# A Data Analytics Framework to Support Training Effectiveness Evaluation

Paul Cummings paul.cummings@icf.com
Chalinda Weerasinghe CD.Weerasinghe@icf.com
Rodney Long rodney.a.long3.civ@mail.mil

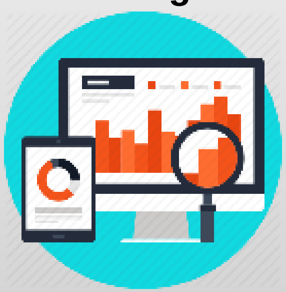
# Introduction

- **Course Level Analytics:** Evaluate course effectiveness as well as student performance within the course. Examine course performance data, learner attributes, and learning assessments and correlations between them.
- **Synthetic class data (authoring tool)**: Given minimal amount of existing intelligent tutoring data, an authoring tool for generating synthetic data would enhance the ability to build models and conduct robust experiments.
- **Report based natural language generation**: Need for **simple, user-intuitive** way will aid course evaluators, instructional designers, and content managers to better assess the effectiveness of a course.

# Data Modeling and Analysis Approach

1. Develop synthetic class assessments and correlations
2. Develop distributions of data based on characteristics of profiles
3. Evaluate data analytics and answer research questions
4. Validate results compared to synthetic data model

**Develop Synthetic Learning Data**

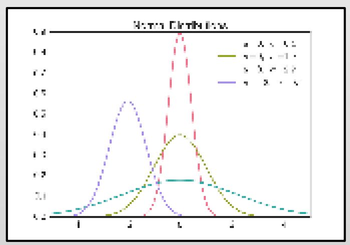


Assessments

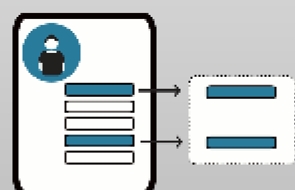
Curriculum

Profile data

**Simulated Data**



Population Data



User Persona Data

**Data Analysis**

Produces answers to research questions

Data Visualization

Descriptive/Inferential Statistics

**Validation**

Compare to initial data model

Are we able to discern the population created by the authoring tool?

# Generating Synthetic Data

The authoring tool is designed to generate a population of data that can be used to model persona attributes and class characteristics and correlations within and between these.

The tool is divided into three major sections:

1 - A **Persona,** which is a representation of a type of student with attributes in a class

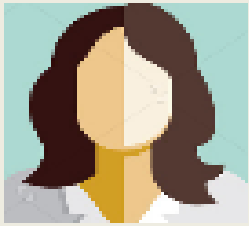
2 - A **Class,** which contains groups of students

3 – **Assessments,** which are ways that the class is evaluated

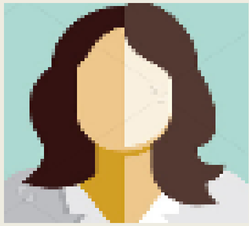
# Developing Populations with Personas

A **Persona** is an aggregate representation of a group of individuals each containing similar random **attributes** such as age, experience, and other demographics

Senior Scientist
Age: 45-65
Experience: 20+
Education: PhD



Mid Level PM
Age: 25-30
Experience: 6+
Education: MA/PMP



Sr. Level ISD
Age: 35-42
Experience: 8+
Education: MA/PhD

## Example Attributes

**Age**

- Nearly one-half (49.6%) of Active Duty enlisted personnel are 25 years of age or younger, with the next largest age group being 26 to 30 years (22.1%),
- 31 to 35 years (14.0%), 36 to 40 years (8.8%), and those 41 years or older (5.6%).
- More than one-quarter (25.7%) of Active Duty officers are 41 years of age or older, with the next largest age group being 26 to 30 years (22.5%),
- 31 to 35 years (20.7%), 36 to 40 years (17.8%), and those 25 years or younger (13.4%). Overall, the average age of the Active Duty force is 28.6 years.

