIPP Model to Guide Service-learning Project

• First, **Context Evaluation identified the needs** of pre-service teachers to gain firsthand experience of working with students from diverse background and the elementary school RTI at-risk readers’ needs for individualized assistance on reading.

• Next, **Input Evaluation incorporated input** from experts, practitioners, and various stakeholders, and prescribed an effective tutoring project.
Use the CIPP Model to Guide Service-learning Project

• First, Context Evaluation identified the needs of pre-service teachers to gain firsthand experience of working with students from diverse background and the elementary school RTI at-risk readers’ needs for individualized assistance on reading.

• Next, Input Evaluation incorporated input from experts, practitioners, and various stakeholders, and prescribed an effective tutoring project.

• Then Process Evaluation helped monitor the project implementation process and provided ongoing feedbacks for needed adjustments to the project.
Use the CIPP Model to Guide Service-learning Project

• First, Context Evaluation identified the needs of pre-service teachers to gain firsthand experience of working with students from diverse background and the elementary school RTI at-risk readers’ needs for individualized assistance on reading.

• Next, Input Evaluation incorporated input from experts, practitioners, and various stakeholders, and prescribed an effective tutoring project.

• Then Process Evaluation helped monitor the project implementation process and provided ongoing feedbacks for needed adjustments to the project.

• Finally, Product Evaluation assessed the service-learning project’s impacts and provided judgment of the project’s effectiveness.
## Use CIPP Model to Guide Service-learning Project

<table>
<thead>
<tr>
<th>Context, Input, Process, and Product Model Components</th>
<th>Methods Used in the Service-learning Tutoring Program</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Component I: Context evaluation</strong></td>
<td>• Assessed the setting for the intended service</td>
</tr>
<tr>
<td>Identify the needs, assets and opportunities for addressing the needs</td>
<td>• Interviewed school principal, teachers, and reading specialists</td>
</tr>
<tr>
<td></td>
<td>• Reviewed school records</td>
</tr>
<tr>
<td></td>
<td>• Identified at-risk readers and their needs</td>
</tr>
<tr>
<td></td>
<td>• Conducted initial quantitative assessment of at-risk readers</td>
</tr>
<tr>
<td></td>
<td>• Conducted pre-service teacher focus group interviews</td>
</tr>
<tr>
<td></td>
<td>• Conducted initial quantitative assessments of pre-service teachers</td>
</tr>
<tr>
<td><strong>Component I: Input evaluation</strong></td>
<td>• Reviewed relevant literature</td>
</tr>
<tr>
<td>Describe a project to meet the identified needs and identify and assess project strategies and procedural designs</td>
<td>• Interviewed school principal, teachers, and reading specialists</td>
</tr>
<tr>
<td></td>
<td>• Interviewed university reading faculty and other experts</td>
</tr>
<tr>
<td></td>
<td>• Viewed exemplary projects</td>
</tr>
<tr>
<td></td>
<td>• Consulted Learn and Serve America</td>
</tr>
<tr>
<td></td>
<td>• Formed advocate teams</td>
</tr>
<tr>
<td></td>
<td>• Conducted bi-weekly updates</td>
</tr>
<tr>
<td></td>
<td>• Conducted pre-service teacher focus group interviews</td>
</tr>
<tr>
<td><strong>Component III: Process evaluation</strong></td>
<td>• Identified what activities should be monitored</td>
</tr>
<tr>
<td>Monitor project’s process and potential procedural barriers and identify needs for project adjustments</td>
<td>• Received bi-weekly update from service-learning taskforce</td>
</tr>
<tr>
<td></td>
<td>• Observed service-learning activities</td>
</tr>
<tr>
<td></td>
<td>• Kept a log of the activities</td>
</tr>
<tr>
<td></td>
<td>• Interviewed at-risk readers</td>
</tr>
<tr>
<td></td>
<td>• Interviewed pre-service teachers</td>
</tr>
<tr>
<td></td>
<td>• Interviewed school principal, teachers, and reading specialists</td>
</tr>
<tr>
<td></td>
<td>• Reviewed pre-service teachers’ self-reflections</td>
</tr>
<tr>
<td></td>
<td>• Reviewed students’ work samples</td>
</tr>
<tr>
<td></td>
<td>• Conducted debriefing with pre-service teachers</td>
</tr>
<tr>
<td><strong>Component IV: Product evaluation</strong></td>
<td>• Conducted post-project quantitative assessments of pre-service teachers</td>
</tr>
<tr>
<td>Measure, interpret, and judge project outcomes and interpret their merit, worth, significance and priority</td>
<td>• Conducted post-project focus group interview of pre-service teachers</td>
</tr>
<tr>
<td></td>
<td>• Conducted post-project quantitative assessment of at-risk readers</td>
</tr>
<tr>
<td></td>
<td>• Administered at-risk readers survey</td>
</tr>
<tr>
<td></td>
<td>• Interviewed or surveyed other stakeholders including faculty instructor, principal, teacher, reading specialists and parents of at-risk readers</td>
</tr>
</tbody>
</table>
Applied Research Study Series Example #6
Providing physical activity for students with intellectual disabilities

- **A Close Look at the Fitness Level of Elementary Students with Intellectual Disabilities.**
  Published in *Sport Science Review*.

