**Abstracts of Theses**

(1) Graduate research assistant, Venkata Ganesh Ashish Akula, is currently working on his Master's thesis based on the research work from this project. His Master's thesis is expected to be completed by December 2018 and the thesis abstract will be submitted then.

**Data on Scientific Collaborators**

(1) Dr. Chen Ling, Co-Principal Investigator, Associate Professor, University of Akron
(2) Venkata Ganesh Ashish Akula, Graduate Student, University of Akron
(3) Mohammed Abdul Sami, Graduate Student, University of Akron
(4) Kerry Holmes, Undergraduate Student, University of Akron

**Technical Description of Project and Results**

Through a variety of study approaches including (1) Literature review, (2) Review of best practices from other industries, (3) Review of case studies on transferring learning to behavior, (4) Phone interviews, (5) On-site interviews, and (6) Academy Evaluation System (AES) data mining, detailed recommendations were provided for Levels 1-5 evaluation processes based on Kirkpatrick model, on both continuous improvements of the existing processes and suggested future state maps, as well as data analytical strategies. The results are summarized below.

**1. Level 1 Evaluation Process**

The Level 1 evaluation through the EOC survey via AES is a well-established and efficient process. Minor process improvement opportunities are identified are follows:

(1) Expand the use of standardized EOC survey via AES to non instructor-led training courses.
(2) Actively track and enforce the review timelines (14 days for course coordinators/instructors and 30 days for AMA managers) to identify potential risks and take necessary actions in a timely manner.
(3) Use Average Weighted Score instead of Favorability Percentage as the performance metric for survey data analysis. This could be easily done using the existing AES, as the survey data are already in the required format. This change of performance metric will lead to the redesign of trigger point(s) for the feedback mechanism. Based on the other industry's practice, different training courses may use different trigger point(s) that is/are

most suitable for the specific courses. Once the trigger points are decided, they can be coded in AES to enable automatic triggers.

(4) Use advanced analytics (correlation analysis, contingency analysis, logistic regression model, and cluster analysis) to analyze the survey data. This could lead to insightful discoveries, which can be used to drive continuous improvement efforts. This can be done either offline periodically (easy approach) or automated in the AES (challenging approach).

(5) The use of text mining for automating the analysis of individual comments/critiques is currently immature, because the accuracy level is not high enough. It is still recommended for the review team to manually read the comments/critiques.

**2. Level 2 Evaluation Process**

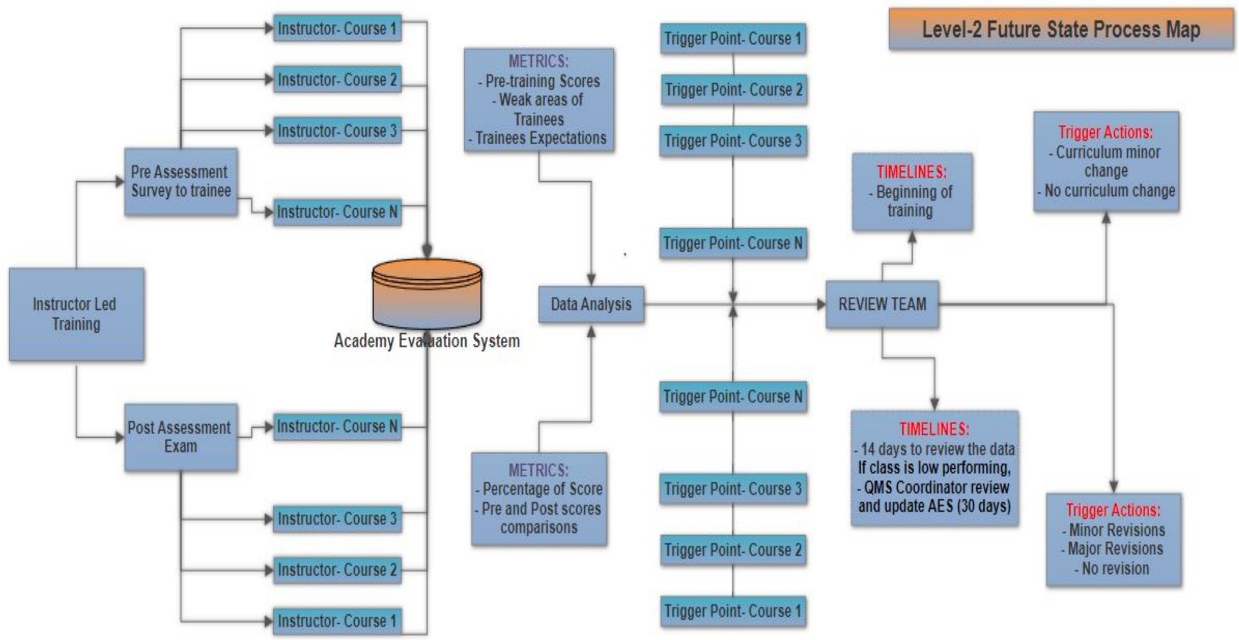
Currently, the Academy does not have a process for collecting and analyzing Level 2 data. The divisions within the Academy use objectives and develop tests/activities to determine if learning occurred in course offerings, but there is no centralized process to collect and analyze Level 2 data.

