Air Force Institute of Technology

AFIT Scholar

Theses and Dissertations

Student Graduate Works

6-2022

Evaluating a Statistical-Based Assessment Tool for Stratifying Risk among U.S. Air Force Organizations

Tiffany A. Low

Follow this and additional works at: https://scholar.afit.edu/etd



Part of the Applied Statistics Commons, and the Risk Analysis Commons

Recommended Citation

Low, Tiffany A., "Evaluating a Statistical-Based Assessment Tool for Stratifying Risk among U.S. Air Force Organizations" (2022). Theses and Dissertations. 5489. https://scholar.afit.edu/etd/5489

This Thesis is brought to you for free and open access by the Student Graduate Works at AFIT Scholar. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of AFIT Scholar. For more information, please contact richard.mansfield@afit.edu.



THESIS

Tiffany A. Low, Major, USAF

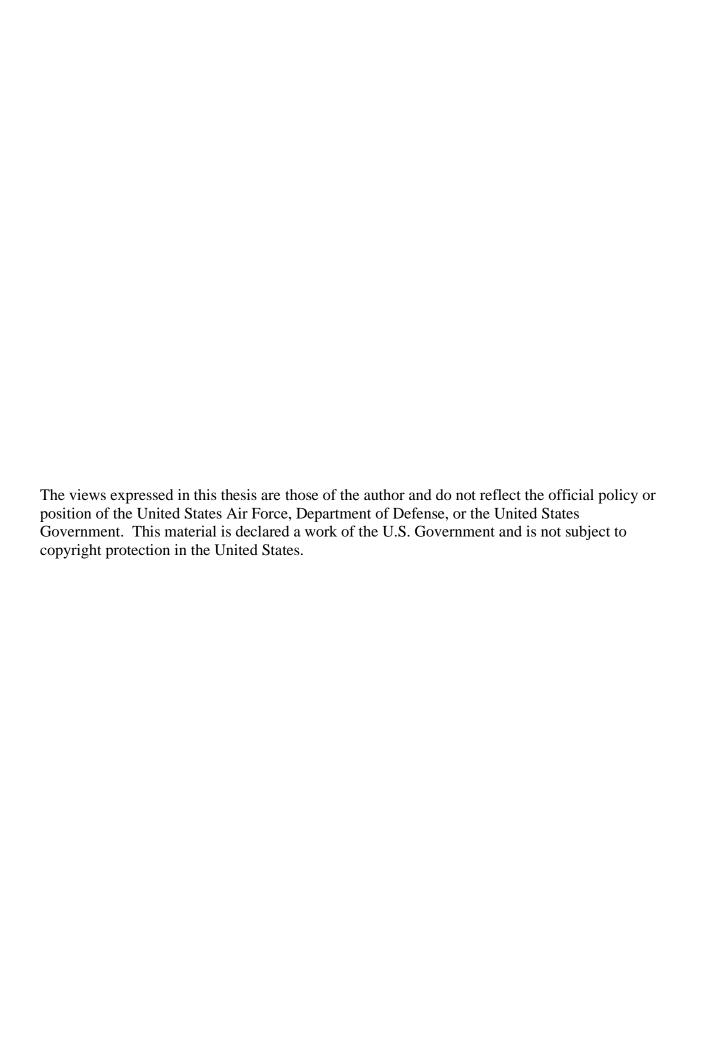
AFIT-ENV-MS-22-J-066

DEPARTMENT OF THE AIR FORCE AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A.APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.



THESIS

Presented to the Faculty

Department of Mathematics and Statistics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Program Management

Tiffany A. Low, BS

Major, USAF

June 2022

DISTRIBUTION STATEMENT A.

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Tiffany A. Low, BS

Major, USAF

Committee Membership:

Dr. Edward D. White Chair

Dr. John J. Elshaw Member

Lt Col Clay M. Koschnick, PhD Member

Abstract

The Air Force Inspection System is a proponent of utilizing a risk-based sampling strategy (RBSS) for conducting inspections from major command levels down to the unit level. The strategy identifies areas deemed most important or risky by commanders and prioritizes them accordingly for an independent assessment by the Inspector General. While Air Force regulation specifies the need to use a RBSS for inspection, the implementation process is delegated to individual commands and, subsequently, wings. The 23rd Wing, the sponsor for this research, directed us to analyze a RBSS tool highlighted as an example from which to adopt for those units within Air Combat Command, the major command for the 23rd. Our analysis entailed both descriptive and inferential measures. The results identified some potential shortfalls for which solutions were proposed. The recommended measures include using a median-based metric instead of a mean-based metric to score risk for organizations, a 3-point Likert scale to evaluate criteria, as well as a scoring system for dichotomous criteria when mixing with either a 3-, 4-, or 5-point Likert scale criteria.

Acknowledgments

First and foremost, I would like to express my deepest appreciation to my advisor and committee chair, Dr. Edward D. White, without whom this thesis would not be possible. Dr. White's patience, guidance, and feedback has been instrumental throughout my research journey. Special thanks to my committee members, Dr. John J. Elshaw and Lt Col Clay M. Koschnick, for finetuning my thesis with their knowledge and expertise. Additionally, I would like to acknowledge my research sponsor, the 23rd Wing, as well as various members of the Inspector General community to include Air Combat Command, 129th Rescue Wing, 163rd Attack Wing, and 366th Fighter Wing for providing crucial information about my research topic. Finally, I would like to thank my friends, family, and running team for their continual encouragement and support.

Tiffany A. Low

Table of Contents

		Page
Abs	stract	iv
List	t of Tables	xi
I. I	Introduction	1
	Problem Statement	1
	Background	1
	Research Questions and Data	2
	Methodology	2
	Thesis Organization	3
II. I	Literature Review	4
	Chapter Overview	4
	Relevant History	4
	Risk-Based Sampling Strategy	8
	23 rd Wing	9
	129th Rescue Wing	10
	163 rd Attack Wing	14
	366th Fighter Wing	21
	Summary	26
III.	Methodology	27
	Chapter Overview	27
	Risk-Based Sampling Strategy Instruments and Components	27
	Assumptions and Limitations	28
	Instruments Used for Analysis	28

Determining Values for Simulation	29
Number of Units	29
Number of Criteria	29
Type of Scale Used	30
Binomial Distribution	35
Data Simulation	38
Standardized Scores.	40
Percent Contribution to the Response Variable	41
Dichotomous Variable	42
Summary	43
IV. Analysis and Results	45
Chapter Overview	45
Standardized Scores	45
Dichotomous Variable	57
Summary	62
V. Conclusions and Recommendations	
Chapter Overview	63
Investigative Questions Answered	63
Conclusions of Research	64
Significance of Research	67
Recommendations for Action	68
Recommendations for Future Research	68
Summary	69
Appendix A. 129th Rescue Wing Risk-Based Sampling Strategy	70

Appendix B. 129 th Rescue Wing Risk-Based Sampling Strategy Criteria Stat	tistics71
Appendix C. 163 rd Attack Wing Risk-Based Sampling Strategy Criteria Stat	istics77
Appendix D. Simulation Graphs for Different Combinations of Likert Scale,	, Criteria, Skewness
Level and Units	81
Bibliography	126

List of Figures

Page
Figure 1. 23 rd Wing Risk-Based Sampling Strategy
Figure 2. 129 th Rescue Wing Risk-Based Sampling Strategy (Truncated)
Figure 3. 129 th Rescue Wing Risk-Based Sampling Strategy Risk Coding
Figure 4. 129 th Rescue Wing Risk-Based Sampling Strategy Risk Coding Under a Normal Curve
Figure 5. 163 rd Attack Wing Risk-Based Sampling Strategy (2020)
Figure 6. 163 rd Attack Wing Risk-Based Sampling Strategy Probability Coding
Figure 7. 163 rd Attack Wing Risk-Based Sampling Strategy Severity Coding
Figure 8. 163 rd Attack Wing Risk Assessment Matrix
Figure 9. 163 rd Attack Wing Risk-Based Sampling Strategy (2018)
Figure 10. 366 th Fighter Wing Risk-Based Sampling Strategy (draft)
Figure 11. 366 th Fighter Wing Risk Score Formula
Figure 12. Theoretical and Actual Binomial Distribution for 3-, 4- and 5-point Likert Scales 37
Figure 13. Data Dominance Comparison Between S-score and M-score
Figure 14. Difference Between Tail Values for S-score and M-score for 3-point Likert Scale 51
Figure 15. Difference Between Tail Values for S-score and M-score for 4-point Likert Scale 51
Figure 16. Difference Between Tail Values for S-score and M-score for 5-point Likert Scale 52
Figure 17. Pareto Analysis Effects of Inputs on S-score Tail Difference
Figure 18. Difference Between Median and Mean for 3-point Likert Scale
Figure 19. Difference Between Median and Mean for 4-point Likert Scale
Figure 20. Difference Between Median and Mean for 5-point Likert Scale 57

Figure 21.	Contribution of Different Dichotomous	Value Ranges for 3-pt Likert Scale 6
Figure 22.	Contribution of Different Dichotomous	Value Ranges for 4-pt Likert Scale 6
Figure 23.	Contribution of Different Dichotomous	Value Ranges for 5-pt Likert Scale 6

List of Tables

Page
Table 1. Waiver Levels
Table 2. 129 th Rescue Wing Risk-Based Sampling Strategy Criteria
Table 3. 163 rd Attack Wing Risk-Based Sampling Strategy Criteria (2020)
Table 4. 163 rd Attack Wing Risk Based Sampling Strategy Criteria (2018)
Table 5. 129 th Rescue Wing Risk-Based Sampling Strategy Metrics
Table 6. 129 th Rescue Wing Risk-Based Sampling Strategy Scaling Criteria Highest Score 32
Table 7. 129 th Rescue Wing Risk-Based Sampling Strategy Scaling Criteria Groups
Table 8. 163 rd Attack Wing Risk-Based Sampling Strategy Metrics
Table 9. 163 rd Attack Wing Risk-Based Sampling Strategy Scaling Criteria Highest Score 34
Table 10. 163 rd Attack Wing Risk-Based Sampling Strategy Scaling Criteria Groups
Table 11. Comparison of S-score and M-score for 3-Point Likert Scale
Table 12. Comparison of S-score and M-score for 4-Point Likert Scale
Table 13. Comparison of S-score and M-score for 5-Point Likert Scale
Table 14. Outliers for M-score and S-score Tail Differences
Table 15. Parameter Estimates for S-score Tail Difference
Table 16. Parameter Estimates for S-score Tail Difference (Rerun)
Table 17. Parameter Estimates for S-score Tail Difference (Sorted Standardized Beta)
Table 18. Determining Optimal Range for a Dichotomous Variable by Likert Scale 59
Table 19. Optimal Range for a Dichotomous Variable

I. Introduction

Problem Statement

According to Air Force Instruction 90-201 (2018), the Air Force Inspection System uses a risk-based sampling strategy for conducting inspections from major command levels down to the unit level. While Air Force regulation specifies the need to use a risk-based sampling strategy (RBSS) for inspection, the implementation process is delegated to individual wings. Without a standardized template from which different units build their tailored strategy, the tools in use can differ drastically across units. Currently the 23rd Wing, the sponsor for this research, uses a risk matrix or heat map to determine which programs or organizational units the wing commander deems as riskiest. However, another organizational unit within Air Combat Command (ACC), 129th Rescue Wing, has a tool that ACC distributes as an example. Our research examines the sample tool currently used and its efficacy.

Background

According to the 23rd Wing Director of Inspections, ACC does not have a unified, quantitative method for implementing their RBSS. Risk-based sampling strategy is defined as "a methodology employed to inspect areas deemed most important by commanders and functional area managers requiring an independent assessment by the Inspector General (IG)" (AFI 90-201, 2018). In other words, it is a decision analysis approach that prioritizes which programs and processes get inspected by the local Inspector General based on its risk to the mission. Since some ACC bases do not have a defined or acceptable RBSS process in place, Headquarters ACC shared an example used by the 129th Rescue Wing for units under its command.

Research Questions and Data

While different units have the flexibility to create their own RBSS that may meet Air Force regulation on the surface, we need to assess whether current tools in use truly meet the intent of a RBSS. We analyze the existing tools, identify shortfalls, and remedy any issues. Our research seeks to answer the following questions:

- 1. What are the current processes different inspection teams use for their risk-based sampling strategy?
- 2. What is the validity of the current risk-based sampling strategies?
- 3. What may be done to correct or improve the current risk-based sampling strategies?
- 4. What may be a better tool to use for a risk-based sampling strategy?
- 5. How can we create a product that is simple to use and explain to users?

We are given two sets of data: one created by the 23rd Wing and one used by the 129th Rescue Wing (RQW) as the ACC RBSS example. Through information tracing, we discovered that the 129th RQW modified a spreadsheet from the 163rd Attack Wing (ATKW), which also modified a tool received from ACC IG believed to originate from the 366th Fighter Wing (FW). We describe how each risk-based sampling strategy is currently employed and identify any shortfalls of the current tools in use, focusing primarily on the 129th RQW RBSS tool since it is an ACC model.

Methodology

First, we examine our data, which are the tools the 23rd Wing currently uses and the one used by the 129th Rescue Wing that ACC views as a benchmark process. Since the 129th RQW is not the originator of their tool, we also consider the preceding tools used by the 163rd ATKW and possibly the 366th FW which serves as the foundational framework for the 129th RQW tool. We

then identify any issues with the current tool and propose adjustments. We explore patterns in actual RBSS data to determine what variables and limits to use when generating data simulations. Through simulation, we highlight the different results using existing tools versus the modified tools. Comparing the results determine which tool better meets the intention of a risk-based sampling strategy.

Thesis Organization

This thesis is organized into five chapters. Chapter 1 introduces the research topic and questions. Chapter 2 reviews the history of the IG process, background information on RBSS including examples of RBSSs in use. Chapter 3 discusses the methodology and data used to investigate our research questions. Chapter 4 presents the analysis and results. Chapter 5 summarizes and discusses conclusions and presents possible recommendations for future research.

II. Literature Review

Chapter Overview

This chapter provides relevant history of the Air Force inspection system, its revision resulting in the need for a risk-based sampling strategy (RBSS), the existing RBSS tool used by the 23rd Wing and the current RBSS tool serving as Air Combat Command's example. This chapter defines what a risk-based sampling strategy is and further highlights current processes used to meet its intent.

Relevant History

When the Air Force was established in 1947, there were six types of inspections; however, by 2010, the Air Force, Department of Defense and other government agencies were conducting over 97 different types of inspections, assessments, and evaluations, leaving a commander less than 50% of days in a year to focus solely on the mission (Camm et al., 2010).

In 2010, the Office of the Secretary of the Air Force, Inspector General created an Inspection System Improvement Tiger Team tasked to improve inspection policy and to reduce the burden on inspected units while improving the quality of relevant information generated for Air Force leadership (Camm et al., 2010). Factors such as budget constraints, reduced staffing, and increased taskings revealed that the old inspection system was unsustainable (Rogers, 2012). In preparation for an inspection, some units perceived the need to "paint their grass green," shifting valuable time and resources from overall mission readiness to focus on compliance with inspection standards (Hyde, 2013).

A new inspection system was needed to shift the focus back to its fundamental purpose, which was to strengthen the command function, improve command effectiveness, promote military discipline, improve unit performance, and management excellence in units and staffs

throughout the chain of command (Rogers, 2012). The very core of the change to the new inspection system was shifting responsibility for inspection from external functional leaders to wing commanders. Wing commanders are responsible for executing their missions; functional leaders are responsible for policy and oversight of their functional areas (Mueller, 2012). Functional inspections now fall under the wing commander's inspection programs, and major command Inspector Generals (MAJCOM IGs) validate and verify compliance (Mueller, 2012). In 2013, the Secretary of the Air Force directed the implementation of the new Air Force Inspection System; Air Force Policy Directive 90-2 as well as Air Force Instruction 90-201 (previously named Inspector General Activities) were revised accordingly.

The revised Air Force Inspection System consists of different tiers of inspection to include management inspection, Unit Effectiveness Inspection (UEI), commander's inspection program and unit self-assessment programs. Instead of relying solely on external inspections, the foundation of the new inspection system relies on internal self-assessments beginning at the unit level. Wing commanders have the flexibility to tailor their Commander's Inspection Program (CCIP), focusing resources in areas deemed most important to them. The wing Inspector General (IG) executes the CCIP and prioritizes those areas for inspections (SAF/IGI, 2016). Changes to the inspection process allow operational wings to focus on daily missions, make continuous improvement, while maintaining a stable state of readiness without increasing workload to facilitate large-scale external inspections.

The intent behind a continual new inspection process is to eliminate inspection preparation practice of painting the grass green since wing performance is now a photo album that captures a unit's performance over the inspection period versus a snapshot in time (Mueller, 2012). The inspection period or UEI cycle is every 24 to 36 months for each Regular Air

Force/Air Force Reserve Command Wing and 48 to 60 months for each Air National Guard Wing. The UEI is an external continual evaluation of Wing performance conducted by MAJCOM IG and/or the Air Force Inspection Agency teams based on four major graded areas:

(a) executing the mission, (b) improving the unit, (c) managing resources, and (d) leading people. Instead of focusing solely on compliance, these four areas give leaders a more holistic view of true capability and inform them where to focus limited resources (Mueller, 2012).

The UEI incorporates elements of compliance and readiness to create a comprehensive assessment of unit effectiveness using a 4-tier rating system: highly effective, effective, marginally effective, and ineffective. An ineffective rating will require a reevaluation within 180 calendar days and can only receive a highest rating of effective instead of the original highest rating of highly effective (AFI 90-201, 2018).

In continuous preparation for a UEI, each wing has its own CCIP executed by its IG. The wing IG's annual inspection plan must include by-law programs and exercise requirements, which are mandatory CCIP inspection areas (AFI 90-201, 2018). Program inspections, also known as horizontal inspections, assess a program's health across a wing. If there is a pattern of multiple organizational units performing poorly in a program such as the unit fitness, then a nonmandatory horizontal inspection of all organizational units' fitness programs may be beneficial and appropriate. On the other hand, unit inspections or vertical inspections assess a specific organizational unit's health. Every organizational unit under the wing commander's purview is subject to at least one vertical inspection during the UEI cycle although this requirement can be waived at the wing commander level or T-3 (see Table 1).

A waiver is a commander's tool to "enhance mission effectiveness at all levels, while preserving resources and safeguarding health and welfare" (AFI 33-360, 2015). An approved

waiver indicates the commander has accepted all the potential risks created by noncompliance (Stuart, 2019). According to AFI 33-360 (2015), there are three specific circumstances for a waiver:

- 1. The cost of compliance creates unacceptable risk to a higher priority task.
- 2. The expected cost of compliance outweighs the benefit.
- 3. Personnel cannot comply with the requirement due to a lack of resources.

 To submit a waiver for a vertical inspection (Stuart, 2019):
- 1. Identify an opportunity to improve operations by simplifying or removing a requirement.
- 2. Identify the governing regulation and the applicable paragraphs for the appropriate waiver level.

Table 1. Waiver Levels

Tier Number	Waiver Authority
	External to the Air Force (e.g. Congress,
T-0	White House)
	MAJCOM Commander who can delegate this
T-1	authority no lower than the MAJCOM Director
	MAJCOM Commander who can delegate this
	authority no lower than the first General
T-2	Officer in the chain of command
	Wing Commander who can delegate this
	authority no lower than the Squadron
T-3	Commander or equivalent level

- 3. Determine who can approve the waiver (Table 1).
- 4. Route AF Form 679 and accompanying electronic Staff Summary Sheet for approval.

 AF Form 679 specifies commander seeking waiver, waiver authority, governing publication, requirement requested to be waived, rationale for the waiver, the duration of the waiver, any applicable risk mitigation measure being taken and the impact of disapproval. However, even if a

waiver is approved as "permanent," a commander can only approve a waiver for the length of his/her tour. The waiver expires 90 calendar days after the commander's change of command unless the new commander chooses to renew it. During the waiver period, the requesting commander/director is required to implement risk controls, work toward compliance (if applicable), continuously reevaluate risk and adjust accordingly.

