

College of Business Sources Standardization

Lean 6 σ Green Belt Project

Tonia Howe | June 1, 2020

Academic Manager, Business

Evaluation Faculty



Problem Statement:

There is variation in how the 225 evaluators within the College of Business (CoB) evaluate the Sources aspect in Performance Assessments. Each Team within the CoB receives direction on how to evaluate this aspect by their current Manager and/or Course Lead. College of Business students are impacted as they unknowingly cross evaluation teams. They are held to different standards for the Sources aspect despite receiving similar directions in the task prompts and being measured by the same or similar rubric requirements.

The variation in guidance and training provided to evaluators within the College of Business Evaluation Department has a negative impact on the calibration of this aspect. The average calibration score of Sources in the CoB in April 2020 was 3.637 with a standard deviation of 1.674. This level of calibration negatively impacts the student experience and student confidence in Evaluation Faculty.

Goal Statement:

In order to reduce the variation of the evaluations of the Sources aspect across CoB teams, the goal is to reduce the standard deviation of the calibration activity by 15-20%. The improved calibration will reduce rework by decreasing the number of resubmissions, reduce appeals due to incorrect scoring, and have a positive impact on student satisfaction.

Current State Analysis

Thought Process Map:

Questions	Stage	Answers
What is the Problem?	Define	The average error rate on the College of Business (CoB) baseline Sources calibration activity was 3.637 with a standard deviation of 1.674.
What is the Goal?	Define	Decrease the standard deviation on the Sources calibration activities by 15-20%. This indicated increased calibration, which will reduce the number of evaluation errors of this aspect, the amount of rework incurred through resubmissions and appeals, and increase student satisfaction.
What is Out of Scope?	Define	The accuracy of applying the correct rubric score is beyond the scope of the project. The errors in the calibration activity will be identified as pass/fail instead of requiring the correct score using the 3-tier rubric. Also, improving the pass rate of Sources for individual tasks is out of scope.
Approach	Define	People leaders were surveyed to determine how they direct their teams to evaluate the Sources aspect. Additionally, all individuals in the CoB participated in a baseline calibration activity in order to identify the current state of calibration across the CoB evaluators. Standardized guidelines on how to handle specific situations were established by the Senior Management team. These guidelines were then shared via a live Sources Department Training with all CoB Evaluation employees. The Standardized Training will take place in two phases. Phase 1 will focus on training and calibrating the People Leaders and Phase 2 will be focused on training the Evaluators. The training process will include individuals participating in the Department Training and Assessment, then all employees will complete a live training to overview the Sr Manager clarifications. After the training events are completed, employees will participate in a second standardized calibration activity to measure the impact of the training. Follow-up trainings and calibration activities will be done with all individuals who score below 80%.
		Reinforcement activities will include bi-weekly calibration activities for the first 3 months after the Standardized Training. Calibration activities will then be held on a quarterly basis. A template will be added to the Quarterly Audit form to address the Sources in the QAs. Standardized Course Resource Document (CRD) notes will be added to each course CRD in the CoB.
What is the Current State?	Define	The average defect rate in the CoB was 3.637 with a standard deviation of 1.67.
What is our Data Type?	Define	Attribute (Qualitative) since we are looking at Pass/Fail data.

Current State Analysis

Thought Process Map (cont.)

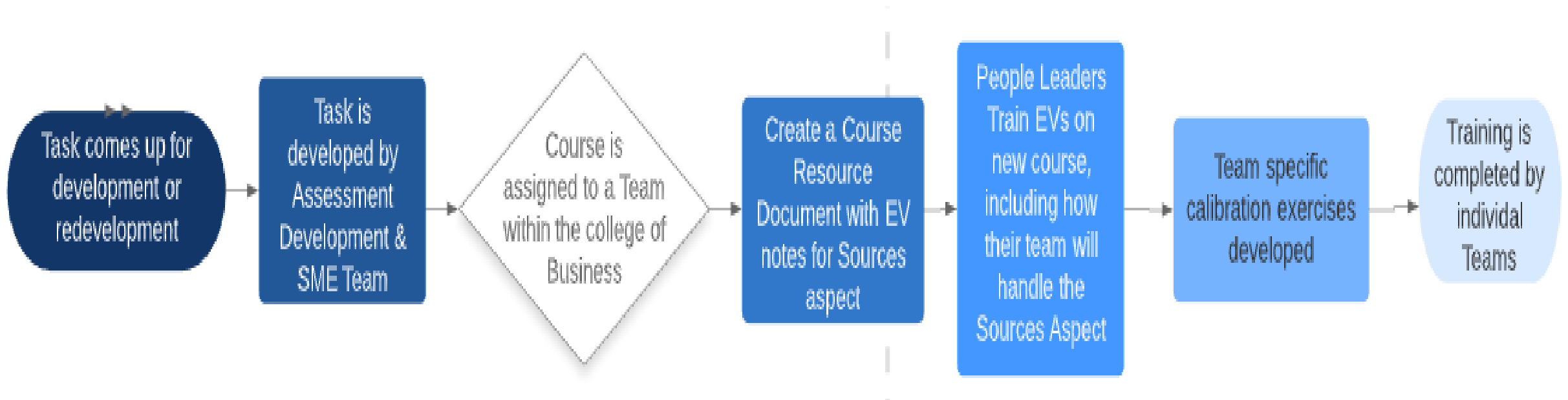
How will measurements be taken?	Measure	Current state metrics will be gathered by engaging all members of the CoB Evaluation team in calibration activities with 5 excerpts and 6 scenarios. The calibration activity will assess whether the individual would accurately pass or fail the situation presented.
What Tools will be used?	Measure	Forms Surveys will be used for the calibration activities MS Teams meetings will be used for Live Training Bridge will be used for the Department Training MS Excel will be used to analyze data MiniTab will be used to perform statistical analysis of data PowerBI and Cognos reports will be used to obtain any other relevant data
How do we determine the Standard Deviation?	Measure	The average defect rate will be calculated using the results of the calibration activity. The standard deviation will then be calculated.
How do we know what areas need improvement?	Analyze	The samples taken will be obtained in the form of concepts and scenarios tested in the calibration activity. A Pareto Chart will be created to identify the concepts tested that have the greatest number of errors (defects).
How do we perform statistical analysis of the data?	Analyze	Create Control Charts (Pareto, np-chart, Boxplot, and 1-sample Z) Calculate the mean and standard deviation of the calibration activities results for the CoB. I will then perform the same analysis on subgroups of the CoB broken down by People Leaders and 4 Senior Manager teams.
What are the solutions needed to address improvement areas identified above?	Improve	Creation of CoB Source guidelines at the Sr Manager level. Department Training on the 'Big 4' rubric is required. Standardized Evaluation Notes in all Course Resource Documents is required. Regular calibration exercises are required to measure EV understanding and identify individuals who need additional training.

Current State Analysis

Thought Process Map (cont.)