**Marital Status:**

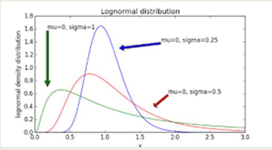
- Just over half (55.3%) of Active Duty military members are married, which is lower than the percentage that were married in 1995 (59.9%).
- In 2014, over half (52.1%) of enlisted members and a majority (69.9%) of officers report themselves as married.
- Over half (57.0%) of Active Duty males and nearly half (45.4%) of Active Duty females are married. In addition,
- 6.4 percent of DoD's Active Duty members are in dual-military marriages.[2]
- During the 2014 fiscal year, 3.5 percent of enlisted personnel and 1.8 percent of officers are estimated to have divorced.

# Developing Assessments

**Assessments** allow the system operator to *simulate performance* at multiple levels of evaluation (Kirkpatrick)

This model uses two of the four levels of evaluation that are used to evaluate performance (Kirkpatrick, 1994):

- Level 1  learner's reaction and satisfaction post-event
- Level 2 knowledge and skill gains that the learner exhibits




- Type
- Continuous/Discrete
- Mean
- SD
- Distribution Types

# Developing a Class

A **Class** is a generic representation of group of students and assessments given to those students

- Distribution of persona types. For example I may select
  - 40% of persona A,
  - 10% of persona B
  - 50% of persona C

- Sequence of assessments: For example, pre-test, post-test, survey

- Assessment and cumulative weighting and grading scales

Personas

Assessments

# Generating Dependencies

**Dependencies** (via correlations) are statistical relationships that describe how two or more random variables are linked to each other.

Correlations can take place
1-Between attributes (age/experience)
2-Between assessments (pre-test/post-test)
3-Between attributes and assessments (experience/pre-test)

Persona

Assessments

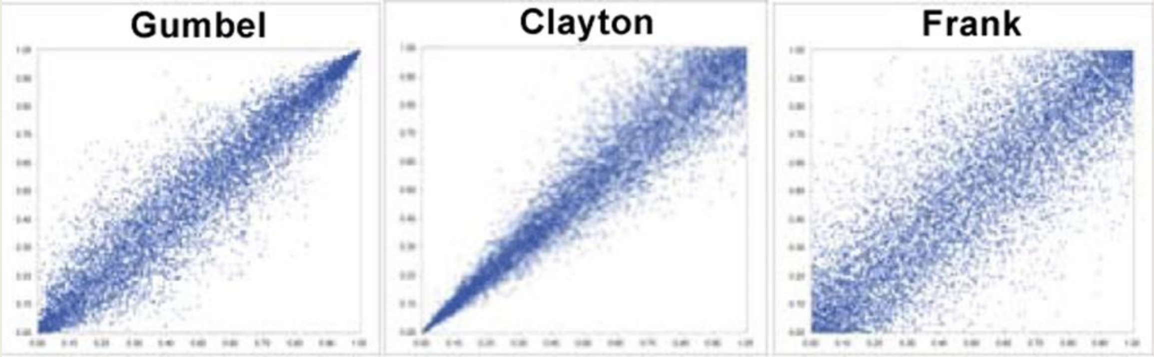
Correlations (attributes - assessments)

Correlate Attributes - Attributes

Correlations (assessments - assessments)

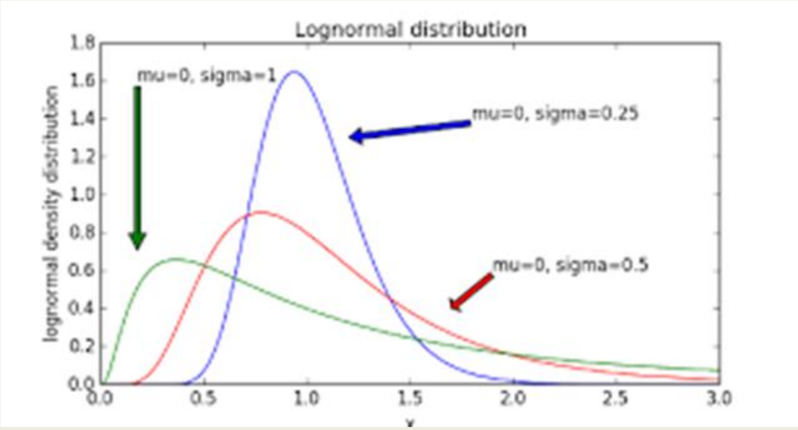
# Simulating Dependency with Copulas

- **Copula functions**: efficient method for describing non-normal (and normal) bivariate/multivariate correlated joint distributions of data.
  - For example, when describing the relationship between Age and Experience, we will want to describe a ‘tightness’ of correlation, direction, as well as a shape and tail behavior.
- Tightness describes the amount of correlation (0-1 with one being perfectly correlated), direction could be positive or negative, and the shape and tail behavior describes the symmetry and extreme probabilities respectively.
- Our model uses three types of Archimedean copula functions; Gumbel, Clayton, and Frank.