- **Providing physical activity for students with intellectual disabilities: The motivate, adapt, and play (MAP) program.**
  Published in *Journal of Physical Education, Recreation & Dance*.

- **Promoting Health-Related Fitness for Elementary Students with Intellectual Disabilities through a Specifically Designed Activity Program.**
  Published in *Journal of Policy and Practice in Intellectual Disabilities*.

In collaboration with Boswell, Davis, Decker, & Hodson.
Applied Research Study Series Example #6

Providing physical activity for students with intellectual disabilities

Background

- In NC, 20% of children are overweight or obese; NC is #5 in the nation for childhood obesity (BeActiveNC, 2006).

- Children with disabilities achieve insufficient physical activity for health purposes (Faison-Hodge & Poretta, 2004).

- Children with disabilities have high risk of Cardiovascular disease, diabetes, and hypertension.

- There has been little help on improving physical activity levels of students with intellectual disabilities.
Participants

- 4 elementary self-contained classes
- 25 students with intellectual disabilities, ages 8-12
- 9 girls, 16 boys
- 11 Caucasian, 11 African American, 3 Hispanic
Pre-test Fitness Data

The pre-test data showed that students with intellectual disabilities had low fitness levels compared to healthy standards set for typical students:

- 31% had an unhealthy weight
- 93% did not meet standard for PACER
- 77% not able to complete the minimum standard for the modified pull-up
- 35% did not meet criteria for sit-and-reach
Applied Research Study Series Example #6

Providing physical activity for students with intellectual disabilities

- **Motivate**: The key is to find out what motivates students with intellectual disabilities to participate (work hard) in physical activity.

- **Adapt**: Teachers adapt all MAP activities to fit the ability/development of each student with intellectual disabilities.

- **Play**: MAP activities were designed to be fun, make students laugh, and continue physical activity. MAP activities included dance, yoga, motor skills, aerobics, and manipulative skills.

***MAP program is designed by: Dr. Boni Boswell, Whitney Woodhall-Smith, Danielle Bogner.***
### Applied Research Study Series Example #6

**Student Performance Before and After the MAP Program (n=25)**

<table>
<thead>
<tr>
<th>Measures</th>
<th>Pre-Post</th>
<th>Mean</th>
<th>Sd</th>
<th>t</th>
<th>df</th>
<th>p-val</th>
<th>ES (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI</td>
<td>Before</td>
<td>19.96</td>
<td>5.92</td>
<td>.056</td>
<td>24</td>
<td>.956</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>After</td>
<td>19.94</td>
<td>5.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PACER</td>
<td>Before</td>
<td>10.28</td>
<td>11.72</td>
<td>2.69</td>
<td>24</td>
<td>.013*</td>
<td>.55</td>
</tr>
<tr>
<td></td>
<td>After</td>
<td>13.24</td>
<td>10.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified Curl-up</td>
<td>Before</td>
<td>7.12</td>
<td>11.80</td>
<td>2.089</td>
<td>24</td>
<td>.048*</td>
<td>.42</td>
</tr>
<tr>
<td></td>
<td>After</td>
<td>12.64</td>
<td>10.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medicine Ball Throw</td>
<td>Before</td>
<td>101.52</td>
<td>73.92</td>
<td>2.081</td>
<td>24</td>
<td>.048*</td>
<td>.42</td>
</tr>
<tr>
<td></td>
<td>After</td>
<td>129.08</td>
<td>80.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sit-and-Reach Right</td>
<td>Before</td>
<td>23.08</td>
<td>7.49</td>
<td>4.278</td>
<td>24</td>
<td>.000*</td>
<td>.882</td>
</tr>
<tr>
<td></td>
<td>After</td>
<td>27.10</td>
<td>8.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sit-and-Reach Left</td>
<td>Before</td>
<td>21.86</td>
<td>7.84</td>
<td>3.29</td>
<td>24</td>
<td>.000*</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>After</td>
<td>26.68</td>
<td>8.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedometer Steps</td>
<td>Before</td>
<td>3185.20</td>
<td>1236.68</td>
<td>.111</td>
<td>24</td>
<td>.913</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>After</td>
<td>3220.72</td>
<td>1740.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: The * indicates statistical significance.*
Podcasting: Engaging preservice teachers in learning about reading education with Web 2.0 technologies.
(with Swaggerty & Atkinson) published in Teacher Education Quarterly.

- This study examines the impact of a podcasting project on preservice teachers’ interests and attitudes in two sections of an undergraduate reading methods course (n=54).