Risk-Based Sampling Strategy

One risk mitigation measure that can be used to justify a vertical inspection waiver is using a risk-based sampling strategy (RBSS). A risk-based sampling strategy is defined as "a methodology employed to inspect areas deemed most important by commanders and functional area managers requiring an independent assessment by the Inspector General" (AFI 90-201, 2018). A wing's risk-based sampling strategy should focus on the wing commander's priorities and areas that are considered high risk. As stated in AFI 90-201 (2018), RBSS should include information sources that are objective performance indicators at the program and unit levels, which collectively provides a reliable assessment on a unit's effectiveness and efficiency. Although there is no formal guidance on what metrics to use or how to use them in an RBSS, some examples of these information sources from AFI 90-201 (2018) are:

- Air and Space Expeditionary Forces Unit Type Code Reporting Tools/Defense Readiness
 Reporting System
- Quality Assurance and Standardization/Evaluation programs
- Functional assessments, inspection results, after-action reports, and meeting minutes
- Individual Medical Readiness reports
- Individual Training Records
- Personal observations and Unit Self-Assessment Program results

- Climate surveys.

While AFI 90-201 (2018) states MAJCOM IG teams will build a tailored inspection risk-based sampling strategy for each Wing, it appears that the wings are expected to build their own risk-based sampling strategy.

23rd Wing

The 23rd Wing interprets RBSS as a sampling technique that looks at the two axes of "Likelihood of Happening" and "Impact to Mission" to prioritize which processes get inspected using a risk matrix or heat map. Beginning with the wing commander's intent and what areas are deemed to be the riskiest, the IG team and subject matter experts map different programs to the risk matrix.



Figure 1. 23rd Wing Risk-Based Sampling Strategy

Figure 1 shows an example of how the risk matrix is used. The horizontal axis shows mission impact, while the vertical axis shows the chance of undetected noncompliance. Programs deemed most likely to occur with the highest consequences are plotted accordingly. For example, as shown in Figure 1, a natural disaster will affect the ability to accomplish the mission, but the chance of undetected noncompliance is low; therefore, it is color-coded yellow. Those in the red

area of the risk matrix are riskiest and should be inspected first, followed by yellow and green.

The risk matrix helps build the basis of the wing inspection plan in the Commander's Inspection

Program, laying out the inspection schedule or inspection priority.

However, Air Combat Command (ACC) determined the 23rd Wing's RBSS is deficient as it is event-based and not risk-based (Figure 1). Due to the lack of a standardized risk-based sampling strategy across the MAJCOM, ACC sent the 23rd Wing an example of an acceptable RBSS used by the 129th Rescue Wing (RQW).

129th Rescue Wing

The 129th RQW's product (Figure 2) helps their Inspector General team prioritize which units and programs to inspect based on risk level. For vertical inspections conducted for specific units, the 129th RQW identifies 24 criteria for evaluation but use only 21 as shown in Table 2. The three criteria not used in the evaluation are identified with an asterisk. They are Readiness Reporting, Physical Training Test Passing Percentages, and Air Reserve Component Network Expeditionary Skills Rodeo Training Report Currency. Readiness reporting data is masked since the information is protected at a higher classification level. Physical training test passing percentage is eliminated due to it not being a priority for the commander. The Air Reserve Component Network Expeditionary Skills Rodeo Training Report Currency is not applicable for the period. The first nine criteria shown in Table 2 are recommended by AFI 90-201 (2018).

Table 2. 129th Rescue Wing Risk-Based Sampling Strategy Criteria

- Wing Commander Priority
- Readiness Reporting
- Quality and Assurance Standardization Evaluation Reports*
- Photo Album Reports [functional assessments, Staff Assistance Visits, After Action Reports, meeting minutes, etc.]
- Individual Medical Readiness
- Individual Training Records
- Airman/Inspections Directorate Personal Observations
- Unit Self-Assessment Program
 (USAP) Findings [Weekly Action
 Report, Commander's Inspection
 Management Board, USAP
 Tracker, Management Internal
 Control Toolset (MICT)]
- Climate Surveys
- Continuous Evaluation Results
- Safety Reports
- Quarterly Inspection Working Group, Guard Inspector General Council/Semi-Annual Inspection Council, Command Interest Items, Special Interest Items, Higher Headquarters Concerns for last year
- Monthly End Strength

- Non-deployable monthly roster from Force Support Flight
- Physical training test passing percentages*
- Performance reports/Airman Comprehensive Assessment timelines
- Air Reserve Component Network Expeditionary Skills Rodeo Training Report Currency*
- Waivers
- Time Since Last Inspection
- Commander/Program Manager Off Station >4 months or Drill Status Guard
- Critical and Significant Inspection Deficiencies from 106th Rescue Wing and 144th Fighter Wing for the last 2 years
- MICT Compliance
- Inspections Directorate MICT Metrics Assessment including quality of responses
- Inspections Directorate Inspector General Evaluation Management System Metrics Assessment

Note: An asterisk (*) identifies criterion not assessed

IGI Member:									1				
low number = inspect/higher vis high number = no inspect/lower vis Green = recommended by 90-201 Blue = complete	Gonkulator gonkulated on:	30-Dec-19											
Criteria	Wing CC (weighted order of	Priority of magnitude x 2)	QA Stan/Eval Reports	Photo Album Reports (Functional assessments, SAVs, AARS, meeting mins,	Amn/IGI Personal Obs	USAP Findings (WAR, CIMB, USAP Tracker, MICT)		Continual Eval Results					
Scaling Criteria	1=very high priority 2= 3= 4= 5= very low priority		reports 2=no	1= very unfavorable reports 2=unfavorable reports 3=neutral/no reports 4=favorable reports 5=very favorable reports	1= very unfavorable obs 2=unfavorable obs 3=neutral obs 4=favorable obs 5=very favorable obs		nt needed	2=no CE 3=medium risk 4=					
Units									SUM	RISK			
WS	3	6	2		1	2	3	1		Highest Ris			
OG Staff (OGV)	2	4	3		4	4	4	2	64	Higher Risk			
OSS	1	2	3		4	4	3	1	57	Highest Ris			
129 RQS	3	6	3		4	4	4	2	76	Least Risk	Mean=	68.2222	
130 RQS	1	2	3		1	2	3	2	66	Higher Risk	Std Dev=	6.52196	
131 RQS	3	6	3		4	4	3	2	73	Less Risk			
MXG Staff	2	4	3	4	3	3	5	2	63	Higher Risk	Highest Risk	61.7003 to	55.1783
AMXS	3	6	3		4	3	4	2	72	Less Risk	Higher Risk	68.2222 to	61.7003
MXS	3	6	3	3	3	3	4	2	65	Higher Risk		68.2222	
MOF	3	6	3		4	3	4	2	73	Less Risk	Less Risk	74.7442 to	68.2222
MSG Staff	4	8	2		4	4	4	2	78	Least Risk	Least Risk	81.2661 to	74.7442
LRS	3	6	3		4	5	4	3	75	Least Risk			
SFS	1	2	4	4	3	4	4	3	69	Less Risk			
CF	2	4	3		3	3	3	1	61	Highest Ris			
CEF	3	6	3		3	3	4	1	68	Higher Risk			
FSF	2	4	2		4	4	2	5	71	Less Risk			
MDG	2	4	3		1	2	2	1	61	Highest Ris			
CPTF	4	8	4		4	4	4	5	77	Least Risk			

Figure 2. 129th Rescue Wing Risk-Based Sampling Strategy (Truncated)

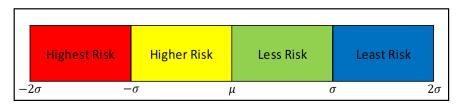


Figure 3. 129th Rescue Wing Risk-Based Sampling Strategy Risk Coding

Figure 2 shows a screenshot to illustrate the ACC-approved (129th RQW) RBSS tool whose sheer size would make it otherwise difficult to portray. See Appendix A for the full RBSS tool. For unit inspections, the wing commander priority is given twice the importance of all other criteria, but all other remaining criteria appear to be weighted equally. However, different criteria use a different scoring scale, such as 3, 4, 5, or 10 points without considerations for standardization. All the criteria are summed for each unit to give a total score where the lower the sum the higher the risk. Thus, higher point scale items bear more weight in the decision process. The mean and standard deviation are calculated separately for the different units. These means and standard deviations are then used to assess risk.

Units within one standard deviation of the mean are deemed higher risk (yellow) or less risk (green). Units two standard deviations of the mean are considered highest risk (red) or lowest risk (blue). Although not explicitly stated, this tool is likely using the Empirical Rule, which states that approximately 68% of scores will fall within 1 standard deviation of the mean, approximately 95% of scores will fall within 2 standard deviations of the mean, and approximately 99.7% of scores will fall within 3 standard deviations of the mean (McClave et al., 2018). The mathematical notation is $\bar{x} \pm ks$, where \bar{x} is the sample mean, s is the sample standard deviation, and k is the number of standard deviations from the sample mean. The tool uses k = 1 for yellow or green and k = 2 for red or blue. Units are inspected based on risk level, where higher risk areas (red), which counterintuitively have the lowest total score, are first inspected. Figure 3 shows how the creator of the 129th RQW's RBSS tool determined the risk coding, likely based on the Empirical Rule (Figure 4).

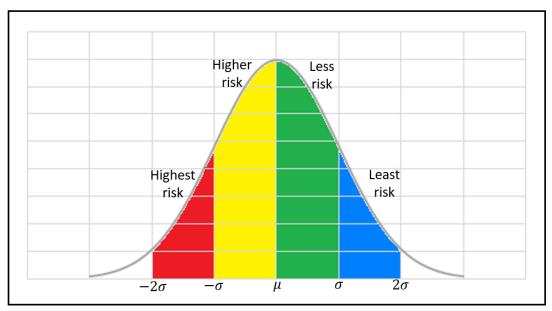


Figure 4. 129th Rescue Wing Risk-Based Sampling Strategy Risk Coding Under a Normal Curve

Upon further research, the 129th RQW was not the creator of their RBSS strategy. The 129th modified an RBSS that was obtained from the 163rd Attack Wing (ATKW) to meet their needs. As such, the overall framework of the 163rd ATKW's RBSS is very similar to its successor. However, there is an added layer of complexity as it incorporates a risk assessment matrix into the tool with input categories for both the probability and severity of the risks.

163rd Attack Wing

In the August 2020 version of the 163rd ATKW RBSS, 13 inputs, categorized by unit, wing and external, determined the probability of risk, and two inputs, one of which was not assessed, determined the severity of the impact to the mission during a vertical inspection. Table 3 showcases the different input criteria for vertical inspections.

Table 3. 163rd Attack Wing Risk-Based Sampling Strategy Criteria (2020)

Probability of occurrence

Unit

- Unit Root Cause Data
- Management Internal Control Toolset (MICT) Open Observations over 180/365 Days
- Waivers Assigned to Unit
- Does the Unit Track
 Observations/Issues to Closure
 Outside of MICT?
- Personnel in Excessive Training (over 36 months) or awaiting training school (over 6 months)

Wing

- Time lapse since last vertical inspection
- Wing Root Cause Data
- Inspector General Evaluation Management System Open Deficiencies over 180 Days
- Repeat Write-ups?
- Unit Health (Defense Organizational Climate Survey/Inspection Findings)

External

- Unit Self-Assessment Program/After Action Reports
- Time Lapse since last external assessment
- Continual Evaluation

Severity of impact to mission

- Completing Mission Assurance Exercise*
- Commander's Priority

Note: An asterisk (*) identifies criterion not assessed.

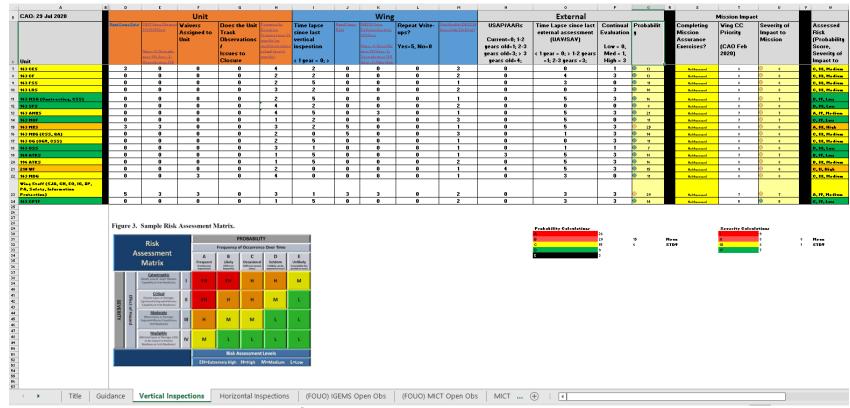


Figure 5. 163rd Attack Wing Risk-Based Sampling Strategy (2020)

At first glance, the 163rd ATKW's RBSS (Figure 5) criteria all appear to be weighted equally and scores are summed when determining the probability of occurrence and severity of mission impact. However, different criteria use a different scoring scale with maximum values such as 3, 4, and 5 points without considerations for standardization. For example, time lapse since last vertical inspection has a maximum value of 5, while continual evaluation has a maximum value of 3. In essence, the 5 will carry more weight than the 3 when they both represent 100% of their respective criterion. All the criteria are summed for each unit to give a total score where the lower the sum the lower the risk. [Note: The 129th RQW modified the tool to reflect the opposite where the lower the sum of the criteria, the higher the risk. This is likely due to metrics used such as climate surveys already have an identified scale where the higher the value, the better the performance.]

The sample mean and standard deviation are calculated for probability and severity inputs. For probability inputs (Figure 6) using the scale A-E in descending order of likelihood from Figure 5, the mean value is set as the lower bound for occasional (or C) probability; values one standard deviation above the mean are deemed likely (or B); two standard deviations above the mean are frequent (or A); values one standard deviation below the mean are seldom (or D); and values two standard deviations below the mean are unlikely (or E). For severity inputs (Figure 7), there are only four possible outputs: catastrophic (I), critical (II), moderate (III), and negligible (IV). The mean is set as the lower bound for moderate severity (III). One standard deviation above the mean is considered critical (II) and two standard deviations above the mean is considered negligible (IV). The outputs for both probability and severity are then mapped to the risk assessment matrix (Figure 8) to determine the level of risk ranging from extremely high to low,

with the higher risk units inspected first. If the two units have the same risk level, then priority will be determined based on available resources such as subject matter experts to assist with the inspection.

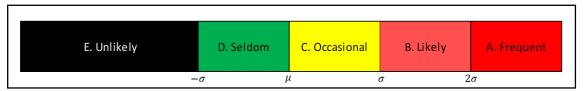


Figure 6. 163rd Attack Wing Risk-Based Sampling Strategy Probability Coding

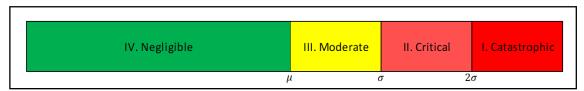


Figure 7. 163rd Attack Wing Risk-Based Sampling Strategy Severity Coding

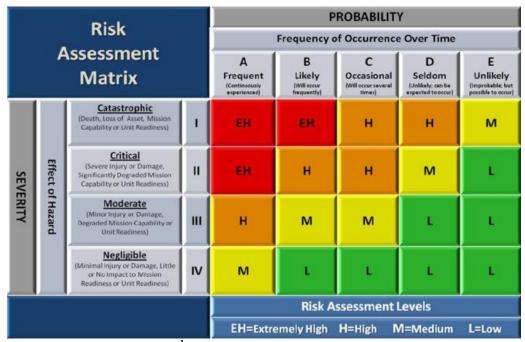
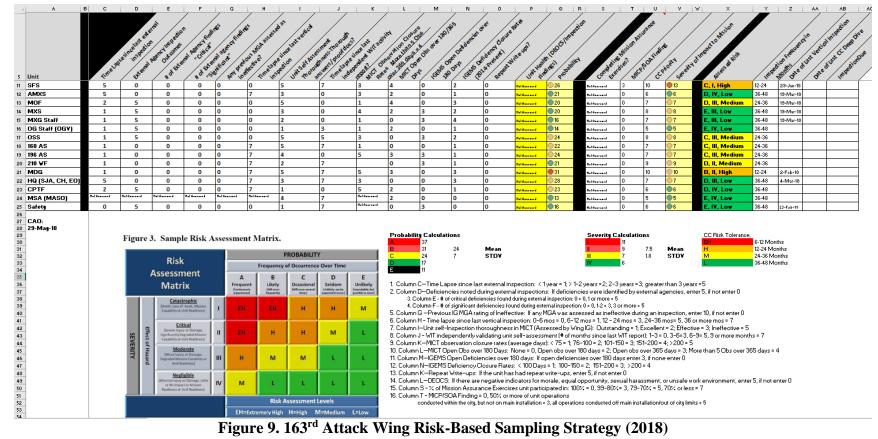


Figure 8. 163rd Attack Wing Risk Assessment Matrix



The June 2018 version of the 163rd ATKW RBSS (Figure 9) is the version the 129th RQW used as the foundation of their RBSS. The primary difference between the June 2018 version and the July 2020 versions are the number of criteria, the criteria selected for evaluation and the categorization of probability inputs (unit, wing, and external) for the latter. The overall framework for calculating risk did not change between the two products.

Table 4. 163rd Attack Wing Risk Based Sampling Strategy Criteria (2018)

Probability of occurrence

- Time lapse since last external inspection
- External agency inspection outcomes
- Number of external agency findings "critical"
- Number of external agency findings "significant"
- Any previous major graded areas assessed as ineffective
- Time lapse since last inspection
- Unit self-assessment thoroughness
- Time lapse since last Wing Inspection Team activity report
- Management Internal Control Toolset (MICT) observation closure rates
- MICT open observations over 180/365 days
- Inspector General Evaluation Management System (IGEMS) open deficiencies over 180 days
- IGEMS deficiency closure rates
- Repeat write ups
- Unit Health*

Severity of impact to mission

- Completing Mission Assurance Exercise*
- Manager's Internal Control Program/Statement of Assurance Findings
- Commander's Priority

Note: An asterisk (*) identifies criterion not assessed.

The June 2018 version used 14 inputs, one of which was not assessed, were considered to determine the probability of a risk to a unit and three inputs, one of which was not assessed, were considered to determine the severity of the impact to the mission during a vertical inspection.

Table 4 lists the different input criteria for vertical inspections.

As we continue to trace the roots of the RBSS strategy, we determined that the 163rd ATKW also modified an RBSS provided to them by ACC IG in 2017. Neither the MAJCOM IG who gave the 163rd ATKW the example nor the person identified as the originator of the Excel file are in the Air Force email global directory service. For further insight, we reached out to ACC IG and was told the spreadsheet may have originated from the 366th Fighter Wing (FW). Due to personnel turnover, we were unable to contact the originator of the spreadsheet but connected with the current 366th FW RBSS subject matter expert (SME). However, the current draft of the 366th FW RBSS tool (Figure 10) does not appear to be the parent tool of the 163rd ATKW tool as it does not resemble what the 163rd ATKW claims to be the unmodified example originally received from ACC. However, due to personnel turnover again and potential product improvements over time, we cannot say with 100% certainty that the roots of the 129th RQW RBSS tool did not originate from the 366th FW. As we were unable to contact the creator of the RBSS, our analysis is limited to our interpretations of the spreadsheet and information from the current SME.

366th Fighter Wing

The 366th FW provided a draft of their current RBSS (Figure 10), which they are in the process of modifying and understanding. It includes objective and subjective metrics and complicated formulas that are difficult to follow to determine a total unit/risk score. The objective metrics include last inspection date (which is turned into a multiplier), the percentage of self-assessment programs completed, the observation/deficiencies ratio (the number of Management Internal Control Toolset or MICT observations in the last 12 months over the number of Inspector General Evaluation Management System or IGEMS deficiencies in the last

24 months), the number of IGEMS deficiency points, and the number of MICT deficiency points.