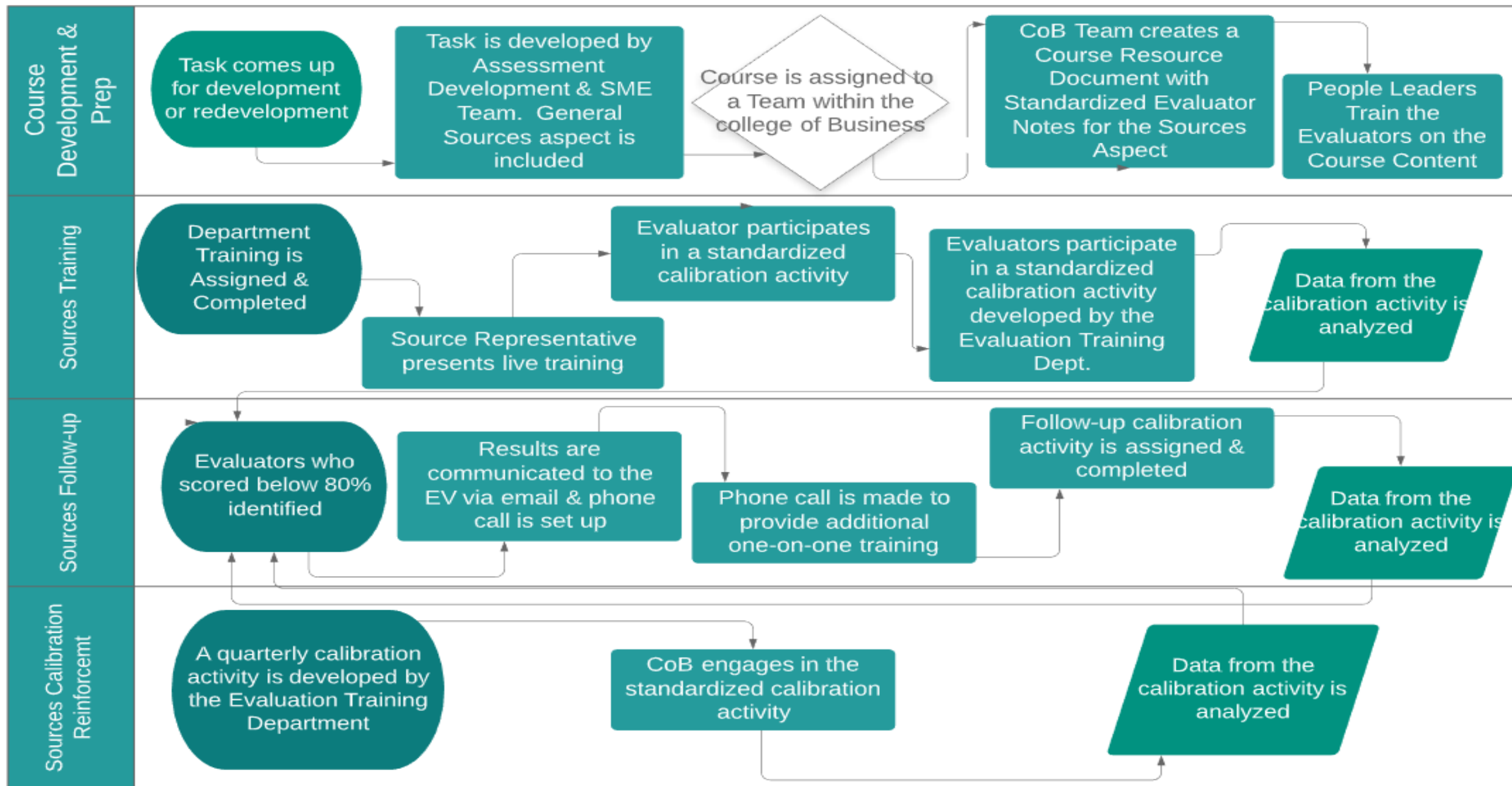
When, where and how will the process changes be implemented?	Improve	<p>In Phase 1 all People Leaders (PL) will be trained during May and June by using a Bridge course, a live training in Teams, and a second calibration exercise. Follow-up will be made with any PL who scores below 80% on the 2nd activity. The follow-up will include an email of activity 2 results to the PL, a phone call with each PL, and engagement in a 3rd calibration activity. Subsequent activities and follow-up will be provided until each PL scores 80% or higher. Quarterly calibration exercises will then be implemented to ensure continued calibration.</p> <p>Evaluators will be trained during the month of July using the same process. Team specific analysis will be done to adapt the live training. Additionally, PLs who did not have defects in Activity 2 will assist with the Team-level training and follow-up using standardized training tools with team specific examples.</p>
What metrics are needed to show efficiencies gained?	Improve	<p>The following metrics will be tracked:</p> <ul style="list-style-type: none"> • Average number of defects in the calibration activity and the standard deviation of each exercise • Estimated Cost Savings calculated using a conservative estimate of monthly evaluation errors of the Sources aspect.
How do we know the new process is being followed?	Control	Regular calibration exercises will be implemented and scored to determine whether Evaluators and People Leaders are still following the new training and guidelines. In Q1 of FY21 the calibration activities will be once a month. Beginning in Q2, the activities will be completed once a quarter. Additionally, the Quarterly Audit (QA) process will be adapted to include a review of the Sources aspect.
What can cause the new and improved process to fail and how can we address it?	Control	Resistance to change by the People Leaders and Evaluators and 'Evaluator drift' are the biggest issues we will encounter. These will be overcome by including the PLs in the EV training phase, providing consistent calibration exercises for Sources to identify EV drift, and supportive training exercises.
When and how often will metrics be gathered?	Control	Calibration results will be analyzed after each activity. Activity 2 metrics will be gathered for the entire CoB during July 2020. Additional metrics will be gathered after each calibration activity and then quarterly.
How do we know we met our objective?	Control	Determine statistical significance by using a 1 Sample Z test to compare future state metrics to the current state metrics.

Current State Process Map – Sources Calibration



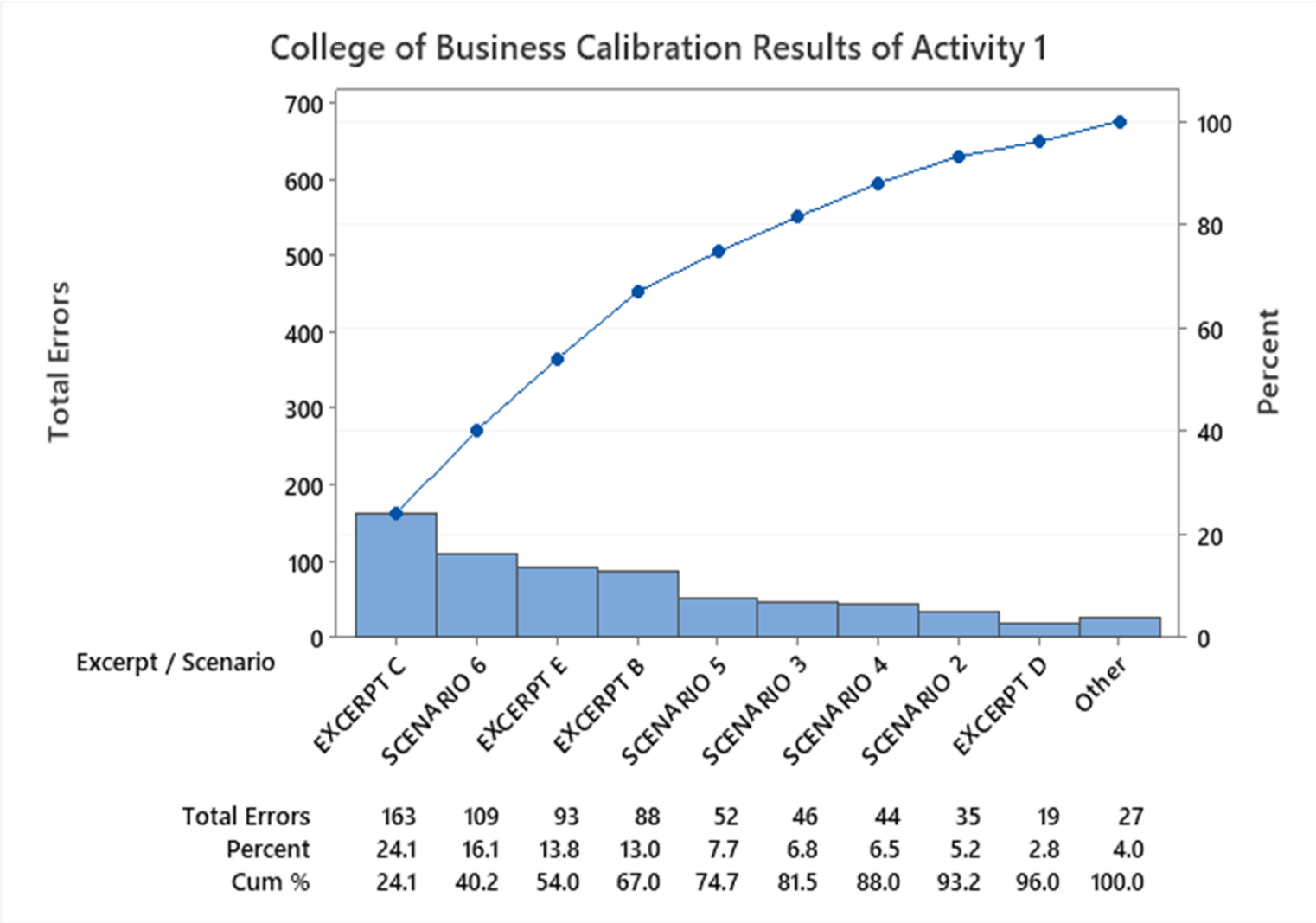
Revised Sources Calibration Process - College of Business

Tonia Howe July 2020



Pareto Chart – College of Business Problem Areas Identification

Activity 1 (Baseline Data)



Analysis:

1. Seven of the 11 scenarios reviewed resulted in more than 40 errors.
2. Three concepts were identified as being particularly problematic with more than 90 defects per item.
3. Six concepts tested resulted in 81.5% of the total errors.
4. 5.5% of evaluators inaccurately failed a perfectly formatted reference & citation.