Kirkpatrick model's Level 2 measures the participants understanding to the course content by testing the knowledge and skills learnt in the training program through a Performance exam at the end of course. There are several data collection methodologies for the evaluation model. The most widely used data collections methods used in best practices in other industries are the pre and post-test assessments, interview before and after the training (which is a very time consuming process). Based on our interviews with AJI-2 customer satisfaction process stakeholders and the industry best practices, we recommend that the pre and post-test assessments is the best option considering the Academy's organizational structure and the number of students that attend the training. It would be difficult in selecting a sample of students and interviewing them to collect Level 2 data.

In the following Figure, we proposed a future state process map for the Level 2 evaluation at the Academy. Both pre-training assessment and post-training assessment need to be conducted. For the pre-training assessment, as each training course has its own training requirements and objectives, it is difficult to use a standardized instrument for pre-training assessment. It is recommended each course instructor conducts his/her pre-training assessment. This could be done using an actual assessment exam or an assessment survey that asks the trainees to answer a series of questions with ratings on their proficient levels on the relevant training tasks. In either case, the pre-training assessment results should be standardized and documented as pre-training assessment scores.

For the post-training assessment, it is recommended that the final course scores be used as the post-training assessment scores. Then, the comparison of the pre-training and post-training assessment scores should be documented as Level 2 data. Each training course could set up its own trigger point for the review process.

It is recommended to use a centralized system, such as AES, to document and track these Level 2 data.




Future State Level 2 Evaluation Process Map

### 3. Level 3 Evaluation Process

The Level 3 evaluation through the EOC survey via AES is a well-established and efficient process. Minor process improvement opportunities are identified are follows:

(1) Expand the use of standardized EOC survey via AES to non instructor-led training courses.
(2) Actively track and enforce the review timelines (30 days for course coordinators/instructors and 30 days for AMA managers) to identify potential risks and take necessary actions in a timely manner.
(3) Use Average Weighted Score instead of Favorability Percentage as the performance metric for survey data analysis. This could be easily done using the existing AES, as the survey data are already in the required format. This change of performance metric will lead to the redesign of trigger point(s) for the feedback mechanism. Based on the other industry's practice, different training courses may use different trigger point(s) that is/are most suitable for the specific courses. Once the trigger points are decided, they can be coded in AES to enable automatic triggers.
(4) Use advanced analytics (correlation analysis, contingency analysis, logistic regression model, and cluster analysis) to analyze the survey data. This could lead to insightful discoveries, which can be used to drive continuous improvement efforts. This can be done either offline periodically (easy approach) or automated in the AES (challenging approach).

(5) The use of text mining for automating the analysis of individual comments/critiques is currently immature, because the accuracy level is not high enough. It is still recommended for the review team to manually read the comments/critiques.