The subjective assessment component allows subject matter experts (SMEs) and leadership to express their concerns about a unit using a scale of 0-3 where 3 represents very concerned. The panel of subjective assessors include the wing commander, IG, IG director of inspections, IG superintendent, inspection planner, assistant inspection planner, leader exercise, assistant exercise, wing self-assessment program manager, and Wing Inspection Team Manager. Each SME opinion is equally weighted except for the fighter wing commander's assessment which has a weight of 3, the IG's assessment which has a weight of 2.25 and the IG director of inspections which has a weight of 1.75. The weighted average of SME opinions become one subjective value, which then translates into subjective inspection priority score. This subjective assessment value, time since last inspection multiplier, and observations and deficiencies metrics are used to determine the total unit/program risk score using an unintuitive complicated formula that a new RBSS officer would have difficult deciphering. Figure 11 shows the risk score formula and relevant columns used in the formula.

While we cannot determine whether the 366th FW draft RBSS precedes the 163rd ATKW and the 129th RQW's, it further indicates the need to have a standardized RBSS process that is easily to use, comprehend and turnover to somebody else, so that resources are not wasted to decipher a complicated product or to reinvent the wheel every time there is a turnover.

4	В	C	D	E	F	G	Н	1	J	K	L	M	N	0
1	Organization	Total Unit/Program Risk Score	Last Inspection Date	Mulitplier	SAP Completed Rubric Score	% SAP Completed (last 12 months) = # compl/total SACs	Observation/Deficiencies Ratio	Subjective Assessmen t	# of IGEMS Deficiencies (Internal + External) (Last 24 Months)	IGEMS Deficiency Points Critical 5 pts ea Significant 3 pts ea (Last 24 months)	# of MIC-T Observations (last 12 months)	MICT Deficiency Points Critical 5 pts ea Significant 3 pts ea (Last 12 months)	GSU Step Yes = 1 No = 0	AF/MAJCO M SII/CII (Interest item = 5)
3	391 FS	72	9-Jul-20	4	0	100%	4	0	18	5	5 4		0	
4	391 FGS	10	25-Jan-22	1	0	100%	0	0	0	0	6	0	0	
5	390 ECS	15	24-Sep-21	1	0	100%	5	0	9	3	0	0	0	
6	389 FS	42	25-Jun-21	2	0	100%	3	0	22	16	11	0	0	
7	389 FGS	#REF!		4	0	100%	0	0	0	0 0		0	0	
8	428 FS	#REF!	10-Jan-20	4	0	93%	3	0	18	6 5		0	0	
9	366 MXS	#REF!	11-Jun-21	2	0	100%	3	0	34	15	9	5	0	
10	366 MUNS	#REF!	9-Oct-20	3	0	96%	4	0	24	6	4	0	0	
11	366 CES	#REF!	4-Sep-20	4	0	100%	3	0	32	9	16	9	0	
12	366 CS	#REF!	20-Nov-20	3	0	99%	2	0	22	20	44	9	0	
13	366 FSS	#REF!	30-Apr-21	2	0	100%	2	0	29	15	44	9	0	
14	366 FAS	#REF!	28-Feb-20	4	0	98%	2	0	15	15 18		3	0	
15	366 LRS	#REF!	12-Feb-21	3	0	100%	2	0	29	18 32		6	0	
16	366 OSS	#REF!	18-Sep-20	3	0	100%	3	0	32	12	22	18	0	
17	366 SFS	#REF!	7-Aug-20	4	0	100%	3	0	26	3	12	3	0	
18	366 OMRS	#REF!	24-Jul-20	4	0	100%	2	0	17	3	35	3	0	
19	366 HCOS	#REF!	2-Apr-21	2	0	100%	2	0	10	3	17	0	0	
		.al B	С		D	Q	R S	Т	U V	W X	Y Z	AA AB	AC	

-4	В	С	D	Q	R	S	T	U	V	W	X	Υ	Z	AA	AB	AC
1				IGI Subjec	<u>essment</u>		<u>Grading Rubric:</u> 0 = Not Concerned, 1 = Slighty Concerned, 2 = Concerned, 3 = Very Concerned									
2	Organization	Total Unit/Program Risk Score	Last Inspection Date	FW/CC	IG	IGI, Director	IGI, Super	Inspection Planner	Asst Inspection planner	Lead Exercise	Asst Exercise	FW SAPM	WIT Manager	Weighted Average	FW Priority	ABS Value of FW priority
3	391 FS	72	9-Jul-20	0	0	0	0	0	0	0	0	2	0	1	8	10
4	391 FGS	10	25-Jan-22	0	0	0	0	0	0	0	0	2	0	1	8	10
5	390 ECS	15	24-Sep-21	0	0	0	0	0	0	0	0	1	0	1	8	10
6	389 FS	42	25-Jun-21	0	0	0	0	0	0	0	0	2	0	1	8	10
7	389 FGS	#REF!		0	0	0	0	0	0	0	0	2	0	1	8	10
8	428 FS	#REF!	10-Jan-20	0	0	0	0	0	0	0	0	1	0	1	8	10
9	366 MXS	#REF!	11-Jun-21	0	0	0	0	0	0	0	0	2	0	1	8	10
10	366 MUNS	#REF!	9-Oct-20	0	0	0	0	0	0	0	0	1	0	1	8	10
11	366 CES	#REF!	4-Sep-20	0	0	0	0	0	0	0	0	2	0	1	8	10
12	366 CS	#REF!	20-Nov-20	0	0	0	0	0	0	0	0	3	0	1	8	10
13	366 FSS	#REF!	30-Apr-21	0	0	0	0	0	0	0	0	3	0	1	8	10
14	366 FAS	#REF!	28-Feb-20	0	0	0	0	0	0	0	0	2	0	1	8	10
15	366 LRS	#REF!	12-Feb-21	0	0	0	0	0	0	0	0	2	0	1	8	10
16	366 OSS	#REF!	18-Sep-20	0	0	0	0	0	0	0	0	2	0	1	8	10
17	366 SFS	#REF!	7-Aug-20	0	0	0	0	0	0	0	0	1	0	1	8	10
18	366 OMRS	#REF!	24-Jul-20	0	0	0	0	0	0	0	0	1	0	1	8	10
19	366 HCOS	#REF!	2-Apr-21	0	0	0	0	0	0	0	0	1	0	1	8	10

Figure 10. 366th Fighter Wing Risk-Based Sampling Strategy (draft)

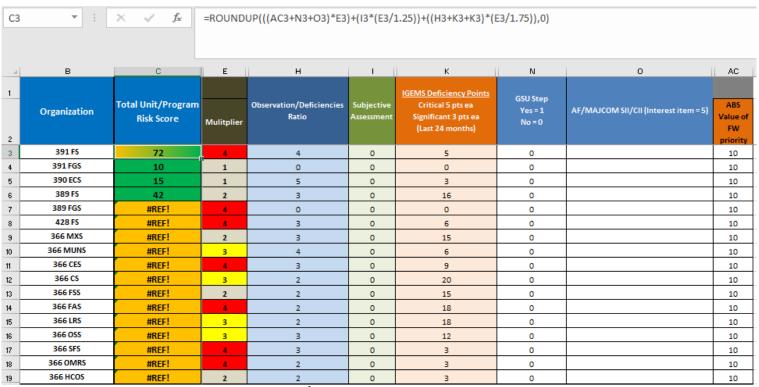


Figure 11. 366th Fighter Wing Risk Score Formula

Lieutenant General Stephen P. Mueller (2014), prior Air Force Inspector General, once described the new Air Force inspection system (AFIS) as the Air Force's "single largest cultural change in the past four decades," yet the Office of the Secretary of the Air Force, Inspector General, Inspections Directorate (SAF/IGI) directed the "implement first, innovate later" strategy (2013, as cited in Craft, 2016). As such, the rushed nature of the implementation of AFIS resulted in open interpretation of how to create a risk-based sampling strategy.

While we consider decision analysis tools such as the multi-objective decision-making process and analytic hierarchy process in our research (Saaty, 1994; Kirkwood, 1995; Eschenbach 2011), it is highly unlikely wings will adopt an unfamiliar and more complex approach. Therefore, we focus our research on modifying and improving an existing ACC-approved tool that has been used as an example forming the foundation of many wing's RBSS.

The general structure of the 129th RQW and 163rd ATKW RBSSs loosely resemble Likert scale processes. To avoid bias, Likert items generally have a balance of both positive and negative responses on a symmetric scale where each successive response option is considered better, although the opposite is true for a reverse Likert scale (Willits et al., 2016). To achieve balance, different response choices for a Likert item should be equidistance apart. If all Likert items use the same scale, then their responses may be summed together to create a score for each respondent (Johns, 2010; Willits et al., 2016). Although both the 129th and 163rd RBSSs do not follow the general guidelines for the Likert scale process, the wings do sum different-scaled unbalanced Likert responses to get a total score for each organizational unit.

Based on the existing tools we examined, there is no standard weighting system resulting in potential skewed results and possibly inaccurate decisions. The extensive number of inputs that need to be manually entered into the tool is time consuming and cumbersome, which may

discourage true implementation to inform decisions. To prevent RBSS from becoming an extension of the previous bad habit of painting the grass green to pass inspection, there needs to be a comprehensive standardized process. In addition, due to personnel turnover, an RBSS that previously passed inspection or sent out as an example may become the gold standard without anybody questioning whether the framework is statistically sound.

Summary

Overall, we were unable to find published literature to substantiate the process implemented by the various discussed units. Furthermore, we were unable to locate sources on how to combine independent criteria scores using different scales while mixing Likert scale variables with dichotomous variables, which we further discuss in Chapter 3. In this chapter, we discussed the relevant history of the Air Force inspection system and the purpose of a risk-based sampling strategy. We reviewed RBSS tools currently used by the 23rd Wing, 129th RQW, 163rd ATKW, and 366th FW. In Chapter 3, we hone in on the haphazard nature of the current processes, focusing on the tools of the 129th RQW and 163rd ATKW.

III. Methodology

Chapter Overview

The purpose of this chapter is to examine the data in the 129th Rescue Wing (RQW) and 163rd Attack Wing's (ATKW) risk-based sampling strategies (RBSSs) focusing on their process for vertical inspections. We begin by describing the current tool and explaining its different components. We then discuss what analysis techniques we use to assess the statistical properties of the current tool. Part of this assessment entails investigating the dichotomous variables used in the two RBSSs. We describe the process of calculating the percent contribution to the response variable and changing the values selected to represent the dichotomous variable to see how it impacts the tool. The descriptive evaluation of the presented data determines what values to use in our simulation as we attempt to reproduce those Likert scale patterns using a binomial distribution. We explain our data simulation process and how it affects the standardized score and how metrics are calculated.

Risk-Based Sampling Strategy Instruments and Components

The 129th RQW and 163rd ATKW RBSS instruments are structurally similar. To better understand how the instruments work, we break them down into their components. The tools consider the following data: number of organizational units, number of criteria, and defined scales for each criterion. The instrument first identifies evaluation criteria for determining unit risk. Each criterion has its respective defined scale, which ranges from 2 to 5 points. The evaluation criteria are then scored and summed for each unit. The aggregate score for all units is then converted into a standardization metric where its sample mean and sample standard deviation are used to categorize units into different risk groups and to determine their inspection priority from highest to lowest risk.

Assumptions and Limitations

Our analysis is limited by the following assumptions: (1) All criteria each wing selects for evaluation are independent of one another so there is no duplication of effort. (2) Organizational units required to be inspected are separate entities and independent of one another. (3) The user wants each criterion to contribute equally to the sum, which is the response of interest. (4) The user wants to keep the instrument simple and does not want to perform complicated calculations or scalings. (5) The user prefers an Excel spreadsheet format. (6) All independent criteria are modeled as binomial random variables with the same probability of success, which allow us to investigate the overall data dominance patterns and not have contradictory responses that cancel each other out and conceals the pattern. (7) We only consider half of the bivariate model, the portion of the data labeled as probability of occurrence, in the 163rd ATKW RBSS (Table 3) in our analysis since those are the values summed to help inform the decision and the portion used as the framework for the 129th RQW RBSS. (8) For simplicity, our analysis will assume all criteria use the same Likert scale. By using one Likert scale, patterns become more apparent. (9) The process for the nonmandatory horizontal inspections focusing on programs instead of units use the same overall framework and process as vertical inspections for determining inspection priority but fewer evaluation criteria. As such, we focus on the process used for mandatory vertical inspections.

Instruments Used for Analysis

For our analysis, we use two different software tools: Excel spreadsheet and JMP Pro 15 software. Excel spreadsheet is the chosen tool for current RBSSs as it is intuitive and straightforward to use. JMP Pro 15 software allows easy data simulation. We use JMP Pro 15 to generate random data changing the values for the number of units, number of criteria and type of

scale used to examine how the response is impacted. The simulation process reveals the theoretical responses while varying inputs. Descriptive statistical data from both the 129th RQW and 163rd ATKW help define the range of values used for simulation.

Determining Values for Simulation

Four input variable values are used in data simulation: the number of organizational units, the number of criteria, the type of Likert scale used, and the percentage of successful trials. We will discuss the last input variable in the section regarding the binomial distribution. The range of values selected for each input variable are based on actual data observed in the 129th RQW and 163rd ATKW RBSSs. These input ranges are used in JMP Pro 15 to simulate different scenarios and to determine the theoretical distribution for our response variable, which is the sum of a unit's score.

Number of Units

Both the 129th RQW and 163rd ATKW coincidentally have the same number of units to inspect, 18. To include a third data point, we also consider the number of units the 366th FW is required to inspect, which is 17, as we have that information available. For our simulation, we select a range of 10 to 20 units and keep the increments evenly spaced (five). We use 10 units as the minimum value to account for smaller organizations.

Number of Criteria

To determine the number of criteria to use in our analysis, we examine the data of both 129th RQW and 163rd ATKW. The 129th RQW's risk-based sampling strategy considered 24 criteria as listed in Table 2 with higher scores equaling to lower risk; however, three were not assessed so only 21 were used. The 163rd Attack Wing's risk-based sampling strategy (2020) used 13 criteria as listed in Table 3 for the probability of occurrence with lower scores equaling

lower risk. Since those are the values summed to help inform the decision and the portion used as the framework for the 129th Rescue Wing's RBSS, we only consider them in our analysis. Of the 13 criteria, two do not make any meaningful contribution as all units scored perfectly on them.

The data shows a range of 13 to 21 for the number of criteria. We use a range of five to 25 in increments of five to cover the full range. The lower bound of five criteria is used instead of 10 to explore the patterns of a simpler RBSS model.

Type of Scale Used

To determine what Likert scale to use, we examine what scales were used by the two units. Both the 129th RQW and the 163rd ATKW combined different scales when calculating the response variable. These different scales range from a 2-point Likert scale or dichotomous scale to a 5-point Likert scale. An in-depth look at the data reveals differences between the theoretical or defined scale and the actual scale used. Intuitively, the highest score should reflect the Likert scale used; however, that is not the case as defined scores are not equidistance apart. For example, in the 163rd ATKW RBSS, a criterion may have possible values of 0, 1, 3, and 5 in the defined scale where some values are one unit apart whereas others are two units apart. To determine the true Likert scale of each criterion, we count the number of defined values in the scale or groups. For both highest scores and defined groups, there were discrepancies between the theoretic metrics and their actual counterparts.

For the 129th Rescue Wing, 87.5% (21) of the 24 criteria were used in the evaluation. Three criteria (individual training records, monthly end strength, and MICT compliance) did not specify or did not fully specify its scale. Since the highest value used in the evaluation was a 5, we assumed that those criteria used a high score of 5. Observed assessment values used for those three criteria were divided into three groups, so we assume the theoretical scale is based on a 3-

point Likert scale matching the actual metrics. Table 5 features the 129th RQW RBSS Metrics revealing the discrepancy between possible scores and scores actually used.

Table 5. 129th Rescue Wing Risk-Based Sampling Strategy Metrics

Criteria	Scores possible	Scaros usad	Highest score possible	Highort coord used	Groups possible	Groups used
					· · · · ·	Groups useu
Wing Commander Priority	2,4,6,8,10	2,4,6,8	10	8	5	4
Quality and Assurance Standardization Evaluation Reports	1,2,3,4	2,3,4	4	4	4	3
Photo Album Reports	1,2,3,4,5	2,3,4,5	5	5	5	4
Individual Medical Readiness	1,2,3	2,3	3	3	3	2
Individual Training Records	Unknown	3,4,5	5	5	Unknown (assume 3)	3
Airman/Inspections Directorate Personal Observations	1,2,3,4,5	1,3,4	5	4	5	3
Unit Self-Assessment Program Findings	1,2,3,4,5	2,3,4,5	5	5	5	4
Climate Surveys	1,2,3,4,5	2,3,4,5	5	5	5	4
Continuous Evaluation Results	1,2,3,5	1,2,3,5	5	5	4	4
Safety Reports	1,2,3,4,5	2,3,4	5	4	5	3
Quarterly Inspection Working Group, Guard Inspector						
General Council/Semi-Annual Inspection Council,						
Command Interest Items, Special Interest Items, Higher						
Headquarters concerns for last year	1,2,3	1,2,3	3	3	3	3
Monthly End Strength	Unknown	2,3,5	5	5	Unknown (assume 3)	3
Non-deployable monthly roster	1,2,3,4,5	2,5	5	5	5	2
Performance reports/Airman Comprehensive						
Assessment timelines	1,2,3,4,5	2,4	5	4	5	2
Waivers	1,3	1,3	3	3	2	2
Time Since Last Inspection	1,2,3,4	1,2,3,4	4	4	4	4
Commander/Program Manager Off Station	1,5	1,5	5	5	2	2
Critical and Significant Inspection Deficiencies from	1,2,3	1,2,3	3	3	3	3
MICT Compliance	4,5 defined	3,4,5	5	5	Unknown (assume 3)	3
IGI MICT Metrics Assessment	1,2,3,4,5	1,2,3,4,5	5	5	5	5
Inspections Directorate Inspector General Evaluation						
Management System Metrics Assessment	1,2,3,4,5	1,2,3,4,5	5	5	5	5

Table 6 shows the highest score possible and highest score used as well as the discrepancy between the two. Of the 21 criteria, one (4.76%) had the highest possible value of 10, 14 (66.67%) had the highest value of 5, two (9.52%) had the highest value of 4 and four (19.05%) had the highest value of 3. The possible scores and the scores used are not necessarily the same. Of the 21 criteria, one (4.76%) used a highest score of 8, eleven (52.38%) used a highest score of 5, five (23.81%) used a highest score of 4, and four (19.05%) used a highest score of 3. In other words, one-third of the criteria (seven) did not have any units achieve the theoretical high score. The asterisk by the score of 8 in Table 6 denotes it is the same criterion with a possible score of 10. The total discrepancy only counted that criterion once.

Table 6. 129th Rescue Wing Risk-Based Sampling Strategy Scaling Criteria Highest Score

Score	Possible (#)	Possible (%)	Used (#)	Used (%)	Discrepancy
3	4	19.05%	4	19.05%	0.00%
4	2	9.52%	5	23.81%	14.29%
5	14	66.67%	11	52.38%	14.29%
8*	0	0.00%	1	4.76%	4.76%
10	1	4.76%	0	0.00%	4.76%
Total	21	100.00%	21	100.00%	33.33%

Note: The asterisk (*) denotes the score of 8 is the same criterion as the one with a possible score of 10. Its contribution to the total discrepancy is not double counted.

Table 7. 129th Rescue Wing Risk-Based Sampling Strategy Scaling Criteria Groups

x-Point Scale	Possible (#)	Possible (%)	Used (#)	Used (%)	Discrepancy
2	2	9.52%	5	23.81%	14.29%
3	6	28.57%	8	38.10%	9.52%
4	3	14.29%	6	28.57%	14.29%
5	10	47.62%	2	9.52%	38.10%
Total	21	100.00%	21	100.00%	76.19%

The highest theoretical score does not translate into the theoretical Likert scale used as the possible values defined are not contiguous. As shown in Table 7, the data reveals that the theoretical values for the scale were clustered into two to five groups. Ten (47.62%) criteria use a 5-point Likert scale, three (14.29%) use a 4-point Likert scale, six (28.57%) use a 3-point Likert scale and two (9.52%) use a dichotomous scale. The actual assessment values reveal two (9.52%) criteria use a 5-point Likert scale, six (28.57%) use a 4-point Likert scale, eight (38.10%) use a 3-point Likert scale and five (23.81%) use a 2-point Likert or dichotomous scale. The actual metrics show a 76.19% discrepancy between the theoretical scale and the actual scale. Most significant are three criteria that were defined with a Likert-scale result in assessed values that are dichotomous, inadvertently increasing dichotomous criteria from two to five.