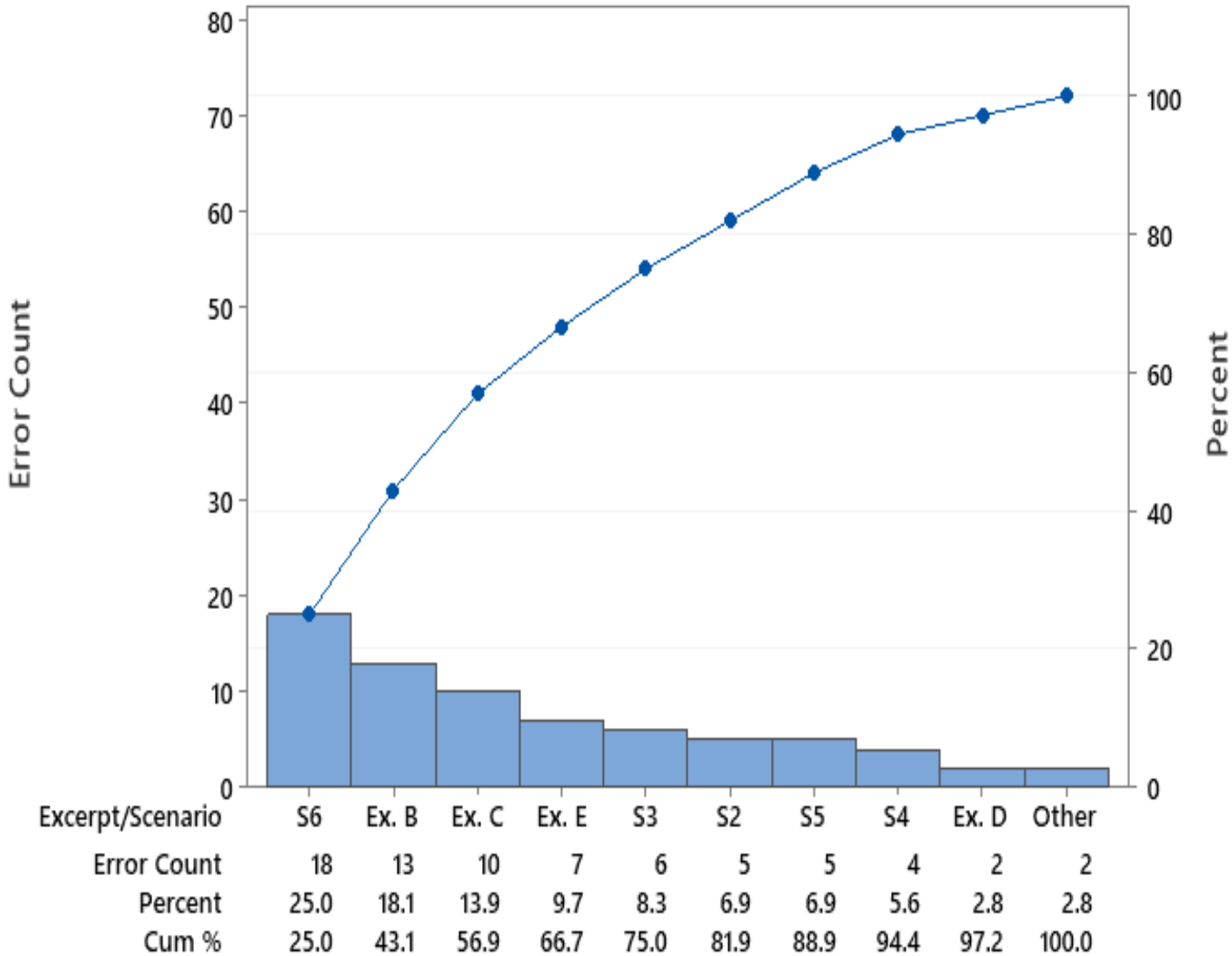
Result:

Given the results of the CoB Baseline Calibration Activity, we proceeded into Phase 1 of the Sources Calibration Initiative.

Pareto Chart – Phase I, People Leader Training

Activity 1

People Leader Calibration Results of Activity 1



Analysis:

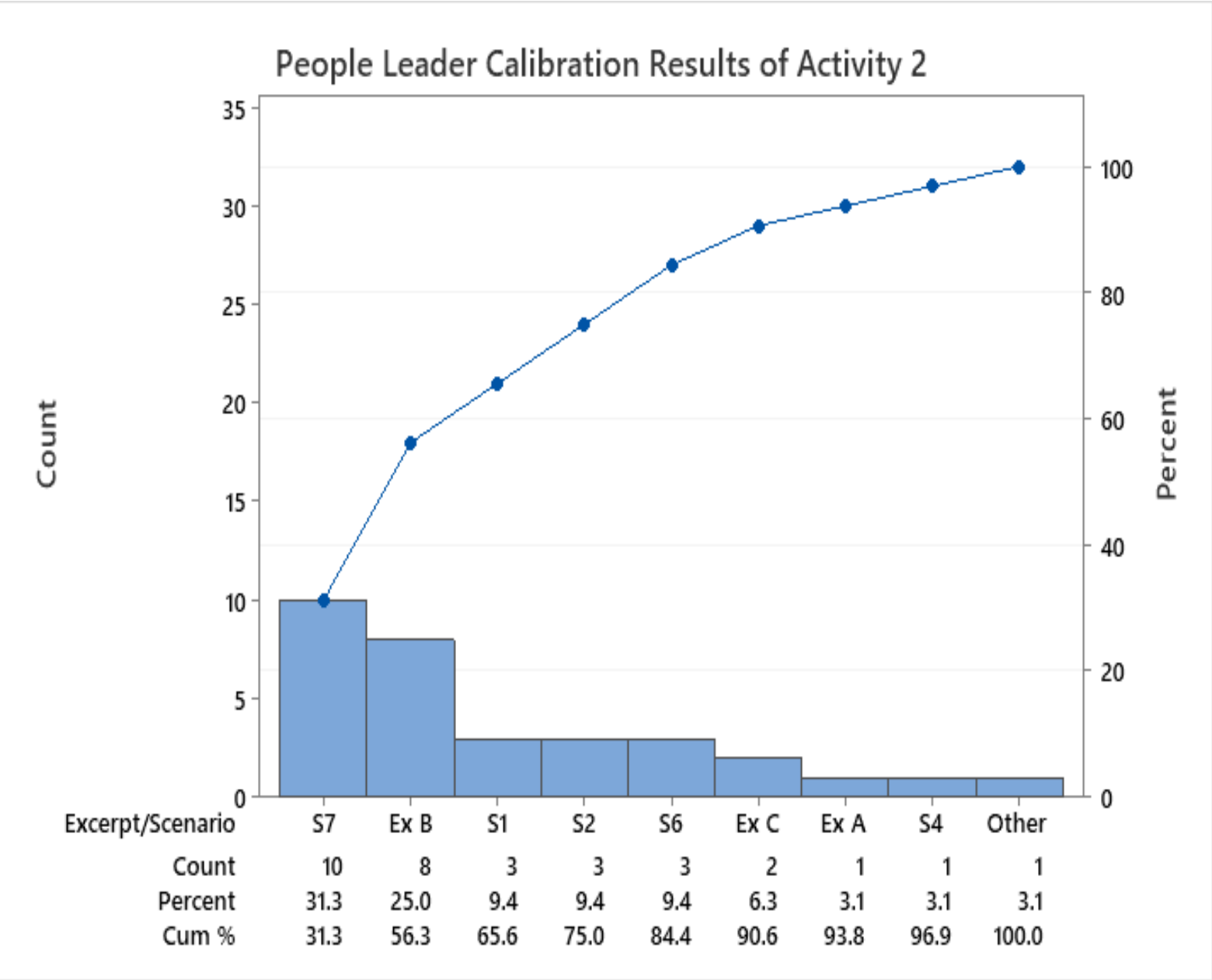
1. Scenario 6 (requiring a reference/citation for a textbook) had the most errors.
- 2.
3. Excerpt A (APA format) had the least number of errors.
4. Six concepts resulted in 81.9% of the total errors.
5. 3 questions were calibrated above 90%
6. Training should emphasize the 6 items resulting in 81.9% of the errors. Special focus will be given on the bottom 3 concepts.