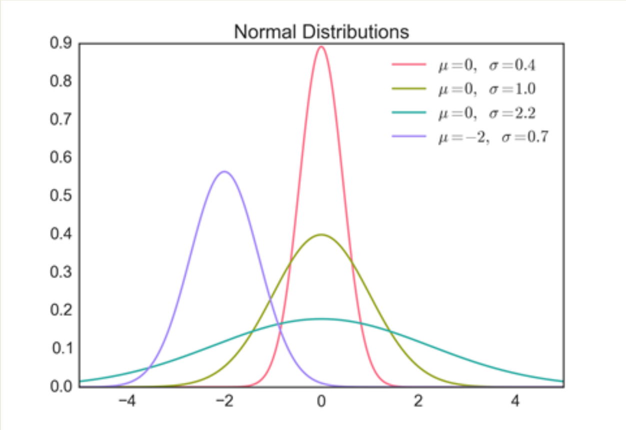


# Creating Random Variables

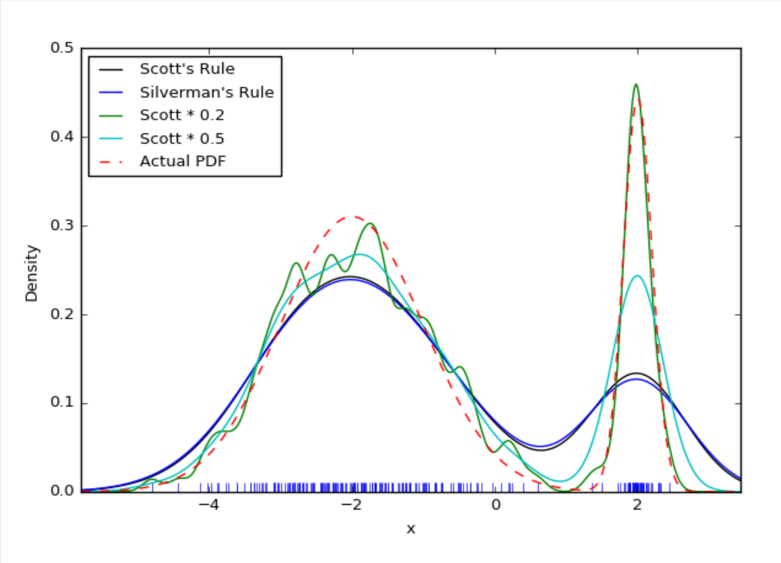
Generate variable distributions with flexibility using authoring tool