For the 163rd Attack Wing, 13 criteria with differing scales were used to calculate the response variable. Two criteria (waivers assigned to units and unit health) did not specify its

scale. Since the highest value used in the evaluation of both criteria was a 3, we assumed that 3 is the theoretical high. Observed assessment values used for waivers were divided into two groups, so we assume the theoretical scale is based on a dichotomous scale matching the actual metrics. Observed assessment values used for unit health were divided into three groups, so we assume the theoretical scale is based on a 3-point Likert scale matching the actual metrics. Table 8 features the 163rd ATKW RBSS Metrics revealing the discrepancy between scores possible and scores used.

Table 8. 163rd Attack Wing Risk-Based Sampling Strategy Metrics

Criteria	Scores possible	Scores used	Highest score possible	Highest score used	Groups possible	Groups used
Unit Root Cause Data	1,2,3,4,5	0*,3,5	5	5	5	3
Management Internal Control Toolset Open						
Observations over 180/365 Days	0,2,3,4	0,3	4	3	4	2
Waivers Assigned to Unit	Unknown	0,3	Unknown (assume 3)	3	Unknown (assume 2)	2
Does the Unit Track Observations/Issues to						
Closure Outside of MICT?	0,5	0	5	0	2	1
Personnel in Excessive Training (over 36 months)						
or awaiting training school (over 6 months)	1,2,3,4,5	1,2,3,4	5	4	5	4
Time lapse since last vertical inspection	0,1,3,5	0,1,2*,5	5	5	4	4
Wing Root Cause Data	1,2,3,4,5	0*,3,5	5	5	5	3
Inspector General Evaluation Management						
System Open Deficiencies over 180 Days	0,2,3,4	0,3	4	3	4	2
Repeat Writeups	0,5	0	5	0	2	1
Unit Health	Unknown	1,2,3	Unknown (assume 3)	3	Unknown (assume 3)	3
Unit Self-Assessment Program/After Action Reports	0,1,3,4,5	0,3,4	5	4	5	3
Time Lapse since last external assessment	0,1,3,4,5	0,1,3,4,5	5	5	5	5
Continual Evaluation	0,1,3	0,1,3	3	3	3	3
Note: An asterisk (*) indicates the score used was not	defined in the sc	ale.				

There were three criteria (MICT Root Cause Data, Time Lapse Since Last Vertical Inspection, and IGEMS Root Cause Data) where the scale used fell outside of the scale defined as indicated by the asterisk in Table 8. Both MICT Root Cause Data and IGEMS Root Cause Data had scores of 0 in the evaluation when the defined scale was 1-5. Time Lapse Since Last Vertical Inspection consistently used a value of 2 instead of the defined value of 3.

Table 9 shows the highest score possible and highest score used as well as the discrepancy between the two for the 163rd ATKW. Of the 13 criteria, eight (61.54%) had the highest possible value of 5, two (15.38%) had the highest value of 4 and three (23.08%) had the

highest value of 3. Like the 129th RQW's RBSS, the possible scores and the scores used are not necessarily the same. Of the 13 criteria, four (30.77%) used a highest score of 5, two (15.38%) used a highest score of 4, five (38.46%) used a highest score of 3, and two used a highest score of 0 (15.38%). There is a 61.54% difference between theoretical and actual highest scores. Although high scores mean higher risk in this case and the user would value lower scores, this metric is purely revealing the discrepancies between theoretical and actual values from what one would assume to be the Likert scale for the associated criterion at first glance.

Table 9. 163rd Attack Wing Risk-Based Sampling Strategy Scaling Criteria Highest Score

Score	Possible (#)	Possible (%)	Used (#)	Used (%)	Discrepancy
	0	0.00%	2	15.38%	15.38%
	3	23.08%	5	38.46%	15.38%
	4 2	15.38%	2	15.38%	0.00%
	5 8	61.54%	4	30.77%	30.77%
Total	13	100.00%	13	100.00%	61.54%

Table 10. 163rd Attack Wing Risk-Based Sampling Strategy Scaling Criteria Groups

x-Point Scale	Possible (#)	Possible (%)	Used (#)	Used (%)	Discrepancy
1	0	0	2	15.38%	15.38%
2	3	23.08%	3	23.08%	0.00%
3	2	15.38%	5	38.46%	23.08%
4	3	23.08%	2	15.38%	7.69%
5	5	38.46%	1	7.69%	30.77%
Total	13	100.00%	13	100.00%	76.92%

Just like its successor the 129th RQW RBSS, the highest theoretical score does not translate into the theoretical Likert scale used as the possible values defined are not contiguous. The data reveals that the theoretical values for the scale were clustered into two to five groups (Table 10). Five (38.46%) criteria use a 5-point Likert scale, three (23.08%) use a 4-point Likert scale, two (15.38%) use a 3-point Likert scale and three (23.08%) use a dichotomous scale. The actual assessment values were clustered into one to five groups. One (7.69%) criterion use a 5-point Likert scale, two (15.38%) use a 4-point Likert scale, five (38.46%) use a 3-point Likert

scale, three (23.08%) use a dichotomous scale, and two (15.38%) use a 1-point scale. The two criteria with an actual 1-point scale reveals that they did not contribute to the decision as all units received the same score. As such, these criteria could be considered for elimination. The actual metrics show a 76.92% discrepancy between the theoretical scale and the actual scale, which is almost identical to discrepancy the 129th RQW had of 76.19%.

Considering how responses to the criteria were grouped, the most common scale used by both wings was 3. For the 129th RQW, the next most common scale used was 4 followed by 2 and 5. For the 163rd ATKW, the next most common scale used was 2, followed by 4, and 5. As the data suggests, we need to consider a 3- to 5-point Likert scale in our simulation. For simplicity, Likert scales will not be mixed in the simulation process. We do not consider using purely dichotomous variables in our simulation for it is very unlikely that operational units will do that since it lacks the detail required to help make informed decisions. Interesting to note, both wings used approximately 23% dichotomous responses in their assessments. We will further explore the dichotomous variable and its optimal values (or ranges) when summed with values from that of either a 3-point, 4-point, or 5-point Likert scale.

Binomial Distribution

We choose a binomial distribution to model the theoretical patterns for each criterion. To use the binomial distribution, two elements are required: the probability of success (p) and the number of trials (n). For our study, we define the number of trials as the number of possible responses in the selected Likert scale and the probability of success as the chance of scoring the theoretical high on the scale. Although the binomial distribution has a set definition of probability of success, our study focuses on the visual patterns of the distributions for varying values of p instead since units may have differing definitions of success. One unit may consider

high scores as success while another may use a reverse Likert scale and consider low scores as success. As such, it would be misleading to always call the low values failures.

To get the theoretical distributions/shape of different probabilities of success, we used JMP Pro 15 and a binomial distribution to simulate data. For a Binomial (n, p) distribution, n is the number of trials (including 0) and p is the probability of success. Starting with a 5-point Likert scale and random sample size of 100,000, we generate a random binomial modifying the formula to: Random Binomial (4, p) + 1. With an n = 4, we get values between 0 and 4. Since we are interested in values between 1 and 5 for a 5-point Likert scale, we add 1 to the formula, essentially shifting the values to match the Likert scale of interest.

We use different values of p, ranging from 0.1 to 0.9 in increments of 0.2 to generate the theoretical distributions varying the degree of skewness in the pattern. Skewness refers to the asymmetry of the distribution due to the mean's position relative to the tails of the distribution. If the mean is closer to the left-tail, the distribution is skewed left and vice versa. However, we want to focus on where the data is dominantly portrayed in the distribution, so we will use the term dominant to describe the varying distributions. When p = 0.1, the distribution is very left-tail dominant where most of the data resides in the lowest value. When p = 0.3, the distribution is somewhat left-tail dominant where most of the data resides in the lower values. When p = 0.5, the distribution is central dominant where most of the data resides in the center. When p = 0.7, the distribution is somewhat right-tail dominant where most of the data resides in the higher values. When p = 0.9, the distribution is very right-tail dominant where most of the data resides in the higher values. The distributions when p = 0.1 and p = 0.9 are mirror images of each other as are the distributions of p = 0.3 and p = 0.7. We repeated the process for generating the theoretical distributions for the 3-point and 4-point Likert scales.

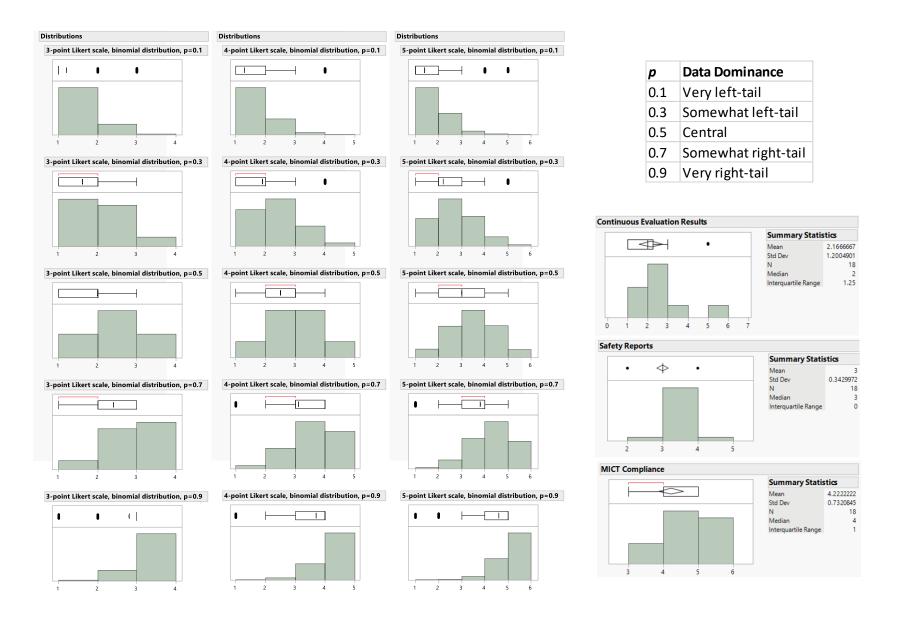


Figure 12. Theoretical and Actual Binomial Distribution for 3-, 4- and 5-point Likert Scales

Figure 12 shows the theoretical distributions for the 3-, 4-, and 5-point Likert scales (from left to right) along with the patterns of p for 0.1, 0.3, 0.5, 0.7, and 0.9 (from top to bottom), a table describing the different data dominance patterns, as well as some examples of actual criteria distributions for easy comparison. The graphs of select criterion on the right side of Figure 12 (from top to bottom) have a p of approximately 0.3, 0.5, and 0.7, respectively, when compared to their theoretical counterparts. See Appendices B and C for the graphs and descriptive statistics of every criterion used by the 129th RQW and 163rd ATKW, respectively.

Aside from displaying the patterns for each individual criterion, the binomial distribution can also illustrate the theoretical pattern on our response variable, the sum of all criteria. As previously stated, we assume that each criterion is an independent binomial with the same probability of success. The sum of independent binomials with the same probability of success is also a binomial (Casella & Berger, 2002).

$$x_1 \sim Bin(n_1, p)$$

$$x_2 \sim Bin(n_2, p)$$

If x_1 and x_2 are independent, then $x_3 = x_1 + x_2$.

$$x_3 \sim Bin(n_1 + n_2, p)$$

In our study, *x* represents an evaluation criterion, *n* represents the possible outcomes of a specific Likert scale (modified to remove the 0 value in the binomial) and *p* represents the probability of success.

Data Simulation

With our four input variables and values of interest defined, we can simulate our data in JMP Pro 15. The number of organizational units have three values (10, 15, and 20). The number of criteria have five values (5, 10, 15, 20, and 25). The Likert scale has three values (3-point, 4-

point, and 5-point). The probability of success (for the binomial random variable) has five values (0.1, 0.3, 0.5, 0.7, and 0.9). We will explain the process for simulating data for a 5-point Likert scale RBSS for a wing who has 10 units to inspect. This process is repeated eight more times to capture data for all nine combinations of number of units and Likert scales.

In JMP Pro 15, we generate a Full Factorial Design structure. The response variable is named sum to represent the sum of all criteria scores per organizational unit. We create three continuous factors: criteria, skewness, and repetitions. Criteria represents the number of evaluation criteria and has five levels: 5, 10, 15, 20, and 25. Skewness represents the different binomial probabilities of success and has five levels: 0.1, 0.3, 0.5, 0.7, and 0.9. Repetitions represent the simulation sample size and we choose a value 10,000. A 5x5x10000 factorial yields 250,000 runs. We keep the run order the same for the experiment. The number of replicates represent one less than the number of units for which we want to simulate data. For example, if we are interested in simulating 10 units, we would use nine replicates as nine plus the original one equals 10 units, which generates 2.5 million rows of data. We would change the number of replicates to 14 for 15 units and 19 for 20 units, which generates 3.75 million and 5 million rows of data, respectively.

Next, we make a table for our factorial design. We insert a formula for our response variable, the sum, and tie in all variables of interest. The generic formula for the sum where n is one less than the Likert point scale of interest, k is the number of criteria and p is the skewness is:

Random Binomial (n*k, p) + k

Following the assignment convention in our factorial design, the formula for the sum using a 5-point Likert scale is:

Random Binomial (4*Criteria, Skewness) + Criteria

We create a new table to obtain the aggregate statistics (mean, median, and standard deviation of the sum) for our data, grouping them by criteria, skewness, and repetition. This new table does not include the simulated individual unit total scores, so we used an inner join structured query language command to merge the original data table and the aggregate statistics so that all data are merged into one table. We next calculate our statistics of interest, the S-score and M-score.

Standardized Scores

The most important metric in the RBSS is how users define risk groups. The current process for both the 129th RQW and the 163rd ATKW is using a standardized score about a mean, which we call the S-score, where *sum* is an individual organizational unit's total score and *mean(sum)* and *stdev(sum)* are the mean and standard deviation of all organizational unit total scores under a wing.

$$S - score = \frac{sum - mean(sum)}{stdev(sum)}$$

For comparison, we also consider the M-score, which replaces *mean(sum)* with the median of all organizational unit total scores under a wing or *median(sum)* in the S-score calculation.

$$M-score = \frac{sum - median(sum)}{stdev(sum)}$$

As we demonstrate in Chapter 4, the S-score can be affected by the non-symmetry of criteria responses. The M-score is invariant to this non-symmetry. After both the S-score and M-score are calculated for each row, we create a metric called data dominance for both the S-score and M-score, which identifies where the data lies relative to the mean or median. If the S-score/M-score is greater than 0, then it will be defined as upper or right-tail dominance. If the S-score/M-score is less than 0, then it will be defined as lower or left-tail dominance. If the S-score/M-score

is 0, which means *sum* is the same as *mean(sum)/median(sum)*, then it will be defined as center or central dominance.

To demonstrate how the S-score and M-score are affected by the pattern of Likert scale responses (for example right or left dominant), we compare the differences between the percentages of sums with upper data dominance to those with lower data dominance. As shown in the next chapter, the S-scores are affected by the tail-dominance of criteria responses while the M-scores are invariant. To determine which factors (i.e., organizational units, criteria, Likert scale values and skewness) affect the non-symmetry of upper and lower percentages, we perform a multiple regression analysis. From there, we present a Pareto Analysis indicating the order and relative magnitude of the significant variables. We use an alpha of 0.05 for this regression analysis.

Percent Contribution to the Response Variable

Although not directly associated with investigating how the S-score performs in RBSS as discussed in Chapter 2, we noted that the values of the dichotomous variable (1 and an upper value) could be overinfluential in the sum of the criteria with respect to percent contribution. Consequently, we investigate varying the upper value of the dichotomous variable to determine which integer value (we assume that users will not be amiable to using fractional values to scale disparate criteria scores) will come closer to the optimal value of what each question should represent of the sum.

To perform this analysis, we use the mean and variance for each respective Likert scale to determine its actual percent contribution to the response. The mean (μ) or expected value (E(x)) of a discrete random variable is

$$\mu = E(x) = \sum x p(x)$$

where x is a random variable and p(x) is its associated probability. Multiplying each possible value of x by its associated probability and summing this product over all possible values of x returns the expected value (McClave et al., 2018). The variance of a discrete random variable is defined as the average of the squared distance of x from the mean and denoted as:

$$\sigma^2 = E[(x - \mu)^2] = \sum (x - \mu)^2 p(x)$$

The percent contribution of a particular criterion can be found using R^2 , which represents the percent of total variation that can be explained by the model.

% Contribution =
$$R^2 = \frac{Explained\ Variance}{Total\ Variance}$$

When using the same Likert scale, all criteria will have an equal probability of occurrence and contribute equally to the model. It is mathematically impossible for a 3-point, 4-point and 5-point Likert scale to contribute equally to a sum without using scaling. For example, if we want a 4-point Likert criterion to contribute equally to a 5-point Likert criterion, we will have to multiply the 4-point Likert criterion by 5 and divide the answer by 4. The more criteria the more complicated this scaling process will be, which makes it less likely for the user to adopt such an approach.

Dichotomous Variable

We can, however, explore the combination of a 3-, 4-, or 5-point Likert scale with a dichotomous variable as users try to force a dichotomous variable into another Likert scale. To match a 3-point scale, one could use the values 1 and 3. To match a 5-point scale, the values of 1 and 5 could be used. It appears from our data that the highest and lowest scores are commonly used to represent the responses "yes" and "no." However, no rule states what values must be used for a dichotomous variable.

For example, consider the case of 10 criteria, eight of which uses a 5-point Likert scale and two of which uses a dichotomous scale. With our assumption of equal contribution, the theoretical contribution of each criterion is 1/10 or 0.1. We can mathematically derive the actual contribution when using a combination of dichotomous and 5-point Likert scales, using the commonly used values of 1 and 5 to represent the dichotomous variable. If we want to know the percent contribution of the 5-point criteria, we calculate the explained variance in the numerator, which is the product of the number of 5-point criterion and its variance, and divide by the denominator, which is the sum of the product of the number of 5-point criteria and its variance and the product of the number of dichotomous criteria and its variance. If we want to know the percent contribution of the dichotomous criteria, we will replace the numerator with product of the dichotomous criteria and its variance, keeping the denominator the same. For a 5-point Likert scale, we can change the values used to represent the dichotomous variable and see if the percent contribution changes. For example, these 10 combinations can be used to represent the dichotomous variable: 1 and 2, 1 and 3, 1 and 4, 1 and 5, 2 and 3, 2 and 4, 2 and 5, 3 and 4, 3 and 5, and 4 and 5.

The process will be repeated varying the number of criteria and the ratio of dichotomous criteria to Likert criteria (3-point and 4-point). If the percent contribution, in fact, changes, we can define rules that will help the user select the most optimal values to represent the dichotomous variable such that all criteria are approximately equally represented in the sum. We discuss those results in Chapter 4.

Summary

In this chapter, we descriptively examine the RBSS data from the 129th RQW and 163rd ATKW to determine the range of values we need to use for our simulation. The variables of

interest include the number of organizational units, number of criteria, scales for each criterion, and the probability of success. Our response of interest is the sum of the criteria responses. We define the S-score and M-score and how we intend to analyze these values to ascertain how they are affected by criteria responses that are non-symmetric. Lastly, we discuss how we intend to investigate varying the upper value of a dichotomous variable and its effect relative to percent contribution to the entire sum of criteria responses. In the next chapter, we present the results of our simulation as well as the recommended range for a dichotomous variable.

IV. Analysis and Results

Chapter Overview

In this chapter, we discuss the results of our simulation and dichotomous variable investigation. We analyze our simulated data using the S-score, our proposed M-score and compare the results of the two standardized scores centering data about the mean and median, respectively. In addition, we reveal whether the values selected to represent the dichotomous variables change its percent contribution to the response variable. We also suggest the optimal range for a dichotomous variable for Likert scales of 3 to 5.