Pareto Chart – Phase I, People Leader Training

Activity 2

Analysis:

- 1. How to handle a submission with an extra reference (S7) had the highest rate of error.
- 2. How to score a submission with 1 of 4 of the required elements missing (Ex B) also had a high error rate.
- 3. Four concepts resulted in 75% of the total errors.
- 4. 8 scenarios/excerpts were calibrated with more than 90% accuracy
- 5.
- 6. Reinforcement Training should emphasize the bottom 2 concepts



Analysis – Phase 1 Training, People Leaders

PL Stats Activity 1 Defects

Basic Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum		Q1	Median	Q3	Maximum
Defects	21	0	3.429	0.369	1.690	1.000	2.500	3.000	5.000	6.000	

PL ACTIVITY 2

PL Stats Activity 2 Defects

Basic Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum		Q1	Median	Q3	Maximum
Defects	30	0	1.067	0.179	0.980	0.000	0.000	1.000	2.000	3.000	

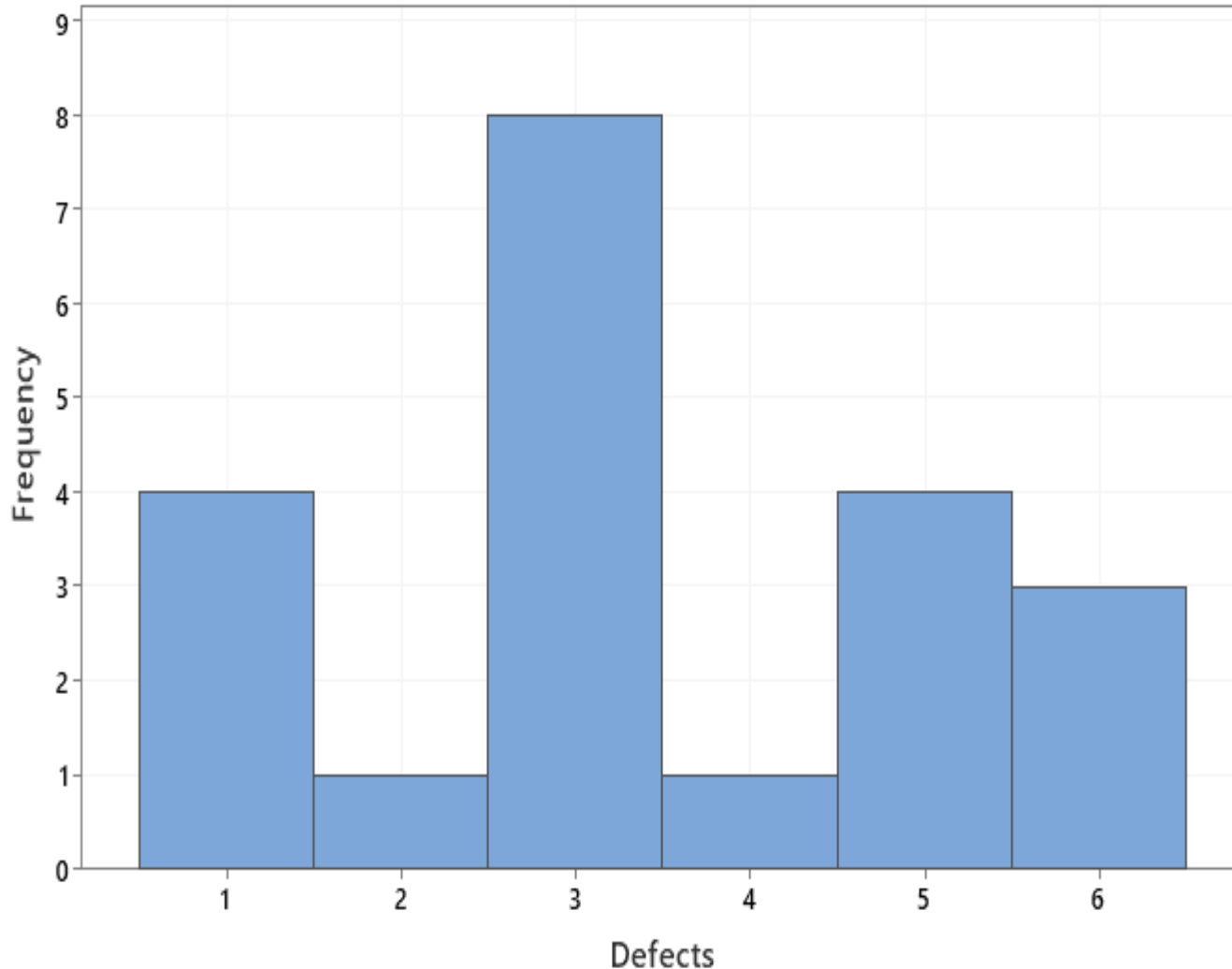
Takeaways:

- 1. Standardized Training & Calibration of the PLs reduced the StDev of the calibration activity by 42.01%.
- 2. Standardized Training & Calibration of the PLs reduced the Mean of the calibration activity by 68.88%.

Analysis – Phase 1 Training, CoB People Leaders

Activity 1 Histogram

People Leaders Activity 1 Histogram of Defects



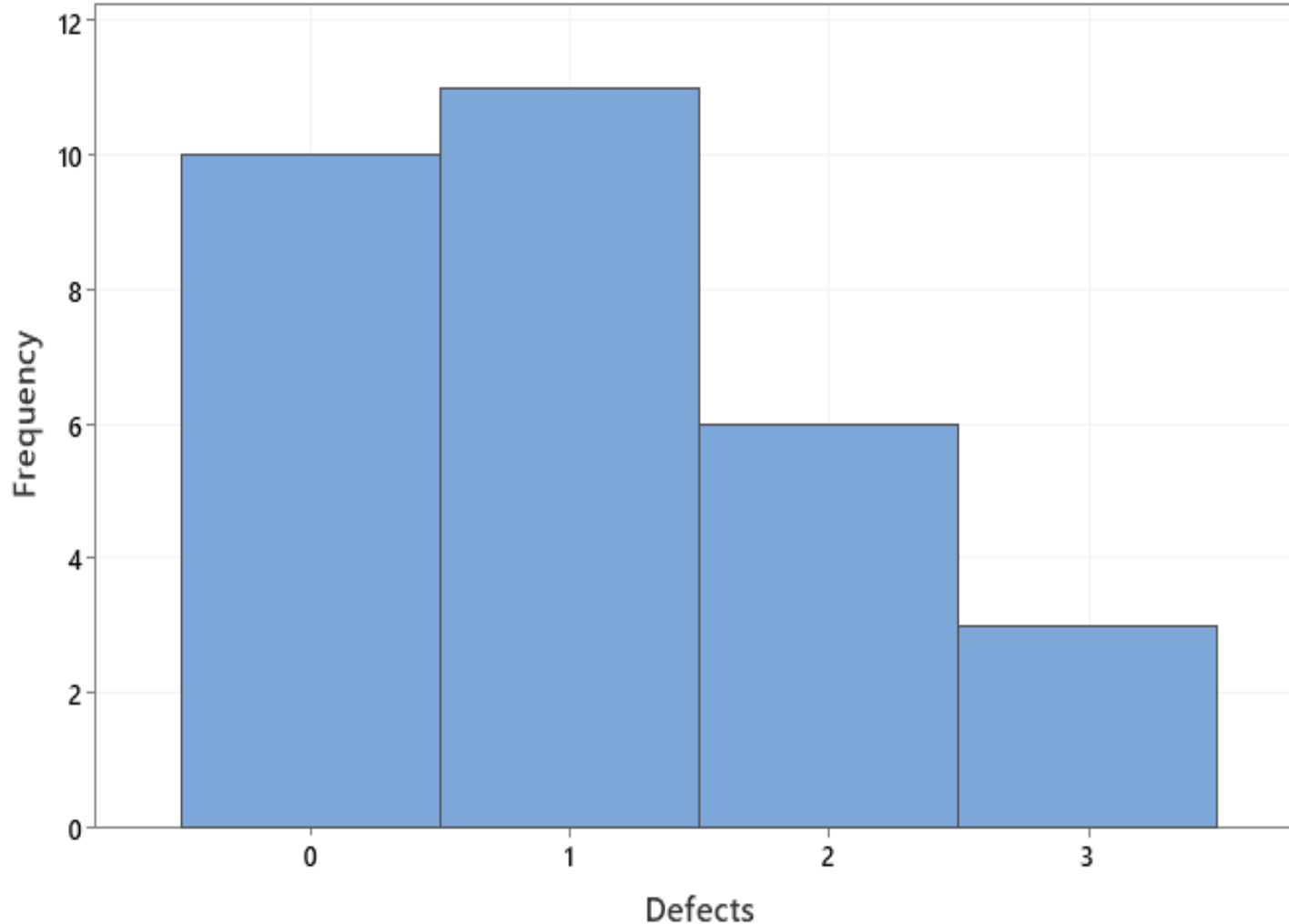
Analysis:

1. The distribution of the errors in the first calibration activity is normal.
2. The majority of the CoB People Leaders had three scoring errors in the first calibration activity.
- 3.
4. The range of errors by the PLs was between 1 and 6 errors.