## 4. Level 4 and Level 5 Evaluation Process

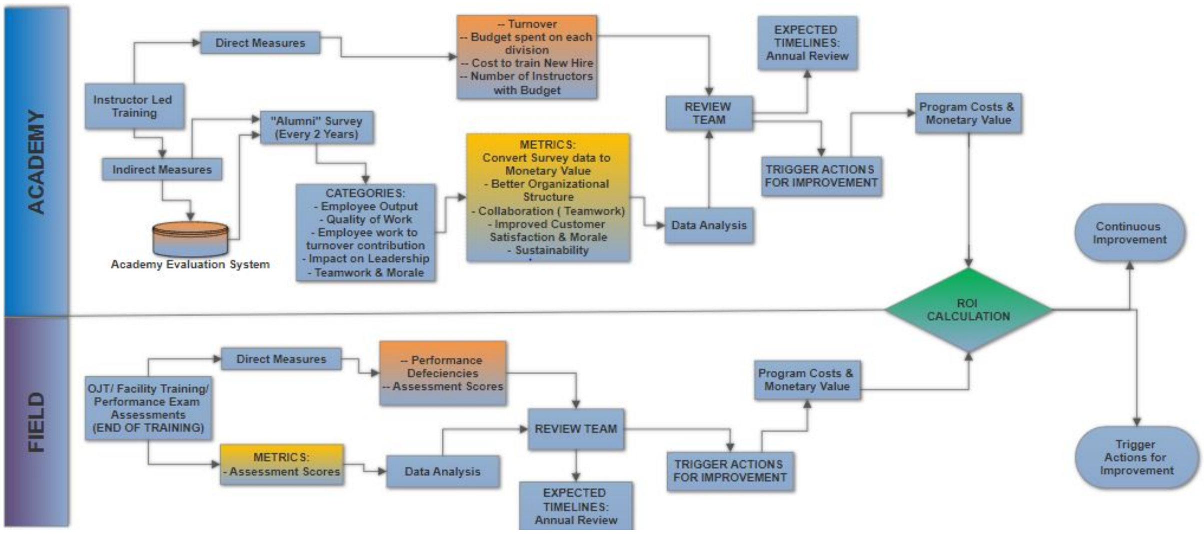
Currently, the AJI-2 and the Academy does not have a process in place to provide Level 4 and Level 5 evaluations. Kirkpatrick model's Level 4 evaluates the business results of the organization on which the training program was conducted. This level of evaluation involves deep understanding of the data obtained from Levels 1-3 and different data analytics have to be performed to obtain the results. Philip's Level 5 Return on Investment (ROI) is the ultimate level of evaluation. It compares the monetary benefits from the program with the program costs. Although the ROI can be expressed in several ways. The Phillips' model evolves from, and can be distinguished from, the earlier Kirkpatrick model by the adoption of return on investment to yield additional, critical insight. ROI allows decision makers to compare the ultimate value of a training investment with other potential investment opportunities.

From the best practices from the other industries, the data collection methodologies for level 4 and level 5 are largely based on online survey, focus group, ROI, value narration, tangible and intangible benefits. In the following Figure, we proposed a future state process map for the Level 4 and Level 5 evaluation process.

To measure the business results (level 4) and ROI (level 5), we propose to use both direct measures and indirect measures. Specifically, at the Academy, the direct measures include trainee turnover rates, division budgets, cost to train new hires, as well as the number of instructors and the associated budgets. The indirect measures can be obtained by sending out a survey to the "alumni" who successfully completed academy training and worked in the field. In the "alumni" survey, proposed questions are designed to gather information about employee work to turnover contribution, organizational structure, team work, morale, improved customer satisfaction and sustainability. Based on other industries' best practices, it is recommended the survey to be conducted every 2 years.