Standardized Scores

As a reminder, we use the following four variables in our simulation: the number of organizational units, number of criteria, scales for each criterion, and the probability of success. The number of units can be 10, 15 or 20. The number of criteria can have values of 5, 10, 15, 20, and 25. We use a 3-point, 4-point or 5-point Likert scale. The probability of success can be 0.1, 0.3, 0.5, 0.7, and 0.9. Simulation resulted in 2.5 million rows of data for 10 units, 3.75 million for 15 units, and 5 million for 20 units. We want to see where the data predominantly lies for our S-score/M-score, which could be below the mean/median, at the mean/median or above the mean/median. We group the results of our data by criteria and skewness, both of which have five possible values. This presents 25 simulation results for each organizational grouping (10, 15, or 20). Each is a culmination of 100,000 outcomes for 10 units, 150,000 for 15 units, and 200,000 for 20 units (since we have 10,000 reps per individual outcome).

When observing the S-score patterns, the percentage of data in the two tail groups are not equal and there are very few values present in the center group, which indicates the distribution is skewed when evaluating about the mean. In other words, the distribution is generally not

normal. Although the 129th RQW and 163rd ATKW never explicitly stated that they assumed a normal distribution, their process for grouping data into risk groups heavily suggests use of the Empirical Rule (Figures 3 and 6), which requires a normal distribution.

As described in Chapter 3, the M-score is a modification of the S-score. Instead of looking at where data lies relative to the mean, we consider where the data lies relative to the median. By definition, the median value is the 50th percentile of a dataset and splits the data in half. With the median centered at 0 for the M-score, it fixes the issue of data skewing in the S-score.

We present the results of the M-score alongside the S-score for side-by-side comparison. Due to the extensive number of pages (45) required to show all graphs (see Appendix D), we have selected one as an example. The selected graph (Figure 13) uses a 3-point Likert scale with 15 criteria and 20 units. It is chosen as an example since both the 129th RQW and 163rd ATKW use a 3-point Likert scale the most, the two averaged closest to 15 criteria in their RBSS, and both have closest to 20 organizational units for inspection. The skewness levels of 0.1 and 0.9 are mirror images of each other as are the skewness levels of 0.3 and 0.7. The data is less skewed as it approaches 0.5 where it finally resembles a normal distribution.

When using the S-score, data dominance for both tails are not equal; however, they are relatively equal when using the M-score. Also, using the M-score reveals a more defined central dominance group compared to the S-score. Tables 11-13 displays the data dominance percentages for both S-scores and M-scores for all graphs.



Figure 13. Data Dominance Comparison Between S-score and M-score

Table 11. Comparison of S-score and M-score for 3-Point Likert Scale

		I	able 1.	ı. Comp	arison (di 8-scoi	re ana M	-score for	3-Point 1	Likert Scal	le	
		Skewness	Distance							Difference between	S-score tail	M-score tail
Units	Criteria	Level	from Center	S-score < mean	S-score = mean	S-score > mean	M-score < median	M-score = median	M-score > median	median and mean	difference	difference
10	5	0.1	0.4	0.5083	0.05751	0.43419	0.31335	0.36976	0.31688	0.31225	0.07411	
10	5	0.3	0.2	0.50201	0.02907	0.46892	0.37618	0.24277	0.38105	0.2137	0.03309	0.00487
10	5	0.5	0	0.48859	0.02538	0.48603	0.38471	0.22927	0.38602	0.20389	0.00256	0.00131
10	5	0.7		0.47089	0.03047	0.49864	0.37874		0.3757	0.21509	0.02775	
10	5	0.9		0.43757	0.05442	0.50801	0.31907	0.36741	0.31352	0.31299		
10	10	0.1	0.4	0.51686	0.02998	0.45316	0.36262	0.26547	0.37191	0.23549	0.0637	0.00929
10	10	0.3	0.2	0.50065	0.02045	0.4789	0.40932	0.17863	0.41205	0.15818	0.02175	0.00273
10	10	0.5	0	0.48992	0.01868	0.4914	0.41892	0.1626	0.41848	0.14392	0.00148	0.00044
10	10	0.7	0.2	0.47857	0.02124	0.50019	0.41135	0.17822	0.41043	0.15698	0.02162	0.00092
10	10	0.9		0.45358	0.03207	0.51435	0.3719		0.36498	0.23105	0.06077	
10	15	0.1		0.51379	0.02591	0.4603	0.38733		0.3919			
10	15	0.3		0.50038	0.01668	0.48294	0.42769		0.42834	0.12729		
10	15	0.5			0.01469	0.49111	0.43365		0.43433	0.11733	0.00309	
10	15	0.7		0.48239	0.01698	0.50063	0.42761	0.14537	0.42702	0.12839	0.01824	
10	15	0.9		0.46019	0.02532	0.51449	0.39092		0.38609	0.19767	0.0543	
10	20	0.1		0.51386	0.02174	0.4644	0.40072		0.40404	0.1735		
10	20	0.3		0.50235	0.01397	0.48368	0.43528		0.43544	0.11531	0.01867	0.00016
10	20	0.5			0.01275	0.49516	0.44176		0.4425	0.10299		
10	20	0.7		0.48335	0.01467	0.50198			0.43704	0.11119		
10	20	0.9	0.4	0.46503	0.02088	0.51409	0.40605		0.40394	0.16913	0.04906	
10	25	0.1	0.4	0.51124	0.02041	0.46835	0.41254		0.41406	0.15299		
10	25	0.3		0.50163	0.01332	0.48505	0.44327	0.11365	0.44308	0.10033	0.01658	
10	25	0.5			0.01267	0.49423	0.44738		0.44751	0.09244	0.00113	
10	25	0.7		0.48834	0.01227	0.49939	0.44363		0.44426			
10	25	0.9		0.47108	0.0198	0.50912	0.41549		0.41255	0.15216		
15	5	0.1		0.51374	0.04537	0.44089	0.29082		0.28223	0.38158		
15	5	0.3		0.50731	0.02003	0.47267	0.34239		0.34583	0.29176		
15	5	0.5		0.48978	0.01849	0.49173	0.35157	0.29564	0.35279	0.27715		
15	5	0.7		0.47385	0.02086	0.50529	0.34483		0.34295	0.29136		
15	5	0.9		0.44036	0.04476	0.51488	0.28327	0.42694	0.28979	0.38218		
15	10	0.1		0.52003	0.02365	0.45631	0.32909		0.33671	0.31055		
15	10	0.3		0.50457	0.01339	0.48205	0.37753		0.37795	0.23113	0.02252	
15	10	0.5			0.013	0.49491	0.38579		0.38506			
15	10	0.7		0.48156	0.01306	0.50538		0.24535	0.37613	0.23229		
15	10	0.9		0.45415	0.02395	0.5219	0.33769		0.32729	0.31107	0.06775	
15	15	0.1		0.51949	0.01697	0.46353	0.35287	0.28827	0.35885	0.2713	0.05596	
15	15	0.3		0.50385	0.01161	0.48453	0.39207	0.21347	0.39446			
15	15	0.5			0.00998	0.49581	0.39977	0.20119	0.39904	0.19121	0.0016	
15	15	0.7		0.48434	0.01135	0.50431	0.39371	0.21414	0.39215	0.20279		
15	15	0.9		0.465	0.01713	0.51787	0.35995		0.35155		0.05287	
15	20	0.1		0.51736	0.01499	0.46765	0.36858		0.3721	0.24433	0.04971	
15	20	0.3		0.50429	0.00942	0.48629	0.40203		0.40298			
15	20	0.5			0.00873	0.49445	0.4093		0.4084	0.17357	0.00238	
15	20	0.7		0.48747	0.00937	0.50316	0.40377	0.19427	0.40197	0.1849		
15	20	0.9		0.46671	0.01475	0.51853	0.37271		0.36706	0.24548		
15	25	0.1		0.5154	0.01234	0.47226			0.38034	0.22839	0.04314	
15	25	0.3		0.50265	0.00828	0.48907	0.40795		0.4096			
15	25	0.5		0.49566	0.00725	0.49709	0.41415		0.4137	0.1649		
15	25	0.7		0.48701	0.00793	0.50505	0.41035		0.40853	0.1732	0.01804	
15	25	0.9		0.47081	0.013	0.51619			0.37754	0.22871	0.04538	
20	5	0.1		0.52031	0.03638	0.44332	0.32378		0.29394	0.3459		
20	5	0.3		0.50602	0.01672	0.47727	0.3724	0.25592	0.37169	0.2392	0.02875	
20	5	0.5		0.49219	0.01519	0.49263	0.38204		0.3831	0.21968		
20	5	0.7		0.47859	0.01673	0.50469	0.37147	0.25442	0.37412	0.23769	0.0261	
20	5	0.9			0.03595	0.52085	0.29429		0.32376			
20	10	0.1			0.02047	0.45899			0.36106			
20	10	0.3		0.50554	0.01046	0.484			0.40883			
20	10	0.5			0.00865	0.49501	0.41535		0.41609			
20	10	0.7			0.01025	0.50523			0.40758			
20	10	0.9			0.01861	0.52347			0.36206			
20	15	0.1		0.51948	0.01487	0.46566			0.38663			
20	15	0.3		0.50525	0.00781	0.48695			0.42602			
20	15	0.5			0.0072	0.49559			0.43158			
20	15	0.7		0.48633	0.00794	0.50574			0.42369			
20	15	0.9		0.46413	0.01307	0.52281	0.38897		0.38306			
20	20	0.1			0.01089	0.46856			0.40157			
20	20	0.3		0.5051	0.00707	0.48784			0.43502			
20	20	0.5			0.00672	0.49754			0.43953			
20	20	0.7			0.00762	0.50512			0.43379			
20	20	0.9		0.47086	0.01123	0.51792			0.39899			
20	25	0.1		0.51682	0.00928	0.47391	0.40996		0.41252			
20	25	0.3		0.50345	0.00588	0.49067	0.4417		0.44152			
20	25	0.5			0.00644	0.49621	0.44638		0.44708			
20	25	0.7			0.00602	0.50391	0.4423		0.44191			
20	25	0.9	0.4	0.47136	0.00951	0.51914	0.41305	0.17816	0.4088	0.16865	0.04778	0.00425

Table 12. Comparison of S-score and M-score for 4-Point Likert Scale

		1	able 12	z. Comp	arison (di 8-scoi	re and M	-score for	4-Point 1	Likert Scal	le	
		Skewness	Distance							Difference between	S-score tail	M-score tail
Units	Criteria	Level	from Center	S-score < mean	S-score = mean	S-score > mean	M-score < median	M-score = median	M-score > median	median and mean	difference	difference
10	5	0.1		0.53035	0.0293	0.44035	0.33718	0.30675	0.35607	0.27745	0.09	
10	5	0.3	0.2	0.49905	0.0251	0.47585	0.39641	0.20576	0.39783	0.18066	0.0232	0.00142
10	5	0.5	0	0.48804	0.02292	0.48904	0.40469	0.18828	0.40703	0.16536	0.001	0.0023
10	5	0.7	0.2	0.47701	0.02311	0.49988	0.39684	0.20771	0.39545	0.1846	0.02287	0.00139
10	5	0.9		0.44241	0.02842	0.52917	0.35305	0.31131	0.33564		0.08676	
10	10	0.1		0.51591	0.02425	0.45984	0.38609	0.22136	0.39255		0.05607	
10	10	0.3		0.49944	0.01616	0.4844	0.42728	0.14405	0.42867	0.12789	0.01504	
10	10	0.5	0	0.49286	0.01446	0.49268	0.43311	0.13326	0.43363	0.1188	0.00018	0.00052
10	10	0.7	0.2	0.48234	0.01616	0.5015	0.42694	0.14601	0.42705	0.12985	0.01916	0.0001
10	10	0.9		0.4594	0.02577	0.51483	0.39284	0.21874	0.38842		0.05543	
10	15	0.1		0.51362	0.02066	0.46572	0.40794	0.18133	0.41073		0.0479	
10	15	0.3		0.50182	0.01291	0.48527	0.43858	0.12229	0.43913		0.01655	
10	15	0.5		0.49433	0.01259	0.49308	0.44648	0.10713	0.44639		0.00125	
10	15	0.7		0.48597	0.01314	0.50089	0.44166	0.1176	0.44074		0.01492	
10	15	0.9		0.46683	0.02176	0.51141	0.41054	0.18281	0.40665		0.04458	
10	20	0.1		0.5107	0.01776	0.47154	0.42049	0.15688	0.42263		0.03916	
10	20	0.3		0.5002	0.01137	0.48843	0.44837	0.10404	0.44759		0.01177	
10	20	0.5			0.01067	0.49446	0.45382	0.09278	0.4534		0.00041	
10	20	0.7		0.48814	0.01188	0.49998	0.44742	0.10527	0.44731		0.01184	
10	20	0.9	0.4	0.47101	0.01783	0.51116		0.15983	0.41922		0.04015	
10	25	0.1			0.01671	0.47587	0.42713	0.14421	0.42866		0.03155	
10	25	0.3		0.50087	0.0101	0.48903	0.45312	0.09381	0.45307	0.08371	0.01184	
10	25	0.5		0.49516	0.0093	0.49554	0.45865	0.08299	0.45836		0.00038	
10	25	0.7		0.49104	0.01012	0.49884	0.45462	0.09107	0.45431		0.0078	
10	25	0.9		0.47645	0.01624	0.50731	0.43017	0.13994	0.42989		0.03086	
15	5	0.1		0.53857	0.01601	0.44543	0.29233	0.37773	0.32993		0.09314	
15	5	0.3		0.50655	0.01468	0.47877	0.36161	0.27227	0.36613		0.02778	
15	5	0.5		0.49397	0.01389	0.49214	0.37091	0.25497	0.37412		0.00183	
15	5	0.7		0.47977	0.01557	0.50467	0.36461	0.27125	0.36414		0.0249	
15	5	0.9		0.44499	0.01509	0.53992	0.32931	0.3775	0.29319		0.09493	
15	10	0.1		0.51898	0.01843	0.46259	0.35184	0.28671	0.36145		0.05639	
15	10	0.3		0.50379	0.01035	0.48586	0.39362	0.21219	0.39419		0.01793	
15	10	0.5			0.01044	0.49371	0.39902	0.20055	0.40043		0.00214	
15	10	0.7		0.48488	0.01112	0.504	0.39417	0.21337	0.39247		0.01912	
15	10	0.9		0.46306	0.01769	0.51925	0.36016	0.28929	0.35055		0.05619	
15	15	0.1		0.51641	0.01384	0.46975	0.37451	0.24843	0.37706		0.04666	
15	15	0.3		0.50446	0.0085	0.48704	0.40659	0.18678	0.40663		0.01742	
15	15	0.5		0.4957	0.00841	0.49589	0.41123	0.17639	0.41238		0.00019	
15	15	0.7		0.48735	0.00814	0.50451	0.40675	0.18751	0.40574		0.01716	
15	15	0.9		0.47037	0.01323	0.51639	0.37713	0.24936	0.37351		0.04602	
15	20	0.1		0.51433	0.01198	0.47369	0.38569	0.22692	0.38739		0.04064	
15	20	0.3		0.50239	0.00744	0.49017	0.4139	0.17185	0.41425		0.01222	
15	20	0.5			0.00665	0.49701	0.41878	0.16325	0.41797		0.00067	
15	20	0.7		0.48858	0.00732	0.5041	0.41504	0.17059	0.41437		0.01552	
15	20	0.9		0.47316	0.0121	0.51474	0.38953	0.22611	0.38437		0.04158	
15	25	0.1		0.51361	0.01097	0.47543	0.39505	0.20891	0.39604		0.03818	
15	25	0.3		0.50189	0.00737	0.49075	0.41993	0.15968	0.42039		0.01114	
15	25	0.5		0.49561	0.0068	0.49759	0.42391	0.15328	0.42281	0.14648	0.00198	
15	25	0.7		0.48921	0.00692	0.50387	0.41957	0.16146	0.41897	0.15454	0.01466	
15	25	0.9		0.47556	0.01074	0.5137	0.39738	0.20914	0.39348		0.03814	
20	5	0.1		0.54446	0.00806	0.44749	0.32117	0.31599	0.36285		0.09697	
20	5	0.3		0.50979	0.01056	0.47966	0.39448	0.20943	0.3961	0.19887	0.03013	
20	5	0.5		0.49565	0.00992	0.49443	0.40193	0.19393	0.40415		0.00122	
20	5	0.7		0.4816	0.01016	0.50825	0.39676	0.21013	0.39312		0.02665	
20	5	0.9			0.0088	0.54629	0.36298		0.31711		0.10137	
20	10	0.1			0.01355	0.4651		0.22837	0.38693			
20	10	0.3		0.50566	0.00829	0.48605		0.1498	0.42569		0.01961	
20	10	0.5			0.00724	0.49611		0.13825	0.43088		0.00055	
20	10	0.7			0.00821	0.5054		0.15068	0.42479		0.019	
20	10	0.9			0.01436	0.51966		0.22946	0.38543		0.05367	
20	15	0.1		0.52047	0.00976	0.46977	0.40477	0.18653	0.40871		0.0507	
20	15	0.3		0.50558	0.00619	0.48823			0.43881		0.01735	
20	15	0.5			0.00636	0.49724		0.11505	0.44233			
20	15	0.7		0.48906	0.00709	0.50386		0.1238	0.43799		0.0148	
20	15	0.9		0.47081	0.00881	0.52039		0.18612	0.40426		0.04958	
20	20	0.1			0.00835	0.47497		0.16378	0.41879		0.04172	
20	20	0.3		0.50407	0.00541	0.49053		0.10813	0.44584		0.01354	
20	20	0.5			0.0048	0.49831	0.45104		0.45078			
20	20	0.7			0.00578	0.50387		0.10748	0.44617		0.01351	
20	20	0.9		0.47525	0.00977	0.51499		0.16315	0.4174		0.03974	
20	25	0.1		0.51553	0.00767	0.47681	0.42551	0.14736	0.42714			
20	25	0.3		0.50315	0.00537	0.49149		0.09721	0.45089		0.01166	
20	25	0.5			0.00496	0.49764		0.0899	0.45472			
20	25	0.7			0.00514	0.50364		0.09822	0.45125		0.01242	
20	25	0.9	0.4	0.47736	0.00784	0.51481	0.42796	0.14649	0.42556	0.13865	0.03745	0.0024