Varying Distribution Types

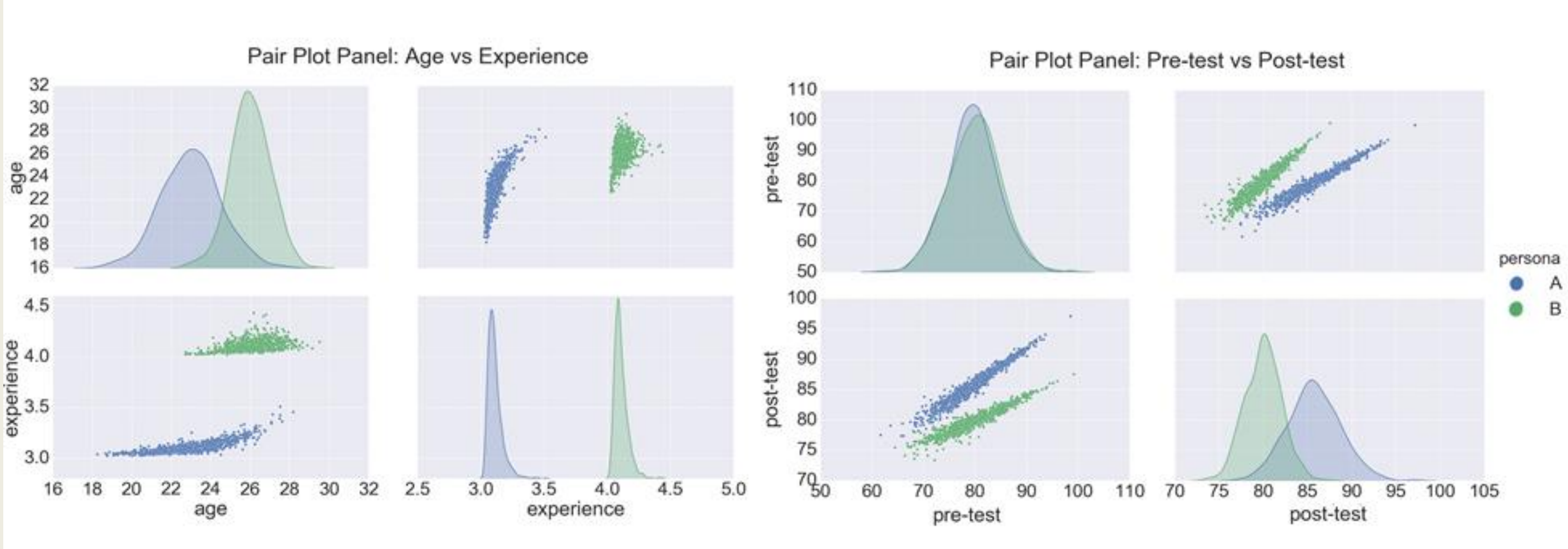



Varying distribution parameters




Modeling joint distributions

# Results




**Sample of Attribute Clusters by Persona.** From top-left to bottom-right, left to right across: the distributions for age; the scatterplot of experience and age; the distributions for pretest scores; the scatterplot of post-test and pre-test scores; the scatterplot of age and experience; the distributions for experience; the scatterplot of pre-test and post-test scores; and the distributions for post-test scores. Kernel distributions shown.

U.S. ARMY RDECOM

# Analytic Engine Validation Levels

ARL

Proposal for Analytics Engine Data Validation-High Level (Non-Predictive)

Data Imported into Engine

## Level 1 Analysis: Overall Class Patterns (Default)

**Does the data show clusters in the aggregate?**
K means clustering- Data is at least interval
Hierarchical clustering- Data is ordinal

**What are the empirical distributions of the variables?**
Histograms/kernel densities- data is at least ordinal
Frequency distributions- data is categorical
Create empirical distributions for the data
Test for common distributions:
Shapiro-Wilkes (Norm)
Kolmogorov-Smirnov- 1 and 2 sample test

## Level 2 Analysis: Between Cluster /Variable Patterns

**Does the data show clustering patterns across variables?**
Bi (and tri?) variate scatter plots- data is at least interval
K means clustering- Data is at least interval
Hierarchical clustering- Data is ordinal

**What are the associations across the variables of interest?**
**Bivariate:** Pearson's r- Data is continuous
Kendall's tau- Data is at least ordinal
Point Biserial correlation- Between binary & continuous data
Phi coefficient- Data is binary
Cramer's V- Data is nominal
**Multivariate**: Vine/bivariate copula analytics

**What are the inferences across the variables of interest?**
Inferences on means- t tests, one way ANOVA, etc.
Inferences on proportions- z test, Chi square test
Inferences on variance- F test
Inferences on medians- nonparam tests
Inferences on independence- Chi square, ANOVA
Inferences on relationships- General linear models, generalized linear models, other models

## Level 3 Analysis: Within Cluster /Variable Patterns

**Does the data show clusters within the variables?**
K means clustering- Data is at least interval
Hierarchical clustering- Data is ordinal

**What are the descriptive statistics of the variables of interest?**
Observation count, class mean, median, mode, standard deviation, min, max, inter-quartile range, skewness, kurtosis

**What are the inferences within the variables of interest?**
Inferences on frequency- Chi squares

**What are the underlying factors within a variable interest?**
Factor analysis

# Use Case: Analytics Framework

Methods and attributes developed in prior research (Long et al., 2016a; Long et al., 2016). Successful detection and validation of these personas and their narratives (Weerasinghe et al., forthcoming).