In the field, when the OJT, facility training, performance exam assessments are conducted, these performance assessment results can be used as Level 4 direct measures. Data analysis on these assessment results can be conducted to obtain metrics such as the assessment score summary statistics (mean, median, standard deviation, etc.), and the first pass yield. Another source of direct measures from the filed could be the documented performance deficiencies (if available). The challenge is that in the field, currently, there is not a centralized unit similar to the AMA-20 in the Academy to coordinate the data collection, analysis and review process. It is recommended further studies to be conducted to investigate the best ways to collect data from the field and establish an effective data sharing and information sharing mechanism among AJI-2, the Academy, and the field.

Furthermore, it is recommended to form a review team consisting of members from different backgrounds to review and analyze the data obtained from direct measures and indirect measures both from the Academy and the field, identify specific trigger points (such as, when employee turnover is more than 10%, or the ROI is less than 1:2, etc.), to take actions. The review cycle is recommended to be every one year.




Future State Level 4 and Level 5 Evaluation Process Map

## 5. Evaluation Data Analytical Strategies

Based on the data analysis results, we recommend the following data analytics strategies for the survey data analysis.

(1) Use Average Weighted Score instead of Favorability Percentage as the performance metric for survey data analysis. This could be easily done using the existing AES, as the survey data are already in the required format. This change of performance metric will lead to the redesign of trigger point(s) for the feedback mechanism. Based on the other industry's practice, different training courses may use different trigger point(s) that is/are most suitable for the specific courses. Once the trigger points are decided, they can be coded in AES to enable automatic triggers.

(2) Use advanced analytics (correlation analysis, contingency analysis, logistic regression model, and cluster analysis) to analyze the survey data. This could lead to insightful discoveries, which can be used to drive continuous improvement efforts. This can be done either offline periodically (easy approach) or automated in the AES (challenging approach).

Specifically, Logistic regression analysis was used to check the relation of the questions with the overall satisfaction rating as every question will play a major role in the trainee's perspective and we will be able to quantify which questions are adding more values to the overall satisfaction score. The outcomes from the logistic regression analysis are that we are able to identify the questions which are most significant and least significant in improving the overall satisfaction rating. The questions with least significant effect on question 14 are Q3, Q6, Q7 and Q12. The countermeasures for these outcomes are to look into the category of the question and determine the root cause for why there is less significance when compared with the overall satisfaction.

Over the three years period when the data was given, for the overall favorability percentages of all the questions, the lowest favorability percentage was 81.2% for Question 9. The improvement in the values of this question will have an impact on the overall satisfaction score, as it is one of the most significant questions in the prediction of the Question 14 responses. And this will in return improve the satisfaction score which can be converted into the tangible benefits as well as intangible benefits which can help calculate the Return-On-Investment (ROI).

The clustering analysis will help FAA in understanding which classroom falls into which cluster with specific cluster means for each of them. By looking at that it can be easily identified over the three year period which classes had fallen into a cluster with low means and high means, which will be resourceful to check what went wrong in those similar types of classes based on the responses, favorability percentage and the average weighted score.

(3) The use of text mining for automating the analysis of individual comments/critiques is currently immature, because the accuracy level is not high enough. It is still recommended for the review team to manually read the comments/critiques.

The text mining technique was most accurate when trained using survey results. The best method for improving its accuracy would be to acquire as many survey results as possible and continually use them to improve the classifier. The confusion matrix showing the results from that training shows that the classifier falsely identifies most of the comments as being negative. This is largely a result of how skewed the dataset is towards negative comments. While acquiring more survey data generally will improve accuracy, increasing the representation of positive and neutral comments will help improve this accuracy. There are other versions of text classifiers, such as random forests, which may perform better, even with the current data but more evaluation is required to determine their viability. In addition, a more rigorous feature extraction process can improve the accuracy of this tool; however, the computation time increases dramatically as the number of features extracted increases.

Another way to improve the accuracy would be to setup the tool as a semi supervised learning system. The system would be allowed to sort out positive comments, and the persons reviewing the negative comments would mark the wrongly classified positive and neutral comments. The system preferentially identifies comments as negative and, consequently, the system is not likely to wrongly classify something as neutral or positive. After the system sorts through the input text, it would be retrained on the updated classifications of all the results it has seen. This allows the system to be utilized now and become more useful overtime.