Table 13. Comparison of S-score and M-score for 5-Point Likert Scale

		1	able 13	s. Comp	arison o	oi 8-scoi	e and M	score for	5-Point 1	Likert Scal	le	
		Skewness	Distance							Difference between	S-score tail	
Units	Criteria			S-score < mean	S-score = mean	S-score > mean	M-score < median	M-score = median	M-score > median	median and mean	difference	
10	5	0.1		0.51577	0.03188	0.45235	0.36246		0.3711	0.23456	0.06342	
10	5	0.3	0.2	0.50058	0.01931	0.48011	0.41153	0.17575	0.41272	0.15644	0.02047	0.00119
10	5	0.5	0	0.4905	0.01784	0.49166	0.41957	0.16181	0.41862	0.14397	0.00116	0.00095
10	5	0.7	0.2	0.47848	0.02027	0.50125	0.41264	0.17497	0.41239	0.1547	0.02277	0.00025
10	5	0.9		0.45563	0.03108		0.36915		0.36195		0.05766	
10	10	0.1		0.51303	0.02132		0.40343		0.40586		0.04738	
10	10	0.3	0.2	0.50056	0.01375	0.48569	0.43624	0.12777	0.43599	0.11402	0.01487	0.00025
10	10	0.5	0	0.49287	0.01295	0.49418	0.44285	0.11411	0.44304	0.10116	0.00131	0.00019
10	10	0.7	0.2	0.4838	0.01472	0.50148	0.43701	0.12704	0.43595	0.11232	0.01768	0.00106
10	10	0.9	0.4	0.46529	0.02246	0.51225	0.40515	0.19262	0.40223	0.17016	0.04696	0.00292
10	15	0.1	0.4	0.51113	0.01734	0.47153	0.41965	0.15747	0.42288	0.14013	0.0396	0.00323
10	15	0.3		0.49948	0.01159	0.48893	0.44707		0.44766		0.01055	
10	15	0.5		0.49449	0.01063	0.49488	0.45368		0.45321	0.08248	0.00039	
10	15	0.7		0.48838	0.01113	0.50049	0.44781		0.44792		0.01211	
10	15	0.9		0.4738	0.01734	0.50886	0.42214				0.03506	
10	20	0.1		0.50928	0.01586	0.47486	0.43183				0.03442	
10	20	0.3		0.50071	0.01028	0.48901	0.45603			0.07752	0.0117	
10	20	0.5	0	0.49635	0.00863	0.49502	0.45861	0.08296	0.45843	0.07433	0.00133	0.00018
10	20	0.7		0.49016	0.01027	0.49957	0.45446		0.45492	0.08035	0.00941	
10	20	0.9		0.47466	0.0166		0.43289			0.11788	0.03408	
10	25	0.1		0.50832	0.01456						0.0312	
10	25	0.3		0.50103	0.00925	0.48972	0.45999				0.01131	
10	25	0.5			0.00815						0.00131	
10	25	0.7		0.49129	0.00851	0.5002					0.00891	
10	25	0.9		0.48061	0.01482	0.50457	0.43947		0.43806		0.02396	
15	5			0.51896	0.02463	0.45641	0.32981			0.31075	0.06255	
15	5			0.50421	0.01374	0.48205	0.37571		0.37767	0.23288	0.02216	
15	5				0.01181	0.49339	0.38483				0.00141	
15	5			0.48057	0.01337	0.50605	0.37957				0.02548	
15	5			0.45591	0.02146		0.33635				0.06672	
15	10	0.1		0.51779	0.01379	0.46843	0.36813			0.24587	0.04936	
15	10	0.3		0.50285	0.00855	0.48861	0.4026		0.40281	0.18604	0.01424	
15	10	0.5			0.00899	0.49319	0.40727		0.40832		0.00463	
15	10	0.7		0.48644	0.00929	0.50427	0.40301			0.18424	0.01783	
15	10	0.9		0.46649	0.01474		0.37176		0.36769		0.05228	
15	15	0.1		0.51507	0.01222	0.47271	0.38592		0.38824		0.04236	
15	15	0.3		0.50073	0.00765	0.49163	0.41347			0.16377	0.0091	
15	15	0.5			0.00731	0.49481	0.41741				0.00307	
15	15	0.7		0.49083	0.00701	0.50215	0.41477		0.4146		0.01132	
15	15	0.9		0.47349	0.01179	0.51473	0.38836		0.38617	0.21368	0.04124	
15	20	0.1		0.51188	0.01061	0.47751	0.39495			0.19607	0.03437	
15	20	0.3		0.5034	0.00644	0.49016	0.42186				0.01324	
15	20	0.5		0.49731	0.00658		0.42428		0.42489		0.0012	
15	20	0.7		0.48963	0.00687	0.50349						
15	20	0.9		0.47728	0.00995	0.51277	0.39864		0.39599		0.03549	
15	25	0.1		0.51226	0.00915	0.47859	0.40301		0.40612		0.03367	
15	25	0.3		0.50201	0.00605	0.49195	0.42573		0.42689		0.01006	
15	25	0.5		0.49532	0.00547	0.49921	0.42973			0.13529	0.00389	
15	25	0.7		0.49209	0.00623	0.50168	0.42621			0.1412		
15	25	0.9		0.47865	0.01007	0.51127	0.40475				0.03262	
20	5			0.52192	0.02004		0.36252				0.06387	
20	5			0.50621	0.01079	0.48301	0.40626		0.40933		0.0232	
20	5			0.49604	0.00855	0.49541	0.41605				0.00063	
20	5			0.48405	0.00977	0.50619			0.40881	0.17233		
20	5				0.01937	0.52332	0.36101			0.25625	0.066	
20	10			0.51947	0.01157		0.40097					
20	10											
20	10	0.5		0.49781	0.0063							
20	10	0.7		0.48879								
20	10			0.47072	0.01037						0.0482	
20	15			0.51605	0.00799							
20	15			0.50344	0.006							
20	15	0.5		0.49699	0.00531						0.00072	
20	15			0.49015							0.0144	
20	15	0.9		0.4768	0.00757						0.03884	
20				0.514	0.00799						0.03599	
20	20											
20	20	0.5		0.49728			0.45686					
20	20			0.4927	0.00546							
20	20			0.47808	0.00721						0.03664	
20	25			0.51306								
20	25	0.3		0.50324								
20	25				0.0043							
20	25											
20	25	0.9	0.4	0.48148	0.00713	0.5114	0.43654	0.12769	0.43578	0.12056	0.02992	0.00076

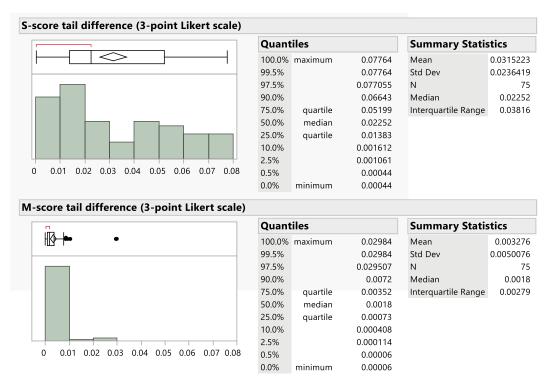


Figure 14. Difference Between Tail Values for S-score and M-score for 3-point Likert Scale

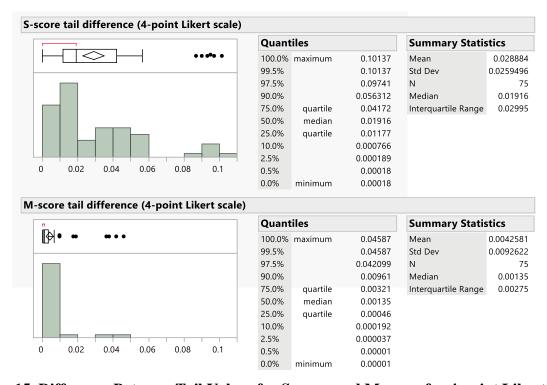


Figure 15. Difference Between Tail Values for S-score and M-score for 4-point Likert Scale

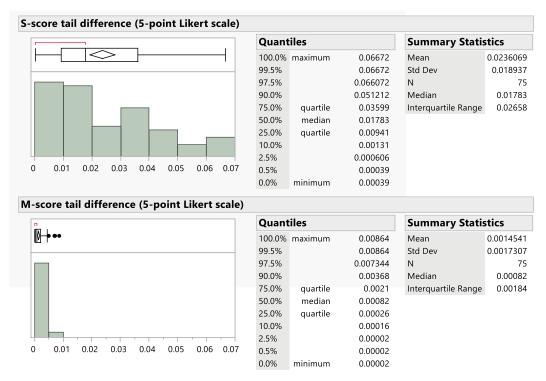


Figure 16. Difference Between Tail Values for S-score and M-score for 5-point Likert Scale

Figures 14-16 show the percentage differences between the tails, that is the absolute value of the difference between the percentage of scores above and below the center value, which is the mean for S-scores and median for M-scores, for Likert scales of 3, 4, and 5. For M-scores, the difference is very close to 0, while S-scores the difference is more variable. This indicates that S-scores are more impacted by skewness.

Table 14 summarizes the outliers present in Figures 14-16 for the percentage difference between the tail groups in the S-score and M-score. A checkmark under S-score or M-score (Table 14) indicates what combination of Likert scale, units, criteria, and skewness level resulted in an outlier for the percentage difference between the tail groups of the distribution. For the M-score, there are a total of 18 outliers: six from the 3-point Likert scale, eight from the 4-point Likert scale, and four from the 5-point Likert scale. For the S-score, there are a total of six outliers, all of which are outliers for the M-score from the 4-point Likert scale as well. It appears

that the overall pattern for outliers is the result of the extreme ends of skewness levels, 0.1 and 0.9, followed by lower numbers of criteria. The M-score tail difference have more outliers than the S-score since the majority of the values are 0 and even the slightest variation above 0 gets flagged as different.

Table 14. Outliers for M-score and S-score Tail Differences

Outliers	S-score	M-score	Likert scale	Units	Criteria	Skewness Level	S-score tail difference	M-score tail difference
1		✓	3	10	10	0.1	0.064	0.009
2		✓	3	15	5	0.1	0.073	0.009
3		✓	3	15	10	0.9	0.068	0.010
4		✓	3	15	15	0.9	0.053	0.008
5		✓	3	20	5	0.1	0.077	0.030
6		✓	3	20	5	0.9	0.078	0.029
7	\checkmark	✓	4	10	5	0.1	0.090	0.019
8	\checkmark	✓	4	10	5	0.9	0.087	0.017
9	\checkmark	✓	4	15	5	0.1	0.093	0.038
10	\checkmark	✓	4	15	5	0.9	0.095	0.036
11	\checkmark	✓	4	20	5	0.1	0.097	0.042
12	\checkmark	✓	4	20	5	0.9	0.101	0.046
13		✓	4	15	10	0.1	0.056	0.010
14		✓	4	15	10	0.9	0.056	0.010
15		✓	5	10	5	0.1	0.063	0.009
16		✓	5	10	5	0.9	0.058	0.007
17		✓	5	15	5	0.1	0.063	0.005
18		✓	5	15	5	0.9	0.067	0.007

To further investigate what factors can best explain the S-score tail difference, we conduct a regression analysis. We create a multiple linear model using the fit model function in in JMP Pro 15. We set the S-score tail difference as the response variable and add Likert scale, organizational units, criteria, and distance from the center as input variables into the model. The parameter estimates of the model (Table 15) indicate that units, which has a *p-value* of 0.2573, is not significant compared to α of 0.05. We remove the insignificant variable and rerun the model (Table 16). The standardized betas for each parameter are then ordered (in magnitude) to show the descending level of significance (Table 17).

Table 15. Parameter Estimates for S-score Tail Difference

Parameter Estimates										
Term	Estimate	Std Error	t Ratio	Prob> t	Std Beta					
Intercept	0.0250337	0.003999	6.26	<.0001 *	0					
Likert scale	-0.003958	0.000726	-5.45	<.0001 *	-0.13983					
Units	0.0001648	0.000145	1.14	0.2573	0.029113					
Criteria	-0.00103	8.378e-5	-12.30	<.0001 *	-0.31526					
Distance from center	0.1324358	0.003958	33.46	<.0001 *	0.857691					

Table 16. Parameter Estimates for S-score Tail Difference (Rerun)

Parameter Estimates										
Term	Estimate	Std Error	t Ratio	Prob> t	Std Beta					
Intercept	0.0275057	0.003357	8.19	<.0001 *	0					
Likert scale	-0.003958	0.000726	-5.45	<.0001 *	-0.13983					
Criteria	-0.00103	8.384e-5	-12.29	<.0001 *	-0.31526					
Distance from center	0.1324358	0.003961	33.44	<.0001 *	0.857691					

Table 17. Parameter Estimates for S-score Tail Difference (Sorted Standardized Beta)

Term	Estimate	Std Error	t Ratio	Prob> t	Std Beta
Intercept	0.0275057	0.003357	8.19	<.0001	0
Distance from center	0.1324358	0.003961	33.44	<.0001	0.857691
Criteria	-0.00103	8.38E-05	-12.29	<.0001	-0.31526
Likert scale	-0.003958	0.000726	-5.45	<.0001	-0.13983

Typically, the parameter estimate shows the effect on the response for every one unit increase in an input. However, since we will not observe a one-unit increase for the input distance from center, we divide the estimate by 10. In other words, for every 0.1-point increase in distance from center, the S-score tail difference will increase by 1.32%. To ensure we do not extrapolate from the bounds of our explanatory variable, our interpretation is limited to values of 0 to 0.4 for distance from center. For a 0 distance from the center or skewness levels of 0.5, the S-score tail difference will not change. For a 0.2 distance from the center or skewness levels of 0.3 and 0.7, the S-score tail difference will increase by 2.64%. For a 0.4 distance from the center or skewness levels of 0.1 and 0.9, the S-score tail difference will increase by 5.28%. These percentages can be interpreted as risk misclassifications. We discuss this in Chapter 5.

Next, we interpret the parameter estimate for criteria, which has a range of five to 25 in increments of five in our analysis. For every 1-point increase for the criteria, the S-score tail difference will decrease by 0.1%. We multiply this value by five to match our data. The S-score tail difference will decrease by 0.5% for 10 criteria, 1% for 15 criteria, 1.5% for 20 criteria, and 2% for 25 criteria.

Finally, we interpret the parameter estimate for Likert scale, which has a range of 3 to 5 in our analysis. For every 1-point increase for the input Likert scale, the S-score tail difference will decrease by 0.4%. The S-score tail difference will decrease by 0.4% for a 4-point Likert scale and 0.8% for a 5-point Likert scale.

In summary, as the number of criteria and Likert scale value increases, the S-score tail difference decreases; and as the data dominance disperses from the center toward the upper or lower bounds of the distribution, the S-score tail difference increases. As shown in Figure 17, the standardized beta values reveal that the distance from the center of the distribution has 2.7 times more impact on the S-score tail difference than criteria, which has 2.3 times more impact than Likert scale.

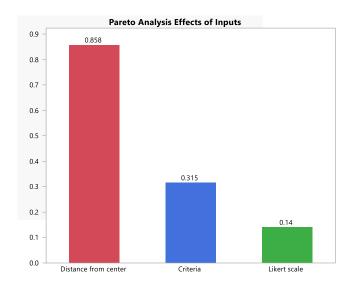


Figure 17. Pareto Analysis Effects of Inputs on S-score Tail Difference

Next, we examine the differences between data central dominance for the S-score and M-score. On average, the M-score has 20% more data central dominance on a 3-point scale, 17% more data central dominance on a 4-point scale, and 15% more data central dominance on a 5-point scale as shown in Figures 18-20. This indicates that M-score, which uses the median, does a better job classifying data in a middle group than an S-score, which uses the mean.

Coincidentally, the two outliers in Figures 18-20 representing the 3-, 4-, and 5-point Likert scales are for the combinations of 15 units, 5 criteria, and 0.1 and 0.9 skewness level. With a lower number of criteria, there is more variability and randomness.

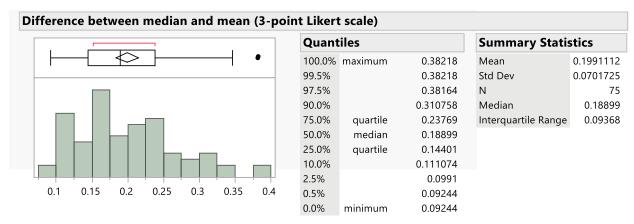


Figure 18. Difference Between Median and Mean for 3-point Likert Scale

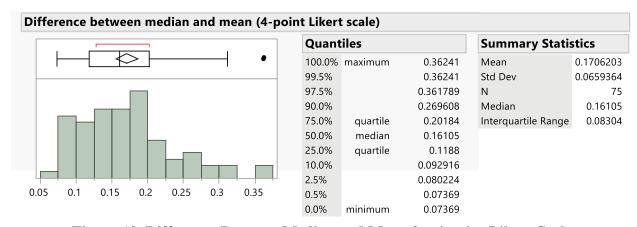


Figure 19. Difference Between Median and Mean for 4-point Likert Scale

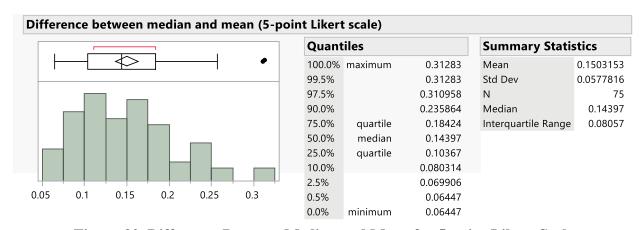


Figure 20. Difference Between Median and Mean for 5-point Likert Scale

Dichotomous Variable

We investigate whether changing the values used to represent a dichotomous variable would change its percent contribution to the response variable. We discover that the values selected to represent the dichotomous variable does skew the response variable. When trying to force a dichotomous variable into a Likert scale as both wings reviewed have done, intuitively the highest and lowest scores are used to represent the responses "yes" and "no." Unbeknownst to those who have never questioned such an approach, choosing the highest and lowest scores could end up giving that criterion a higher weight for the decision process. Through simulation, we realize that the values of the dichotomous numbers do not matter — what matters is the range between the values. For example, using the values of 1 and 5 to represent a dichotomous variable is the same as using the values of 2 and 6 as the range for both are the same (four).

Based on actual data, the 129th RQW and 163rd ATKW both use approximately 23% dichotomous responses when conducting their evaluation. Hence, we use this percentage to approximate the number of criteria theoretically would be dichotomous variables using our five chosen criteria values. For a five-criteria evaluation, one criterion (20%) can be dichotomous. For a 10-criteria evaluation, up to two can be dichotomous (20%). For a 15-criteria evaluation,

up to four (26.7%) can be dichotomous. For a 20-criteria evaluation, up to five (25%) can be dichotomous. For a 25-criteria evaluation, up to six (24%) can be dichotomous.

For a 5-point Likert scale, the maximum range between values is four since possible values are between 1 and 5. For a 4-point Likert scale, the maximum range between values is three. For a 3-point Likert scale, the maximum range between values is two. With this information, we calculate the actual percent contribution of each dichotomous variable and compare it with the ideal theoretical percent contribution. The range of the dichotomous variable that has the closest percent contribution to the theoretical value is the recommended range to use if we want each criterion to approximately contribute equally to the response. The detailed metrics for determining the optimal range of a dichotomous variable shown in Table 18 also reveal that the greater the range between the dichotomous variables, the greater its contribution to the response.