Analysis – Phase 1 Training, CoB People Leaders

Activity 2 Histogram

People Leaders Activity 2 Histogram of Defects

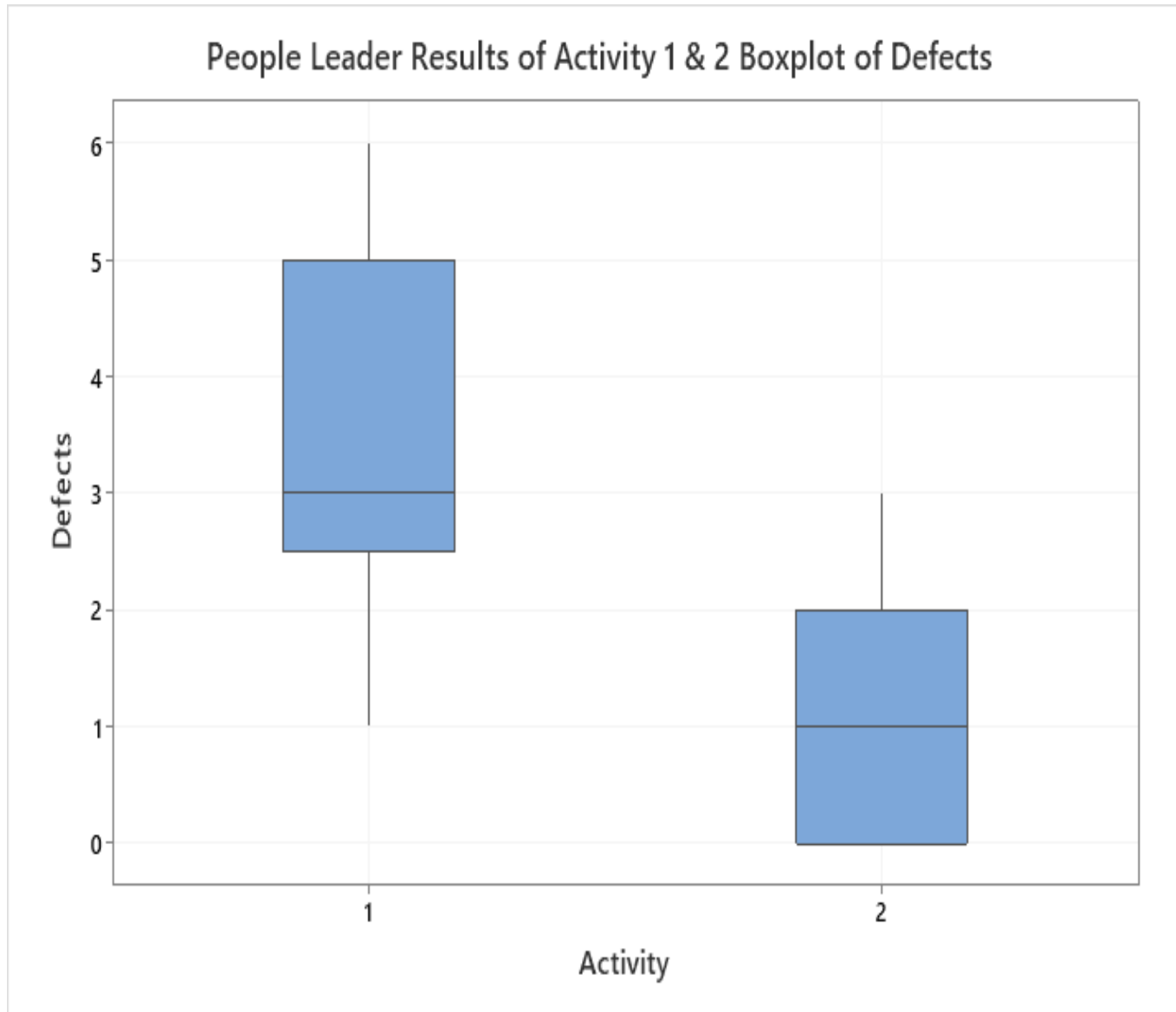


Analysis:

1. The new Standardized Training & Calibration process reduced the number of errors.
2. One error was the most common result in the 2nd calibration activity.
3. The range of errors per person was reduced to a range of 0 – 3.
4. The new distribution of errors for activity 2 resulted in skewed right distribution. This is due to the fact that the error rate on the calibration activity can not be negative.

Analysis – Phase 1 Training, CoB People Leaders

Boxplot Comparison of activities

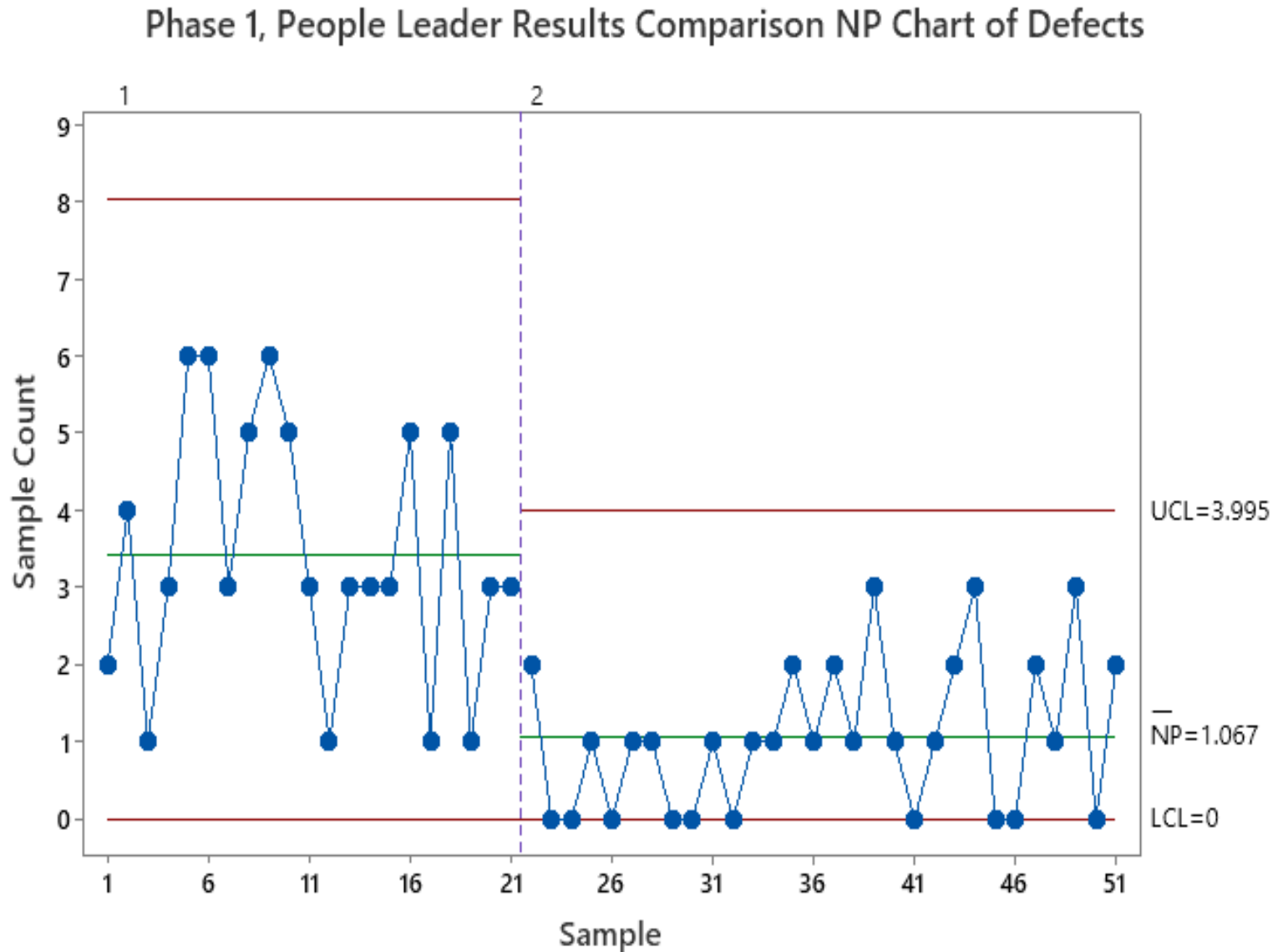


Analysis:

1. The PL Team averaged 3 defects in Activity 1. 50% of the team had between 2.5 & 5 defects.
2. Also, in activity 1, the range of defects was 1 - 6.
3. After going through the new Standardized Training & Calibration process, the average number of defects was reduced to 1 error.
4. The team also demonstrated a reduction in the range of defects (0 – 3) with 75% of the team incurring 2 or less defects after going through the new process.