  Students worked in pairs or groups of three to create an episode of the “Literacy Superstar Podcast” for the course website. Students chose a prominent figure in the field of literacy education, researched the literacy superstar’s life and his or her impact on the field, synthesized the information, and created a script for the project.
### Applied Research Study Example #7

**Comparison of Preservice Teachers’ Interest in Using Podcasts for Educational Purposes before and after the Podcast Project**

![Bar chart showing interest levels before and after the project.](chart)

<table>
<thead>
<tr>
<th>Survey</th>
<th>n</th>
<th>Mean</th>
<th>sd</th>
<th>t</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>54</td>
<td>2.37</td>
<td>0.808</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>After</td>
<td>54</td>
<td>3.28</td>
<td>0.856</td>
<td>6.829</td>
<td>.000</td>
<td>0.932</td>
</tr>
</tbody>
</table>
Selected Grants

Guili Zhang, PI/PD. IGERT: Spatial Ecology and Evolution: Integrative Quantitative Training in Biology, Statistics, and Mathematics, Project evaluation grant. Funded by the National Science Foundation (NSF). The Integrative Graduate Education and Research Traineeship (IGERT) is NSF’s flagship interdisciplinary training program, educating U.S. Ph.D. scientists and engineers by building on the foundations of their disciplinary knowledge with interdisciplinary training.

Guili Zhang, Program Evaluator & Statistical Analyst. SUCCEED (Southeastern University and College Coalition for Engineering Education). Funded by NSF. It takes a holistic approach to curriculum reform and leads the systemic reform in the nation’s engineering education.

- Guili Zhang, Program Evaluator & Statistical Analyst. “Studies Using the Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)”. Funded by NSF. The MIDFIELD database has been a unique and an exceptionally useful source for the nation’s engineering education research.
Selected Grants, cont’d

• **Guili Zhang, Project Evaluator, Teacher Quality Partnership Project (TQP).** Funded by the U.S. Department of Education. The TQP Grants Program seeks to improve the quality of new teachers by creating partnerships among IHEs, high-need school districts (local educational agencies (LEAs)) their high-need schools, and/or high-need early childhood education (ECE) program.


• **Guili Zhang, Co-PI (with Davis, Boswell, Decker, & Hodson). Motivate, Adapt, and Play (MAP).** Funded by Division of Research and Graduate Studies Research Development Grant. This project involves developing, implementing, and assessing a physical activity program for youth with intellectual disabilities in order to increase their health-related fitness level.
Selected Grants, cont’d

- **Guili Zhang, PI. Learning to Teach, Learning to Serve Evaluation Grant.** Funded by Learn & Serve America. Students achieve learning goals while providing meaningful service to the community.

- **Guili Zhang, Co-PI (with Zeller). Learning to Teach, Learning to Serve Evaluation Grant.** Funded by Learn & Serve America. Students achieve learning goals while providing meaningful service to the community.

- **Guili Zhang, Co-PI (with Shea). Learning to Teach, Learning to Serve Program Grant.** Funded by Learn & Serve America. It aims to strengthen teacher preparation programs by incorporating service learning in teacher education curriculum.

- **Guili Zhang, PI. 2012 AERA Institute on Statistical Analysis for Education Policy Grant.** Funded by the American Educational Research Association (AERA) Grants Program which is supported by NSF.
National Leadership Services

- **Chair**, Quantitative Methods TIG, American Evaluation Association.


**Chair**, Outstanding Quantitative Dissertation Award Committee of American Educational Research Association (AERA), 2009-2011.

- **Program Co-Chair**, Quantitative TIG, American Evaluation Association.

- **Chair**, Phenomenology, Hermeneutics, & Critical Theories in Qualitative Research session of the 2012 AERA Annual Meeting in Vancouver, British Columbia, Canada.
National Leadership Services, cont’d

- **Advisory Board Member**: *Journal of Curriculum and Instruction*, 2009-present.

- **Chair**, Research on Evaluation Session, American Educational Research Association Annual Conference.


National Leadership Services, cont’d

- **Chair**, Learning Communities Session, American Institute of Higher Education National Conference, Orlando, FL, April 2008

- **Co-Editor**, *Journal of Curriculum and Instruction*, 2006-2010.

- **Invited Presenter**: Beijing Teacher’s Training Center for Higher Education, People’s Republic of China

- **Advisor**: Shandong Normal University, People’s Republic of China
Selected National Awards & Honors

- **Outstanding Professor and Researcher**
The Government of the United States (USCIS category E17 – Outstanding Professor and Researcher).

- **2010 Edward C. Pomeroy Award for Outstanding Contributions to Teacher Education** (with JoCI Co-editors: Atkinson, O’Connor, Steinweg, Kester, Rodriguez, & Swaggerty)
The American Association of Colleges for Teacher Education.

- **Benjamin J. Dasher Best Paper Award**
  Frontiers in (engineering) Education (FIE)

- **Best Paper Award**
  American Society for Engineering Education (ASEE)
Selected National Awards & Honors

- **2012 AERA Institute on Statistical Analysis for Education Policy Award.**
  American Educational Research Association & National Science Foundation.

- **2010 American Educational Research Association Statistics Institute for Faculty Award**
  National Science Foundation & American Educational Research Association

- **American Educational Research Association Outstanding Quantitative Dissertation Award Finalist**
  American Educational Research Association

- **Excellent Lesson Award**
  National Teaching Competition, Education Ministry of China

- **Excellent Lesson Award**
  The 6th Annual Teaching Research Conference of Eastern China (Six Provinces)
Thank You