Table 18. Determining Optimal Range for a Dichotomous Variable by Likert Scale

					r a Dichou		
		Percentage	Ideal	Range	Contribution (5 pt)		
5	1	0.200	0.200	1	0.030	0.048	0.086
5	1	0.200	0.200	2	0.111	0.167	0.273
5	1	0.200	0.200	3	0.220	0.310	*
5	1	0.200	0.200	4	0.333		
10	1	0.100	0.100	1	0.014	0.022	0.040
10	1	0.100	0.100	2	0.053	0.082	0.143
10	1	0.100	0.100	3	0.111	0.167	*
10	1	0.100	0.100	4	0.182	*	*
10	2	0.200	0.100	1	0.015	0.024	0.043
10	2	0.200	0.100	2	0.056	0.083	0.136
10	2	0.200	0.100	3	0.110	0.155	*
10	2	0.200	0.100	4	0.167	*	*
15	1	0.067	0.067	1	0.009	0.014	0.026
15	1	0.067	0.067	2	0.034	0.054	0.097
15	1	0.067	0.067	3	0.074	0.114	*
15	1	0.067	0.067	4	0.125	*	*
15	2	0.133	0.067	1	0.009	0.015	0.027
15	2	0.133	0.067	2	0.036	0.055	0.094
15	2	0.133	0.067	3	0.074	0.108	*
15	2	0.133	0.067	4	0.118	*	*
15	3	0.200	0.067	1	0.010	0.016	0.029
15	3	0.200	0.067	2	0.037	0.056	0.023
15	3	0.200	0.067	3	0.073	0.103	*
15	3	0.200	0.067	4	0.111	*	*
15	4	0.267	0.067	1	0.111	0.017	0.030
	4			2			0.030
15	4	0.267 0.267	0.067	3	0.038	0.056	0.088 *
15			0.067		0.073	0.099	*
15	4	0.267	0.067	4	0.105	0.010	
20	1	0.050	0.050	1	0.007	0.010	0.019
20	1	0.050	0.050	2	0.026	0.040	0.073
20	1	0.050	0.050	3	0.056	0.087	*
20	1	0.050	0.050	4	0.095	*	*
20	2	0.100	0.050	1	0.007	0.011	0.020
20	2	0.100	0.050	2	0.026	0.041	0.071
20	2	0.100	0.050	3	0.056	0.083	*
20	2	0.100	0.050	4	0.091	*	*
20	3	0.150	0.050	1	0.007	0.011	0.021
20	3	0.150	0.050	2	0.027	0.041	0.070
20	3	0.150	0.050	3	0.055	0.080	*
20	3	0.150	0.050	4	0.087	*	*
20	4	0.200	0.050	1	0.008	0.012	0.021
20	4	0.200	0.050	2	0.028	0.042	0.068
20	4	0.200	0.050	3	0.055	0.078	*
20	4	0.200	0.050	4	0.083	*	*
20	5	0.250	0.050	1	0.008	0.013	0.022
20	5	0.250	0.050	2	0.029	0.042	0.067
20	5	0.250	0.050	3	0.055	0.075	*
20	5	0.250	0.050	4	0.080	*	*
25	1	0.230	0.030	1	0.005	0.008	0.015
25	1	0.040	0.040	2	0.020	0.032	0.015
25	1			3			*
25		0.040	0.040	4	0.045	0.070	*
	1		0.040		0.077	0.000	
25	2	0.080	0.040	1	0.005	0.009	0.016
25	2	0.080	0.040	2	0.021	0.033	0.058
25	2	0.080	0.040	3	0.045	0.068	*
25	2	0.080	0.040	4	0.074	*	
25	3	0.120	0.040	1	0.006	0.009	0.016
25	3	0.120	0.040	2	0.021	0.033	0.057
25	3	0.120	0.040	3	0.044	0.066	*
25	3	0.120	0.040	4	0.071	*	*
25	4	0.160	0.040	1	0.006	0.009	0.017
25	4	0.160	0.040	2	0.022	0.033	0.056
25	4	0.160	0.040	3	0.044	0.064	*
25	4	0.160	0.040	4	0.069	*	*
25	5	0.200	0.040	1	0.006	0.010	0.017
25	5	0.200	0.040	2	0.022	0.033	0.055
25	5	0.200	0.040	3	0.044	0.062	*
25	5	0.200	0.040	4	0.067	*	*
25	6	0.240	0.040	1	0.006	0.010	0.018
25	6	0.240	0.040	2	0.023	0.034	0.018
25	6	0.240		3		0.060	*
			0.040		0.044	*	*
25	6	0.240	0.040	4	0.065	T	*

Table 19. Optimal Range for a Dichotomous Variable

x-Point Likert Scale	Range	Values to Use
3	2	1 and 3
4	2	1 and 3, or 2 and 4
5	3	1 and 4, or 2 and 5

Table 19 summarizes the results from the analysis performed in Table 18. A dichotomous variable for a 3-point Likert scale should have a range of two using values of 1 and 3. A dichotomous variable for a 4-point Likert scale should have a range of two using values of 1 and 3 or 2 and 4. A dichotomous variable for a 5-point Likert scale should have a range of three using values of 1 and 4 or 2 and 5. While it is quite possible to develop an equation and determine the exact range to use for dichotomous variables, the method would be too complicated to be adopted in the field. The aforementioned simplified method works for determining the values of the dichotomous variable if the interested response is the sum of dichotomous variables and same-scaled Likert variables where it is desirable for all inputs to be of relatively equal weight. Figures 21-23 show the actual contribution for different dichotomous value ranges against its ideal contribution.

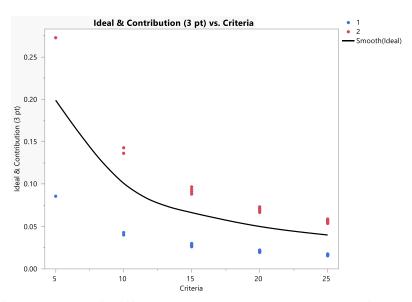


Figure 21. Contribution of Different Dichotomous Value Ranges for 3-pt Likert Scale

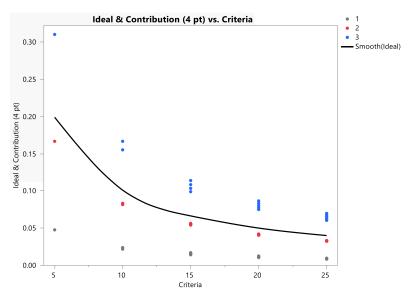


Figure 22. Contribution of Different Dichotomous Value Ranges for 4-pt Likert Scale

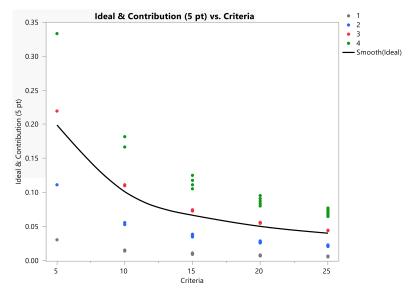


Figure 23. Contribution of Different Dichotomous Value Ranges for 5-pt Likert Scale

Summary

In this chapter, we analyze our data using the S-score, the M-score and compare the results of the two standardized scores centering data about the mean and median, respectively. Using the S-score skews the data dominance in the tails of the distribution and misclassifies data that should be in the center group. The M-score fixes the shortfalls of the S-score. We reveal the values selected to represent the dichotomous variables change its percent contribution to the response. While the values themselves do not matter, the range between the two do. The greater the range between the two values, the greater its contribution to the response. We reveal the optimal range for a dichotomous variable for Likert scales of 3 to 5. Our results reveal that the structure of the two RBSSs from the 129th RQW and 163rd ATKW are haphazard and wings may be unintentionally skewing the decision of inspection prioritization using the existing methods. In the next chapter, we present our conclusions and recommendations.

V. Conclusions and Recommendations

Chapter Overview

In this chapter, we present our conclusions of our research, its significance, recommendations for action and for future research. Our research investigates the viability of an inspection prioritization tool Air Combat Command (ACC) disseminates as an example that meets the intent of a risk-based sampling strategy in compliance with Air Force Instruction 90-201 (2018). A risk-based sampling strategy (RBSS) is a method used to inspect areas deemed most important by commanders requiring an independent assessment by the Inspector General (IG). Currently, individual wings build their own strategy, which may be considerably different as there is no uniform approach across the Air Force. The sponsor for this research, the 23rd Wing, requested that we evaluate the sample tool from the 129th Rescue Wing (RQW) obtained through its major command (MAJCOM), ACC, to see if it is a statistically sound process and recommend improvements as necessary.

Investigative Questions Answered

Our research investigates and examines the validity of the 129th RQW's current RBSS tools as well as its preceding tool from the 163rd Attack Wing, which is also a modified product possibly originating from the 366th Fighter Wing. We primarily focus on the 129th's tool as it is presented to the sponsor as an example of a properly executed RBSS. We identify any problems with the current tools and propose improvements and possible solutions. While it is possible to create a new decision analysis tool, it is easier for users to adopt a process that is familiar, reducing the learning curve. Given that variants of the existing tool have proliferated throughout ACC as an example of an acceptable RBSS, our research focuses on enhancing an existing

product, making any necessary adjustments, and ensuring the method employed is statistically sound. Through our analysis, we answer our investigative questions:

- 1. What are the current processes different inspection teams use for their risk-based sampling strategy?
- 2. What is the validity of the current risk-based sampling strategies?
- 3. What may be done to correct or improve the current risk-based sampling strategies?
- 4. What may be a better tool to use for risk-based sampling strategy?
- 5. How can we create a product that is simple to use and explain to users?

Conclusions of Research

We determine that the 129th RQW's RBSS tool has potential for improvement. Its current structure identifies evaluation criteria, each with its own scale, and scores all organizational units under its command (wing). These individual criteria scores are summed for each organizational unit and then converted into a standardization metric where its mean and standard deviation are used to categorize units into four different risk groups to determine their inspection priority from highest to lowest risk using the color scheme red, amber, green, and blue, respectively.

Although never explicitly stated, the 129th RQW RBSS process appears to follow the Empirical Rule, which requires data to be normally distributed. However, our simulations suggest that the data does not follow a normal distribution. As such, the mean is influenced by skewness. Yet, the tool uses a standardized score about a mean, which we call an S-score in our analysis. By centering data classification about a mean that is generally not equal to 0, organizational units may be improperly classified as risky or vice versa when in fact they belong in between the two extreme groups. To prevent misclassification in the center, we propose an M-

score, which uses a standardized score about a median, equally dividing all organizational units in half.

With an M-score, aside from grouping scores in the upper and lower bounds, we can create a third group for standardized scores equal to the median as it would be inaccurate to force the middle point values into either the upper or lower value groups. In contrast, it is theoretically unlikely for the S-scores to equal the mean, which is an average and typically a fractional value (not a whole number). Not only does the M-score provide a more accurate comparison of organizational performance within a wing, but it also classifies units into three groups: poor, average, and good. By dividing performance into three groups, one can use the intuitive and familiar red, amber, and green traffic light status reporting to highlight organizational units that are the riskiest. Once the initial three groups are identified, wings can repeat the M-score process for the top and bottom groups to further stratify organizational unit performance where the bottom (hottest) of the red group will have top inspection priority and the top (coolest) of the green group could be a potential candidate for a waived vertical inspection.

In its current composition, the 129th RQW RBSS tool mixes and matches Likert scales for criteria ranging from 2 to 5 without consideration for standardization. A 3-point Likert scale is not the same as a 5-point Likert scale. Combining different scales changes its percent contribution to the variable of interest, which is the aggregate of an organizational unit's scores across all criteria. Assuming it is desirable to have each criterion contribute equally to the response variable and the users do not want to perform complicated scaling, the simple fix is to consistently use the same Likert scale.

The tool's defined criteria scale is different than the scores inspectors used in their assessment. While most of the criteria are written to be scored on a 5-point Likert scale, most

criteria typically display 3-point Likert responses, which indicates inspectors favor a 3-point Likert scale. This is likely due to the high turnover rate of IG staff positions where one may have breadth and not depth for (or some basic knowledge of) different organizational units across the wing, making it manageable to distinguish among poor, average, and good, but challenging to differentiate between performance beyond that. As such, we recommend a 3-point Likert scale be used as it is simple and again follows the intuitive red, amber, and green traffic light status reporting that leaders are accustomed to seeing and using.

Although some criteria in the RBSS tool have theoretical Likert scales, their presented responses are dichotomous. While we generally do not recommend combining different Likert scales, there are occasions where a valued criteria may only have dichotomous responses as evident in our descriptive analysis. Instead of throwing out a potentially valuable criteria, we consider its impact to the response variable if matched to the same Likert scale other criteria use. Unlike the standard 3-, 4-, and 5-point Likert scales, the range between dichotomous variables are not set when matching a same-point Likert scale.

We investigated whether the values of the dichotomous response mattered if it is desirable for all input variables to have relatively equal contribution to the response variable. We discover that the ranges between the dichotomous values matter more than the values themselves because if two values have the same range, they have the same percent contribution. In addition, the greater the range between the dichotomous variables, the greater its contribution to the response. While it is mathematically feasible to determine a fractional value for the optimal range of a dichotomous variable, a complicated method will unlikely be employed in the field. To simplify the approach, we approximate the optimal range of a dichotomous variable, which

depends on the Likert scale used. For a 3-point and 4-point Likert scale, the optimal range for the dichotomous values is two. For a 5-point Likert scale, the optimal range is three.

Significance of Research

The Air Force does not have a standardized approach for creating an RBSS, using the current best attempt as the gold standard without consideration for statistical soundness of the process. Through descriptive and inferential statistical analysis, we define rules that can help Air Force wings build a simple intuitive model that meets the intent of an RBSS without being overly cumbersome and complicated. We recognize that organizational unit performance may not be normally distributed, which makes it difficult to standardize scores about a mean without data dominance classification errors and propose standardizing scores about a median, allowing the tails of distributions to have roughly equal percentages of data dominance, and a more prominent central dominance group. The approach of using a median instead of a mean is a simple adjustment that can be performed using the current tool of choice, an Excel spreadsheet.

While wings may have a defined scale to use for their evaluation, our analysis reveals that most criteria assessments have three responses, resembling a 3-point Likert scale and indicating a mismatch between theoretical and actual scales. Instead of imposing a scale that will not be used in practice, we recommend using a 3-point Likert scale as most people can differentiate between poor, average, and good.

While no research indicates what values should be chosen to represent a dichotomous variable, we found that the percent contribution increases as the range between the two chosen values increases. For roughly equal contribution, we should use the values 1 and 3 to represent dichotomous variables for 3- and 4-point Likert scales and the values 1 and 4 to represent dichotomous variables for 5-point Likert scales.

Recommendations for Action

We recommend using one Likert scale throughout the RBSS to maintain the integrity of equal contribution to the response variable. Based on our findings, we recommend using a 3-point Likert scale. If dichotomous variables are used in conjunction with a chosen Likert scale, the range of the two dichotomous values matter – the greater the range, the greater its contribution. For simplicity, use a range of two (or values of 1 and 3) for 3- and 4-point Likert scales; use a range of three (or values of 1 and 4) for 5-point Likert scales. When comparing the sum of criteria scores among organizational units to determine risk groups, we recommend standardizing about a median and using an M-score to categorize into three risk groups.

Recommendations for Future Research

Our research focuses on improving an existing MAJCOM-approved RBSS tool by examining data dominance patterns. We reasonably assume that all criteria are independent in our analysis, but perhaps additional analysis can be performed for criteria that have correlation. We assume that the user prefers each criterion to contribute equally to the response variable and did not consider changing the weighting factor of some criteria that may be deemed more important to the decision process. While we do discuss using a combination of a 3-point Likert scale and dichotomous variables in our research, our analysis does not consider mixing and matching different Likert scales within the same RBSS tool. We examine data for primarily two organizations, the 129th RQW and 163rd ATKW. While both fall under ACC, neither are activeduty units. Other MAJCOMs and active-duty units may have a different approach to their RBSS and their approved RBSSs can potentially be evaluated and compared to the ones in our analysis. Additionally, we focus on enhancing the univariate model used by the 129th RQW as it is the one

ACC provided and do not further evaluate the more complicated bivariate model used by the 163rd ATKW, which could potentially be further researched.

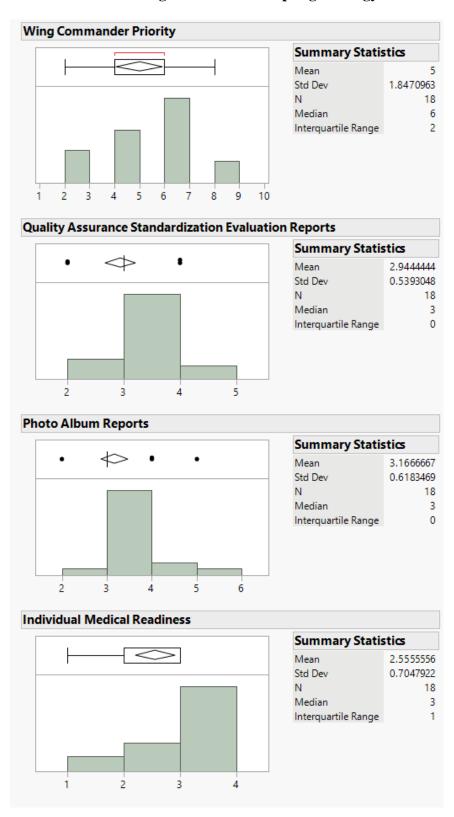
Summary

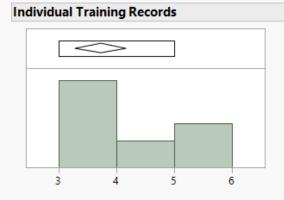
In conclusion, we recommend that wings use an M-score to standardize about a median so that data dominance in the tail groups would be relatively equal and there would be a more prominent central dominance group. This allows wings to categorize organizational units into three groups: good (green), average (amber), and poor (red). Following this same principle, we recommend that wings consistently use a 3-point Likert scale to evaluate their selected criteria in their RBSS as we find it is the most used scale. If combining a dichotomous variable with a 3-point Likert scale, the optimal range between the two variables is two so we should use the values 1 and 3, assuming it is desirable for all criteria to contribute approximately equally to the sum of criteria scores.