| *Persona A* | *Persona B* |
|---|---|
| • Learners' age and experience were positively correlated (moderate to strong, Kendall's τ ≈ 0.6)<br>• At lower ages, there was lower experience with more variation (more probability of falling at this end),<br>• At higher ages, there was higher experience with little variation (less probability of falling at the high end).<br>• The pre-test score was independent of age and experience.<br>• Pre-test score was strongly, positively correlated with post-test assessment (τ ≈ 0.85)<br>• Those who scored lower on the pre-test scored lower with more variation on the post-test (more probability of falling at the low end), while those who scored higher at the pre-test scored higher in the post-test with little variation (less probability of falling at the high end). | • Learners' age and experience were weakly and positively correlated (τ ≈ 0.3)<br>• At lower ages, there was lower experience with less variation (less probability of falling at this end)<br>• At higher ages, there was higher experience with more variation (higher probability of falling at the high end).<br>• The pre-test score was independent of age and experience.<br>• Pre-test score was strongly positively correlated with post-test assessment (τ ≈ 0.8).<br>• Those who scored lower on the pre-test scored lower with more variation (more probability of falling at this end). Those who scored higher on the pre-test scored higher in the post-test with little variation (less probability of falling at the high end). |

## Research Questions for Analytics Engine to Answer

Example Question: How did the class(es) perform?

1. What are the statistical characteristics of the question?
2. How can we begin by distilling question into statistical and/or computational approach?
3. What does the 'natural language' answer give the operator?

U.S. ARMY RDECOM

# Analytics Framework: Research Question Reporting

ARL 25 YEARS U.S. ARMY RESEARCH LABORATORY 1992-2017

## LP Question 1: How did the class(es) perform?

### System Default Metrics/Goals

Based on your grading scale and how you wanted to weight the assessment scores, here is breakdown of the scores for the class:

Frequency of Students in Grading Categories

| Grade Category | Pre-Test | Post-Test | Final Percent Score |
|---|---|---|---|
| A | 49 | 75 | 46 |
| B | 936 | 1408 | 1275 |
| C | 960 | 517 | 679 |
| D | 55 | 0 | 0 |
| F | 0 | 0 | 0 |

Percentage of Students in Grading Categories

| Grade Category | Pre-Test | Post-Test | Final Percent Score |
|---|---|---|---|
| A | 2.45% | 3.75% | 2.30% |
| B | 46.80% | 70.40% | 63.75% |
| C | 48.00% | 25.85% | 33.95% |
| D | 2.75% | 0.00% | 0.00% |
| F | 0.00% | 0.00% | 0.00% |

**NLF:** Based on the grade distributions in your class, 46 students (2.3%) got A's, 1275 students (63.75%) got B's, and 679 (33.95%) got C's. There were no D's or F's. Please see the Tables for breakdown on the assessments.

The Bar graph below shows the grades on the Pre-test



The Nation's Premier Laboratory for Land Forces

# Example Natural Language Reporting Q3

## LP Question 3: Were there factors related to success or failure within the class?

**System Default Metrics/Goals**

**Class Factors**

There are several factors related to the **overall success/failure** of your students in the class. Here is a Table that shows these factors:

| Factors Related to Success/Failure | Positive/Negative | Strength | Statistical Significance |
|---|---|---|---|
| Age | Negative | Low | Very Significant |
| Experience | Negative | Low | Very Significant |

**NLF:** Based on this Table, an increase in the final score for students in your class was associated with a decrease in age (negative relationship). That is, on average older students tended to do worse than younger students. The effect was low in strength (the relationship is not strong between age and final score). This relationship is statistical very significant: the probability is very high that the relationship observed could not happen by chance alone.

**NLF:** Based on this Table, an increase in the final score for students in your class was associated with a decrease in experience (negative relationship). That is, on average more experienced students tended to do worse than less experienced students. The effect was low in strength (the relationship is not strong between experience and final score). This relationship is statistical very significant: the probability is very high that the relationship observed could not happen by chance alone.