Analysis – Phase 1 Training, CoB People Leaders

np chart of Defects comparing Activity 1 & 2



Analysis:

1. The process was in a state of statistical control for both activities.
2. Standardized training & calibration reduced the UCL from 8.037 to 3.995, which is a reduction of 50.29%.
3. Standardized training & calibration reduced the mean defect rate from 3.429 to 1.067, which is 68.88%.

Analysis – Phase 1 Training, CoB People Leaders

Hypothesis Testing

Descriptive Statistics

N	Mean	SE Mean	95% Upper Bound for μ
10	1.067	0.310	1.577

μ : population mean of Sample
Known standard deviation = 0.98

Test

Null hypothesis	$H_0: \mu = 3.429$
Alternative hypothesis	$H_1: \mu < 3.429$

Z-Value	P-Value
-7.62	0.000

- The p-value for the PL results is less than 0.05
- If the *p* is low, null must go
- Hypothesis testing proves that the new standardized training & calibration process improvements are not random.
- We can conclude with 95% confidence that the new Standardized Training & Calibration process for Sources had a **statistically significant impact** on the mean defect rate for People Leaders. This indicates there will be less variation across the CoB team guidance on how to evaluate the Sources aspect.

Analysis – College of Business Activity 1

- 1. Baseline data for the College of Business will be compared with the results of Phase 2.
- 2. During Phase 2 we will train approximately 200 evaluators using the new Standardized Training & Calibration process.

Activity 1 – Entire College of Business Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Defects	201	0	3.637	0.118	1.674	0.000	2.000	3.000	5.000	8.000

Activity 1 - CoB by Team Statistics

Variable	Team	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Defects	Acct/ITM	16	0	3.250	0.359	1.438	1.000	2.250	3.000	4.750	5.000
	Marketing		48	0	3.333	0.213	1.478	0.000	2.000	3.000	4.000
	MBA	91	0	3.626	0.185	1.768	0.000	3.000	3.000	5.000	8.000
	UG Core	46	0	4.109	0.249	1.690	1.000	2.000	4.500	6.000	6.000

Analysis – College of Business

Phase 2 Evaluator Training

Activity 2

Activity 2 – Entire College of Business

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Defects	190	0	1.526	0.101	1.394	0.000	0.000	1.000	2.000	6.000

Activity 2 - CoB by Team

Statistics

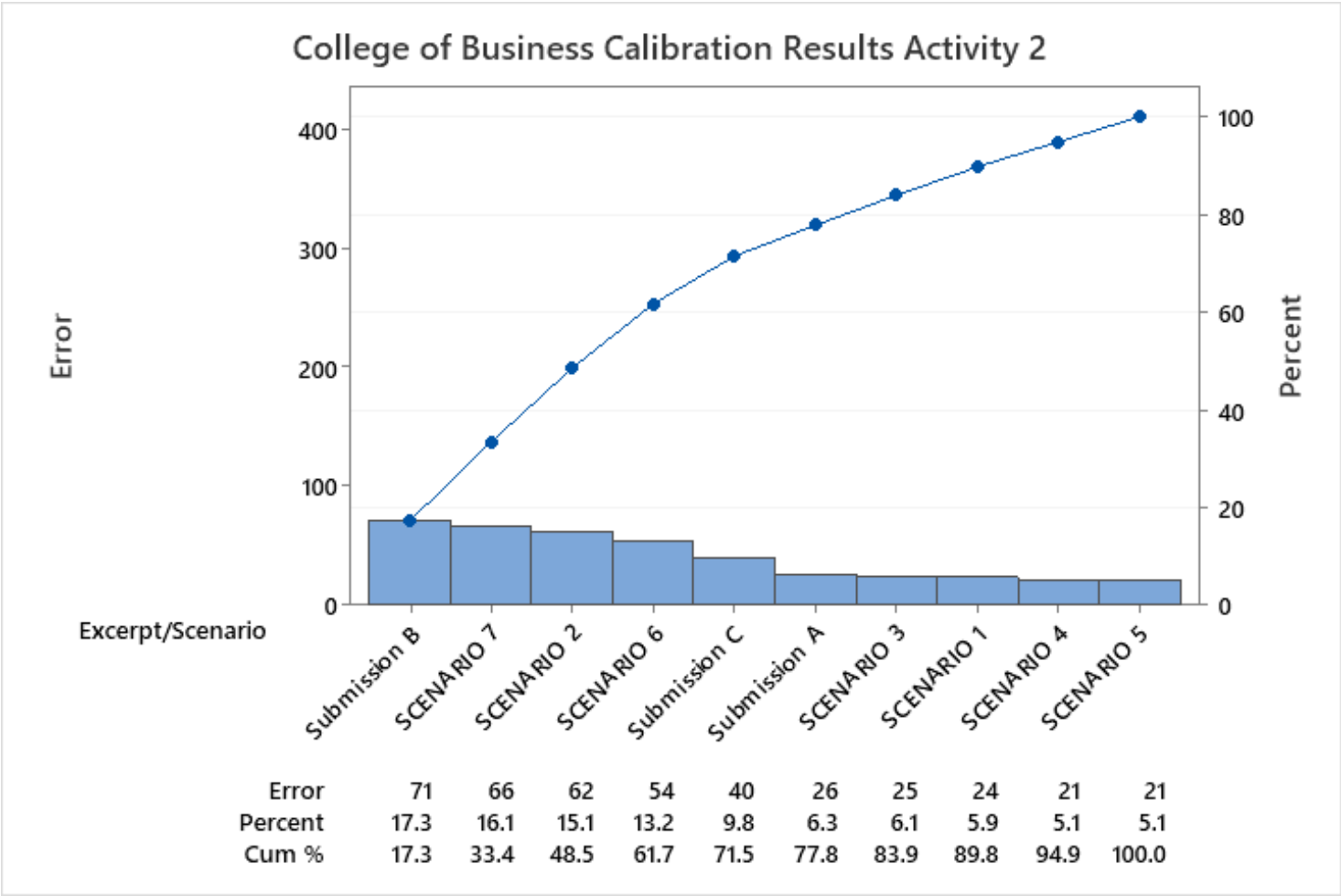
Variable	Team	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Defects	Acct/ITM	16	0	1.063	0.193	0.772	0.000	0.250	1.000	2.000	2.000
	Marketing	59	0	1.424	0.166	1.276	0.000	0.000	1.000	2.000	5.000
	MBA	83	0	1.723	0.173	1.572	0.000	1.000	1.000	3.000	6.000
	UG Core	32	0	1.438	0.233	1.318	0.000	0.000	1.000	2.750	5.000

Analysis:

1. The standardized training and calibration for the College of Business improved the standard deviation by 16.73% in Activity 2.
2. The standardized training and calibration for the College of Business improved the mean by 58.04%.

Analysis – Phase 2 Training, CoB Evaluators

Activity 2 Pareto Chart

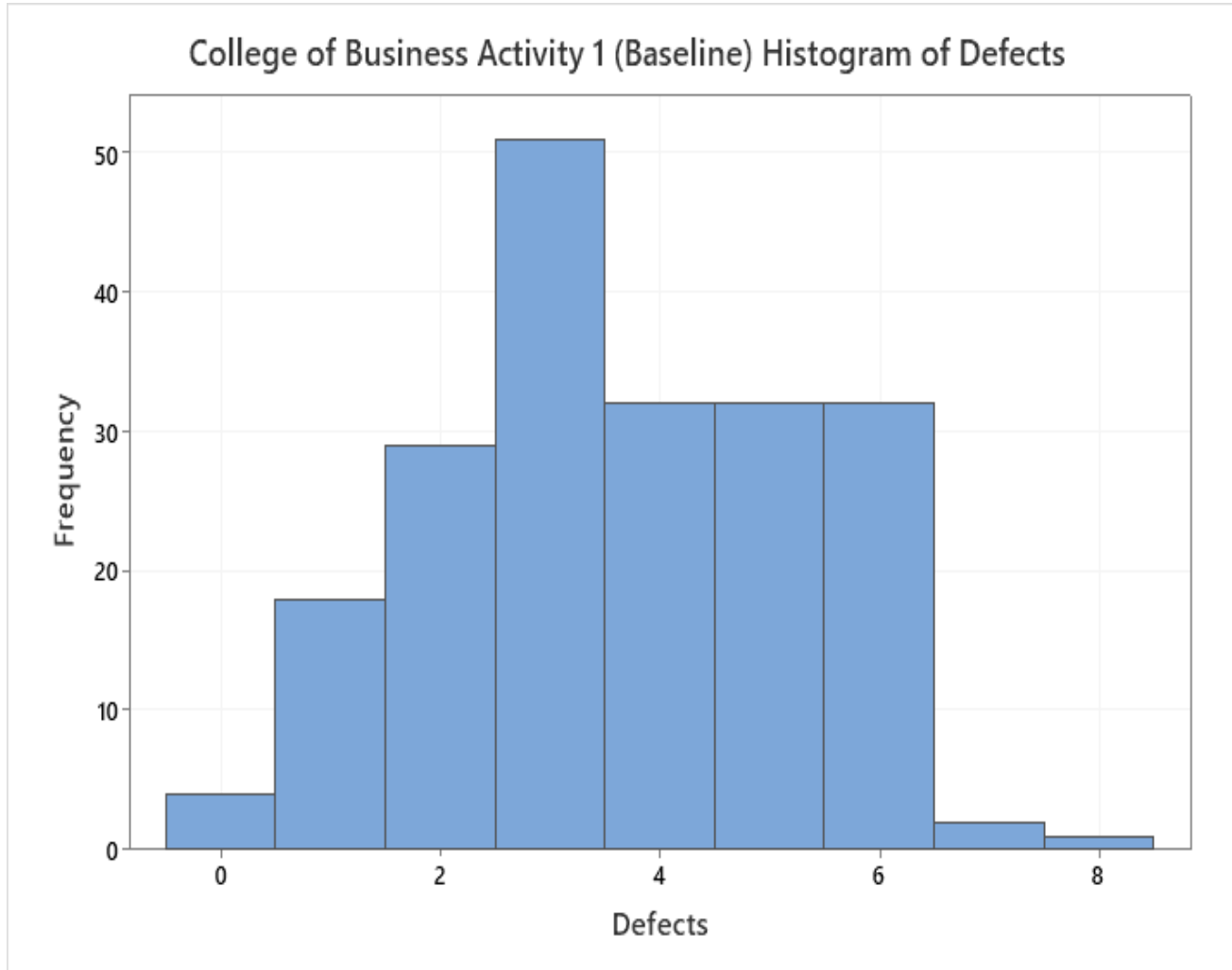


Analysis:

1. Six of the 10 concepts tested in the calibration activity resulted in 77.8% of the total errors.
2. Three concepts were identified as being particularly problematic resulting in close to 50% of the total errors.
3. The August activity will focus on the six concepts that resulted in 77.8% of the total errors.

Analysis – Phase 2 Training, CoB Evaluators

Activity 1 Histogram

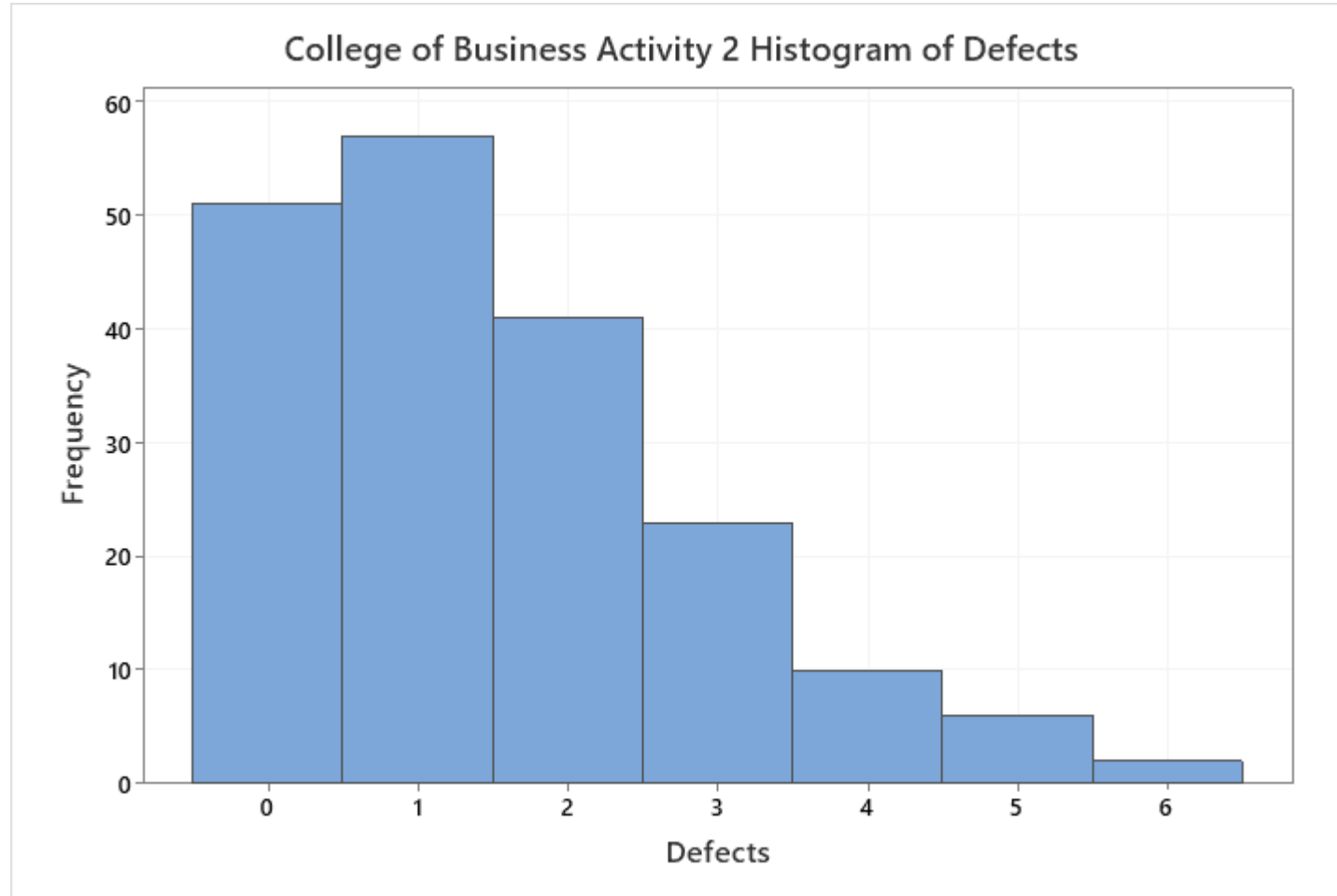


Analysis:

1. Peaks – Three errors was the most common number of errors in activity 1.
2. Spread – The range of defects was 0 - 8.
3. Symmetry – The histogram shows a normal distribution with minor skew to the right, indicating a relatively high number of errors in the first calibration activity.