**Assessment Factors**

There are several factors related to the success/failure of your students in the **various assessments** given. Here is a Table that shows these factors:

| Assessment Name | Factors Related to Success/Failure | Positive/Negative | Strength | Statistical Significance |
|---|---|---|---|---|
| Pretest | Age | Positive | Very Low | Very Insignificant |
| | Experience | Positive | Very Low | Insignificant |
| Posttest | Age | Negative | Moderate | Very Significant |
| | Experience | Negative | Moderate | Very Significant |

**NLF:**

**For the Pretest:**

Based on this Table, an increase in the pretest score for students in your class was associated with an increase in age (positive relationship). That is, on average older students tended to do better than younger students. The effect was very low in strength (the relationship is almost nonexistent between age and pretest). This relationship is statistical very insignificant: the probability is very low that the relationship observed could not happen by chance alone.

Based on this Table, an increase in the pretest score for students in your class was associated with an increase in experience (positive relationship). That is, on average more experienced students tended to do better than less experienced students. The effect was very low in strength (the relationship is almost nonexistent between experience and pretest). This relationship is statistical insignificant: the probability is low that the relationship observed could not happen by chance alone.

**For the Posttest:**

Based on this Table, an increase in the posttest score for students in your class was associated with a decrease in age (negative relationship). That is, on average older students tended to do worse than younger students. The effect was moderate in strength (the relationship is quite strong between age and posttest). This relationship is statistical very significant: the probability is very high that the relationship observed could not happen by chance alone.

Based on this Table, an increase in the pretest score for students in your class was associated with a decrease in experience (negative relationship). That is, on average more experienced students tended to do worse than less experienced students. The effect was moderate in strength (the relationship is quite strong between experience and pretest). This relationship is statistical very significant: the probability is very high that the relationship observed could not happen by chance alone.

**Instructor-Defined Metrics/Goals**

**Assessment Goals:**

**Goal 1:** What is the most important factor relating to pretest performance?

**NLF:** The most important factor relating to pretest performance is experience. See explanation above.

**Goal 2:** What is the most important factor relating to posttest performance?

**NLF:** The most important factor relating to posttest performance is experience. See explanation above.

**Overall Class Goals:**

**Goal 1:** What is the most important factor relating to overall class performance?

**NLF:** The most important factor relating to posttest performance is experience. See explanation above.

# Discussion

- **Detection**: The results indicated that the automated analysis tool for GIFT successfully detected the persona groups from the learner population at each level of validation.

- **Usability**: Design creates a fairly intuitive way of representing both data generation (personas, classes, assessments, correlation) and data presentation

- **Scalable complexity**: Given the success of this relatively simple model, the data authoring tool can be used to develop more complex interdependencies for larger numbers of variables.
  - Also able to design random experiments for better inferential and predictive analysis.
  - Copulas have been demonstrated to be a highly scalable tool for modeling relationships with a number of desirable statistical properties.

# Future Research

- While initial development focuses on **descriptive** and preliminary inferential analysis, future extensions will easily transition to deeper **inferential** methods and predictive applications.
- Use of **multivariate copulas** (using vines) in more general dependency settings (Bedford & Cooke, 2002; Kurowicka & Joe, 2011).
- While the initial demographic analysis variables are limited, it is straightforward to **scale up** and include more variables and inputs in the data authoring tool and to allow for more variation.

# Briefing References


- Bedford, T.J. & Cooke, R.M. (2002). Vines- A new graphical model for dependent random variables. *Annals of Statistics 30*(4): 1031–1068.

- Kurowicka, D. & Joe, H. (2011). *Dependence modeling: Vine copula handbook.* World Scientific, Singapore.

- Long, R., Smith, M., Dass, S., Dillon, C., & Hill, K. (2016a). *Data Analytics: Techniques and Applications to Transform Army Learning*. ModSim World Conference, NDIA, 2016.

- Long, R., Smith, M., Dass, S., Dillon, C., & Silverman, J. (2016b). *Authoring Tools to Assess Training Effectiveness in Intelligent Tutors.* I/ITSEC Conference, NDIA, 2016.

- Weerasinghe, C., Dillon, C., Cummings, P., Smith, M., & Long, R. (forthcoming). A Data Analytics Framework to Support Training Effectiveness Evaluation. GIFT Users Symposium, 2017.

# Backup

# Example Research Questions

- How much learning took place?
- What is the difference in results between this course offering and another course offering?
- How did attitude affect the results?
- Who did better in course?
- Who benefits most from the course?
- What affect performance?
- How do the results vary by instructor? (this seems more an admin question)