Analysis – Phase 2 Training, CoB Evaluators

Activity 2 Histogram



Analysis:

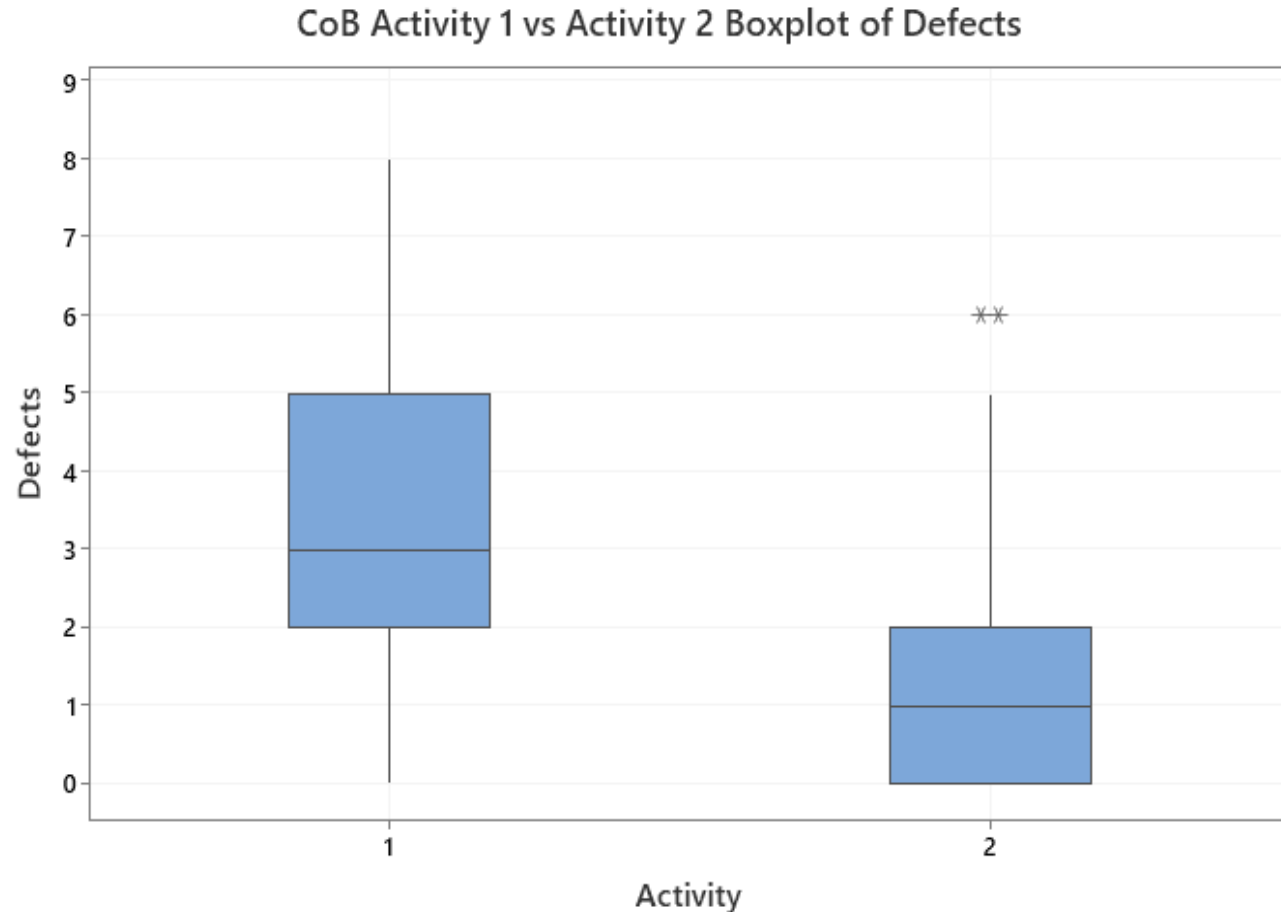
1. The new Standardized Training & Calibration process reduced the number of errors across the CoB.
2. A total of 1 error was the most common result in the second activity.
3. The number of errors per person was reduced to a range of 0 – 6.
4. The distribution of errors for activity 2 is skewed right. The distribution indicates an increased understanding on how to handle various situations related to Sources by the CoB Evaluators.

Analysis – Phase 2 Training, Evaluators

Boxplot Comparison of activities

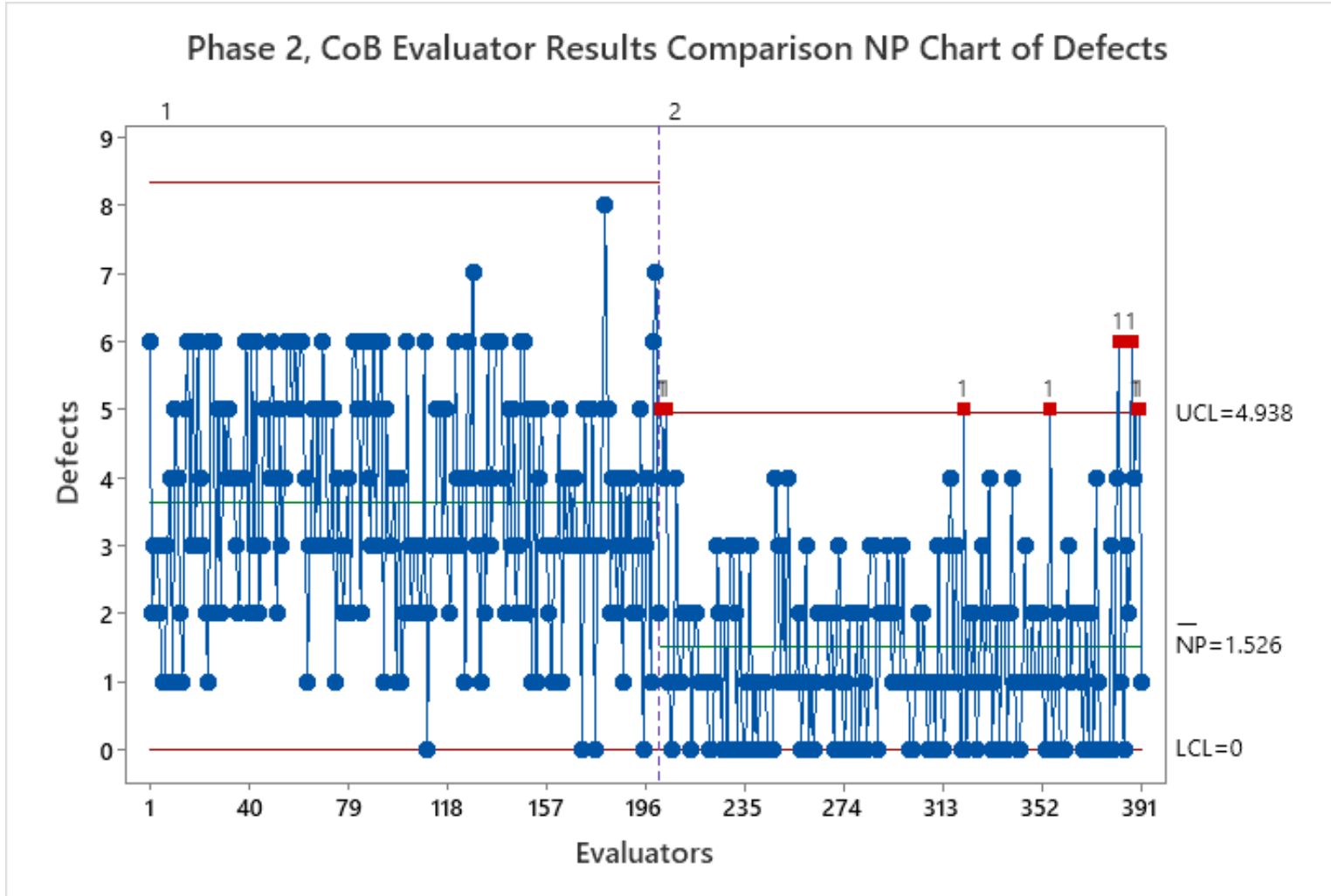
Analysis:

1. The CoB averaged 3 defects in Activity 1. 50% of the team had between 2 & 5 defects.
2. The range of errors in the baseline activity was 0 - 8.
3. After going through the new Standardized Training & Calibration process, the median number of errors was reduced to 1.
4. The team also demonstrated a reduction in the range of errors (0 – 5) with 75% of the team incurring 2 or less errors after going through the new training process.



Analysis – Phase 2 Training, CoB Evaluators

np chart of Defects comparing Activity 1 & 2



Analysis:

1. The process was in a state of statistical control during the first activity. Activity 2 had 8 points that fell outside the upper control limit. These outliers warrant additional review.
2. Standardized training & calibration reduced the UCL from 8.318 to 4.938, which is a reduction of 40.63%.
3. Standardized training & calibration reduced the average error rate from 3.637 to 1.526, which is 58.04%.

Analysis – Phase 2 Training, CoB Evaluators

Hypothesis Testing

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
190	1.526	1.394	0.121	(1.288, 1.764)

μ : population mean of Defects
Known standard deviation = 1.674

Test

Null hypothesis $H_0: \mu = 3.637$
Alternative hypothesis $H_1: \mu \neq 3.637$

Z-Value	P-Value
-17.38	0.000

- The p-value for PL results is less than 0.05
- If the *p* is low, null must go
- Hypothesis testing proves that the new standardized training & calibration process improvements are not random.
- We can conclude with 95% confidence that the new Standardized Training & Calibration process for Sources had a **statistically significant impact** on the mean error rate of the Evaluators in the CoB. This indicates there will be less variation in the evaluation of the Sources aspect across CoB Teams.



WGU®

www.wgu.edu