Appendix A. 129th Rescue Wing Risk-Based Sampling Strategy

Α	В	C	В	E	F	G		н 1		K	L	М	N	0	P	Q	F
IGI Member:		Taute		Taute	Johnston	Johnston	Tat	ekawa Morg		Taute	Taute	Johnston	Johnston	Taute	Blancas	Blancas	
v number = inspect/higher vis gh number = no inspect/lower vis reen = recommended by 90-201 ue = complete	go	Gonkulator nkulated on:	30-Dec-19					numboer (d fob) imprav samod incr rod numboe	4:5 ero-4								
				Deathern	OA Constitued	Disease Alberta Deserve	ts IMB	Laborat.	. AIICID101-	: USAP Findings (WAR	Chart	Control Foot	C-(D	OIL IC ICICCICAICICII	Manualita	No. destable	PITe
Criteria	(we	Wing CC P ighted order of r		Readiness Reporting	QA Stan/Eval Reports	Photo Album Repor (Functional assess SAVs, AARS, meeti mins, etc.)	ments,	Individu Trning Records	I Amn/IGI Personal Obs	CIMB, USAP Tracker, MICT)	Surveys	Continual Eval Results	Safety Reports	QIWG/GIGC/SAIC/CII/ SII/HHQ Concerns for the last year	End	Non-deployable monthly roster from FSF	Passi Peree
Scaling Criteria	priori 3= 4=	ty 2= ery low	olumn C × 2)	1=very low readiness 2=low readiness 3=neutral readiness 4=high readiness 5=very high readiness	reports 2=no reports 3=neutral	1= very unfavorable reports 2=unfavorat reports 3=neutral/nc reports 4=favorable reports 5=very favorable reports	ble below o 90% 3 above	2=80- report 3=90% & "AFSC	2=unfavorable obs 3=neutral obs red 4=favorable obs 5=very favorable obs	os 1= very unfavorable obs 2=unfavorable obs 3=neutralfno obs 4=favorable obs 5=very favorable obs	1=improveme nt needed 2=caution 3=no survey 4=adequate 5=excellent	2=no CE 3=medium risk 4=	1= very unfavorable 2=unfavorable 3=neutral/none 4=favorable 5=very favorable	1=3 or more trend(s)/note(s) exist 2=1-2 trend(s)/note(s) exist 3=no trend(s)/note(s) exist		1= very unfavorable 2=unfavorable 3=neutral/none 4=favorable 5=very favorable	pero
nits																	
S																	
G Staff (OGV) BS				D											-		
BRQS				Α													
RQS				т													
RQS																	
G Staff XS		—— <u></u>		Α											-		-
3				-													-
F				М													
G Staff				Α													
3				s													
•				1 -													
:				K													
				E													
IG TF				D													
۵		T		V	W	×	V	7 00	6R 6C 6D	AF AF	46	āH.	61 6.1	0K 01	am l	ăN	٥٢
gh number = no inspect/lower vis reen = recommended by 90-201	S Blancas	T Morgan	U Trujillo	V Mann	W Taute	X Taute	Morgan	Z AA MICT MICT Open Critical Obs Obs	AB AC AD Taute MICTT- MICTT- MICTT- 0 Obs 1 Obs 2 Obs	- MICT # of MICT # of overdue SAC obs assessment Suspense	MICT # of missing CAPs for	overdue ove olosure CAI suspense sus	Taute MS#of IGEMS rdue #of overdue pense CAP	AK AL IGEMS # of IGEMS # of Crit Sig deficiency deficiencie s s	AM	AN	AC
ow number = inspect/higher vis igh number = no inspect/lower vis Green = recommended by 90-201		ARCNet ESI Training Rep Currency (CBRNE &	Trujillo R Waivers		Commander/Pro C gram Mgr off E station > 4 F	Taute Tit & Sig Inspection leficiencies from 106 RQW & 144 FW for the	Morgan MICT Complianc	MICT MICT Open Critical Obs Obs	Taute MICTT- MICTT- MICTT-	- MICT # of MICT # of overdue obs assessment Suspense s/validation dates	MICT # of missing CAPs for Obs	IGEMS # of IGE overdue ove olosure CAI suspense sus dates dat	Taute MS#of IGEMS rdue #of Overdue pense CAP	IGEMS#of IGEMS#of Crit Sig deficienci deficiencie es s	AM	AN	A
ow number = inspectifrigher vis sigh number = no inspectiflower vis sieen = recommended by 90-201 silue = complete	Performance Reports/ACAs timeliness 1= very unfavorable 2=unfavorable	ARCNet ESI Training Rep Currency (CBRNE & SARC) NA for thi period du to inadequa accoutab	Trujillo Rivers In our waivers 3= waiver(s)	Mann Time Since Last Inspection (CCIP/UEI) 1= > 2 years 2= 1 yr < time < 2 yrs	Commander/Pro C gram Mgr off station > 4 F months or DSG I l=yes to above 1 S=no to above 2 3	Taute Zit & Sig Inspection Jeficiencies from 106	Morgan MICT Complianc	MICT MICT Open Critical Obs Obs	Taute MICTT- MICTT- MICTT- 0 Obs 1 Obs 2 Obs CT Metrics Assessment incl e metrics trics	- MICT # of MICT # of overdue obs assessment Suspense s/validation dates	MICT # of missing CAPs for Obs	IGEMS # of IGE overdue ove olosure CAI suspense sus dates dat	Taute MS # of IGEMS Indue # of overdue Dense ECDs GGEMS Metrics Asse metrics ics ics	IGEMS#of IGEMS#of Crit Sig deficienci deficiencie es s	AM	AN	AL
rumber = Inspectiviliphervis phrumber = notipectiliprervis seen = recommended by 90-201 us = complete Criteria Scaling Criteria	Performance Reports/ACAs timeliness 1= very unfavorable 2=unfavorable 3=neutral/non 4=favorable	ARCNet ESI Training Rep Currency (CBRNE & SARC) NA for thi period du to inadequa accoutab	Trujillo Rivers In our waivers 3= waiver(s)	Mann Time Since Last Inspection (CCIP/UEI) 1=> 2 years 2= 1 yr < time < 2 yrs 3 = 6 mo < time < 1 yr	Commander/Pro C gram Mgr off station > 4 F months or DSG I l=yes to above 1 S=no to above 2 3	Taute Cit & Sig Inspection Deficiencies from 106 GW & 144 FW for the ast 2 years Imany deficiencies Is little to no.	MICT Compliance percentag e: 95- 100%=5 90-	MICT MICT Open Critical Obs Obs IGI M 1= very unfavorable m 3=neutral / NO m 4=favorable metravorable metrav	Taute MICTT- MICTT- MICTT- 0 Obs 1 Obs 2 Obs CT Metrics Assessment incl e metrics trics	- MICT # of MICT # of overdue obs assessment Suspense s/validation dates	MICT # of missing CAPs for Obs	IGEMS # of IGE overdue ove closure CAI suspense dat suspense dat IGI 1= very unfavorable 2= unfavorable met ric- 4= favorable metric-	Taute MS # of IGEMS Indue # of overdue Dense ECDs GGEMS Metrics Asse metrics ics ics	IGEMS#of IGEMS#of Crit Sig deficienci deficiencie es s	SUM	BISK	A
r number = inspecifivilipher visigh number = no inspecifive evisige even = recommended by 90-201 tue = complete Criteria Scaling Criteria	Performance Reports/ACAs timeliness 1= very unfavorable 2=unfavorable 3=neutral/non 4=favorable	ARCNet ESI Training Rep Currency (CBRNE & SARC) NA for thi period du to inadequa accoutab	Trujillo Rivers In our waivers 3= waiver(s)	Mann Time Since Last Inspection (CCIP/UEI) 1=> 2 years 2= 1 yr < time < 2 yrs 3 = 6 mo < time < 1 yr	Commander/Pro C gram Mgr off station > 4 F months or DSG I l=yes to above 1 S=no to above 2 3	Taute Cit & Sig Inspection Deficiencies from 106 GW & 144 FW for the ast 2 years Imany deficiencies Is little to no.	MICT Compliance percentag e: 95- 100%=5 90-	MICT MICT Open Critical Obs Obs IGI M 1= very unfavorable m 3=neutral / NO m 4=favorable metravorable metrav	Taute MICTT- MICTT- MICTT- 0 Obs 1 Obs 2 Obs CT Metrics Assessment incl e metrics trics	- MICT # of MICT # of overdue obs assessment Suspense s/validation dates	MICT # of missing CAPs for Obs	IGEMS # of IGE overdue ove closure CAI suspense dat suspense dat IGI 1= very unfavorable 2= unfavorable met ric- 4= favorable metric-	Taute MS # of IGEMS Indue # of overdue Dense ECDs GGEMS Metrics Asse metrics ics ics	IGEMS#of IGEMS#of Crit Sig deficienci deficiencie es s	SUM 0	BISK lighest Risk	A
rumber = Inspectivelyer vs ghrumber = not pectivelyer vs ghrumber = not pectively vs green = recommended by 90-201 us = complete Criteria Scaling Criteria	Performance Reports/ACAs timeliness 1= very unfavorable 2=unfavorable 3=neutral/non 4=favorable	ARCNet ESI Training Rep Currency (CBRNE & SARC) NA for thi period du to inadequa accoutab	Trujillo Rivers In our waivers 3= waiver(s)	Mann Time Since Last Inspection (CCIP/UEI) 1=> 2 years 2= 1 yr < time < 2 yrs 3 = 6 mo < time < 1 yr	Commander/Pro C gram Mgr off station > 4 F months or DSG I l=yes to above 1 S=no to above 2 3	Taute Cit & Sig Inspection Deficiencies from 106 GW & 144 FW for the ast 2 years Imany deficiencies Is little to no.	MICT Compliance percentag e: 95- 100%=5 90-	MICT MICT Open Critical Obs Obs IGI M 1= very unfavorable m 3=neutral / NO m 4=favorable metravorable metrav	Taute MICTT- MICTT- MICTT- 0 Obs 1 Obs 2 Obs CT Metrics Assessment incl e metrics trics	- MICT # of MICT # of overdue obs assessment Suspense s/validation dates	MICT # of missing CAPs for Obs	IGEMS # of IGE overdue ove closure CAI suspense dat suspense dat IGI 1= very unfavorable 2= unfavorable met ric- 4= favorable metric-	Taute MS # of IGEMS Indue # of overdue Dense ECDs GGEMS Metrics Asse metrics ics ics	IGEMS#of IGEMS#of Crit Sig deficienci deficiencie es s	SUM 0	RISK tighest Risk lighest Risk	A
rumber = Inspectivelyer visit promoter = Inspective visit promoter = Inspective visit ceen = recommended by 90-201 us = complete Criteria Scaling Criteria Scaling Criteria	Performance Reports/ACAs timeliness 1= very unfavorable 2=unfavorable 3=neutral/non 4=favorable	ARCNet ESI Training Rep Currency (CBRNE & SARC) NA for thi period du to inadequa accoutab	Trujillo Rivers In our waivers 3= waiver(s)	Mann Time Since Last Inspection (CCIP/UEI) 1=> 2 years 2= 1 yr < time < 2 yrs 3 = 6 mo < time < 1 yr	Commander/Pro C gram Mgr off station > 4 F months or DSG I l=yes to above 1 S=no to above 2 3	Taute Cit & Sig Inspection Deficiencies from 106 GW & 144 FW for the ast 2 years Imany deficiencies Is little to no.	MICT Compliance percentag e: 95- 100%=5 90-	MICT MICT Open Critical Obs Obs IGI M 1= very unfavorable m 3=neutral / NO m 4=favorable metravorable metrav	Taute MICTT- MICTT- MICTT- 0 Obs 1 Obs 2 Obs CT Metrics Assessment incl e metrics trics	- MICT # of MICT # of overdue obs assessment Suspense s/validation dates	MICT # of missing CAPs for Obs	IGEMS # of IGE overdue ove closure CAI suspense dat suspense dat IGI 1= very unfavorable 2= unfavorable met ric- 4= favorable metric-	Taute MS # of IGEMS Indue # of overdue Dense ECDs GGEMS Metrics Asse metrics ics ics	IGEMS#of IGEMS#of Crit Sig deficienci deficiencie es s	SUM 0 1	RISK tighest Risk lighest Risk lighest Risk lighest Risk	
rumber = Inspective/orevis frumber = notipective vis frumber = notipective vis een = recommended by 90-201 us = complete Criteria Scaling Criteria Signati (OGV) Signat	Performance Reports/ACAs timeliness 1= very unfavorable 2=unfavorable 3=neutral/non 4=favorable	ARCNet ESI Training Rep Currency (CBRNE & SARC) NA for thi period du to inadequa accoutab	Trujillo Rivers In our waivers 3= waiver(s)	Mann Time Since Last Inspection (CCIP/UEI) 1=> 2 years 2= 1 yr < time < 2 yrs 3 = 6 mo < time < 1 yr	Commander/Pro C gram Mgr off station > 4 F months or DSG I l=yes to above 1 S=no to above 2 3	Taute Cit & Sig Inspection Deficiencies from 106 GW & 144 FW for the ast 2 years Imany deficiencies Is little to no.	MICT Compliance percentag e: 95- 100%=5 90-	MICT MICT Open Critical Obs Obs IGI M 1= very unfavorable m 3=neutral / NO m 4=favorable metravorable metrav	Taute MICTT- MICTT- MICTT- 0 Obs 1 Obs 2 Obs CT Metrics Assessment incl e metrics trics	- MICT # of MICT # of overdue obs assessment Suspense s/validation dates	MICT # of missing CAPs for Obs	IGEMS # of IGE overdue ove closure CAI suspense dat suspense dat IGI 1= very unfavorable 2= unfavorable met ric- 4= favorable metric-	Taute MS # of IGEMS Indue # of overdue Dense ECDs GGEMS Metrics Asse metrics ics ics	IGEMS#of IGEMS#of Crit Sig deficienci deficiencie es s	SUM 01 01 01 01 01 01 01 01 01 01 01 01 01	RISK Highest Rick Highest Rick Highest Rick Highest Rick Highest Rick	
r number = Inspectifylgher vis gh number = no inspectifylgher vis gh number = no inspectifylgher vis een = recommended by 90-201 ue = complete Criteria Scaling Criteria 3 Staff (OGV) 35 SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS	Performance Reports/ACAs timeliness 1= very unfavorable 2=unfavorable 3=neutral/non 4=favorable	ARCNet ESI Training Rep Currency (CBRNE & SARC) NA for thi period du to inadequa accoutab	Trujillo Rivers In our waivers 3= waiver(s)	Mann Time Since Last Inspection (CCIP/UEI) 1=> 2 years 2= 1 yr < time < 2 yrs 3 = 6 mo < time < 1 yr	Commander/Pro C gram Mgr off station > 4 F months or DSG I l=yes to above 1 S=no to above 2 3	Taute Cit & Sig Inspection Deficiencies from 106 GW & 144 FW for the ast 2 years Imany deficiencies Is little to no.	MICT Compliance percentag e: 95- 100%=5 90-	MICT MICT Open Critical Obs Obs IGI M 1= very unfavorable m 3=neutral / NO m 4=favorable metravorable metrav	Taute MICTT- MICTT- MICTT- 0 Obs 1 Obs 2 Obs CT Metrics Assessment incl e metrics trics	- MICT # of MICT # of overdue obs assessment Suspense s/validation dates	MICT # of missing CAPs for Obs	IGEMS # of IGE overdue ove closure CAI suspense dat suspense dat IGI 1= very unfavorable 2= unfavorable met ric- 4= favorable metric-	Taute MS # of IGEMS Indue # of overdue Dense ECDs GGEMS Metrics Asse metrics ics ics	IGEMS#of IGEMS#of Crit Sig deficienci deficiencie es s	SUM 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0	RISK ligheat Risk ligheat Risk ligheat Risk ligheat Risk ligheat Risk	S
rumber = Inspectivelyers is promised in the pr	Performance Reports/ACAs timeliness 1= very unfavorable 2=unfavorable 3=neutral/non 4=favorable	ARCNet ESI Training Rep Currency (CBRNE & SARC) NA for thi period du to inadequa accoutab	Trujillo Rivers In our waivers 3= waiver(s)	Mann Time Since Last Inspection (CCIP/UEI) 1=> 2 years 2= 1 yr < time < 2 yrs 3 = 6 mo < time < 1 yr	Commander/Pro C gram Mgr off station > 4 F months or DSG I l=yes to above 1 S=no to above 2 3	Taute Cit & Sig Inspection Deficiencies from 106 GW & 144 FW for the ast 2 years Imany deficiencies Is little to no.	MICT Compliance percentag e: 95- 100%=5 90-	MICT MICT Open Critical Obs Obs IGI M 1= very unfavorable m 3=neutral / NO m 4=favorable metravorable metrav	Taute MICTT- MICTT- MICTT- 0 Obs 1 Obs 2 Obs CT Metrics Assessment incl e metrics trics	- MICT # of MICT # of overdue obs assessment Suspense s/validation dates	MICT # of missing CAPs for Obs	IGEMS # of IGE overdue ove closure CAI suspense dat suspense dat IGI 1= very unfavorable 2= unfavorable met ric- 4= favorable metric-	Taute MS # of IGEMS Indue # of overdue Dense ECDs GGEMS Metrics Asse metrics ics ics	IGEMS#of IGEMS#of Crit Sig deficienci deficiencie es s	SUM 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	RISK Highest Rick Highest Rick Highest Rick Highest Rick Highest Rick	S
rumber = Inspectivelyers is promised in processor of the community of the	Performance Reports/ACAs timeliness 1= very unfavorable 2=unfavorable 3=neutral/non 4=favorable	ARCNet ESI Training Rep Currency (CBRNE & SARC) NA for thi period du to inadequa accoutab	Trujillo Rivers In our waivers 3= waiver(s)	Mann Time Since Last Inspection (CCIP/UEI) 1=> 2 years 2= 1 yr < time < 2 yrs 3 = 6 mo < time < 1 yr	Commander/Pro C gram Mgr off station > 4 F months or DSG I l=yes to above 1 S=no to above 2 3	Taute Cit & Sig Inspection Deficiencies from 106 GW & 144 FW for the ast 2 years Imany deficiencies Is little to no.	MICT Compliance percentag e: 95- 100%=5 90-	MICT MICT Open Critical Obs Obs IGI M 1= very unfavorable m 3=neutral / NO m 4=favorable metravorable metrav	Taute MICTT- MICTT- MICTT- 0 Obs 1 Obs 2 Obs CT Metrics Assessment incl e metrics trics	- MICT # of MICT # of overdue obs assessment Suspense s/validation dates	MICT # of missing CAPs for Obs	IGEMS # of IGE overdue ove closure CAI suspense dat suspense dat IGI 1= very unfavorable 2= unfavorable met ric- 4= favorable metric-	Taute MS # of IGEMS Indue # of overdue Dense ECDs GGEMS Metrics Asse metrics ics ics	IGEMS#of IGEMS#of Crit Sig deficienci deficiencie es s	SUM 01	RISK Highest Flick Highest Flick Highest Flick Highest Flick Highest Flick Highest Flick Highest Flick Highest Flick	S gher gher ean =
r number = inspectifylphervis j h number = no inspectifylphervis j h number = no inspectifylphervis j een = recommended by 90-201 ue = complete Criteria Scaling Criteria Signatify (IGV) S	Performance Reports/ACAs timeliness 1= very unfavorable 2=unfavorable 3=neutral/non 4=favorable	ARCNet ESI Training Rep Currency (CBRNE & SARC) NA for thi period du to inadequa accoutab	Trujillo Rivers In our waivers 3= waiver(s)	Mann Time Since Last Inspection (CCIP/UEI) 1=> 2 years 2= 1 yr < time < 2 yrs 3 = 6 mo < time < 1 yr	Commander/Pro C gram Mgr off station > 4 F months or DSG Is 1=yes to above 2 S=no to above 3	Taute Cit & Sig Inspection Deficiencies from 106 GW & 144 FW for the ast 2 years Imany deficiencies Is little to no.	MICT Compliance percentag e: 95- 100%=5 90-	MICT MICT Open Critical Obs Obs IGI M 1= very unfavorable m 3=neutral / NO m 4=favorable metravorable metrav	Taute MICTT- MICTT- MICTT- 0 Obs 1 Obs 2 Obs CT Metrics Assessment incl e metrics trics	- MICT # of MICT # of overdue obs assessment Suspense s/validation dates	MICT # of missing CAPs for Obs	IGEMS # of IGE overdue ove closure CAI suspense dat suspense dat IGI 1= very unfavorable 2= unfavorable met ric- 4= favorable metric-	Taute MS # of IGEMS Indue # of overdue Dense ECDs GGEMS Metrics Asse metrics ics ics	IGEMS#of IGEMS#of Crit Sig deficienci deficiencie es s	SUM 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	RISK Highest Risk Highest Risk Highest Risk Highest Risk Highest Risk Highest Risk Highest Risk Highest Risk Mighest Risk	ghe- gheens
w number = Inspectivelyers vs ghrunder en cinpective vs chromatic entrement of the commended by 90-201 luse = complete Criteria Scaling Criteria Scaling Criteria state	Performance Reports/ACAs timeliness 1= very unfavorable 2=unfavorable 3=neutral/non 4=favorable	ARCNet ESI Training Rep Currency (CBRNE & SARC) NA for thi period du to inadequa accoutab	Trujillo Rivers In our waivers 3= waiver(s)	Mann Time Since Last Inspection (CCIP/UEI) 1=> 2 years 2= 1 yr < time < 2 yrs 3 = 6 mo < time < 1 yr	Commander/Pro C gram Mgr off station > 4 F months or DSG Is 1=yes to above 2 S=no to above 3	Taute Cit & Sig Inspection Deficiencies from 106 GW & 144 FW for the ast 2 years Imany deficiencies Is little to no.	MICT Compliance percentag e: 95- 100%=5 90-	MICT MICT Open Critical Obs Obs IGI M 1= very unfavorable m 3=neutral / NO m 4=favorable metravorable metrav	Taute MICTT- MICTT- MICTT- 0 Obs 1 Obs 2 Obs CT Metrics Assessment incl e metrics trics	- MICT # of MICT # of overdue obs assessment Suspense s/validation dates	MICT # of missing CAPs for Obs	IGEMS # of IGE overdue ove closure CAI suspense dat suspense dat IGI 1= very unfavorable 2= unfavorable met ric- 4= favorable metric-	Taute MS # of IGEMS Indue # of overdue Dense ECDs GGEMS Metrics Asse metrics ics ics	IGEMS#of IGEMS#of Crit Sig deficienci deficiencie es s	SUM 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0	RISK Highest Hink Highest Rink Highest Rink Highest Rink Highest Rink Highest Rink Highest Rink Highest Rink Highest Rink Highest Rink Highest Rink	Sighes gher gear
w number = inspecifylighers vis giphnumber = no numberollower vis even = recommended by 90-201 turn = complete Criteria Scaling Criteria Signature = complete Criteria	Performance Reports/ACAs timeliness 1= very unfavorable 2=unfavorable 3=neutral/non 4=favorable	ARCNet ESI Training Rep Currency (CBRNE & SARC) NA for thi period du to inadequa accoutab	Trujillo Rivers In our waivers 3= waiver(s)	Mann Time Since Last Inspection (CCIP/UEI) 1=> 2 years 2= 1 yr < time < 2 yrs 3 = 6 mo < time < 1 yr	Commander/Pro C gram Mgr off station > 4 F months or DSG Is 1=yes to above 2 S=no to above 3	Taute Cit & Sig Inspection Deficiencies from 106 GW & 144 FW for the ast 2 years Imany deficiencies Is little to no.	MICT Compliance percentag e: 95- 100%=5 90-	MICT MICT Open Critical Obs Obs IGI M 1= very unfavorable m 3=neutral / NO m 4=favorable metravorable metrav	Taute MICTT- MICTT- MICTT- 0 Obs 1 Obs 2 Obs CT Metrics Assessment incl e metrics trics	- MICT # of MICT # of overdue obs assessment Suspense s/validation dates	MICT # of missing CAPs for Obs	IGEMS # of IGE overdue ove closure CAI suspense dat suspense dat IGI 1= very unfavorable 2= unfavorable met ric- 4= favorable metric-	Taute MS # of IGEMS Indue # of overdue Dense ECDs GGEMS Metrics Asse metrics ics ics	IGEMS#of IGEMS#of Crit Sig deficiencie es s	SUM 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0	RISK lighest flisk lighest flisk	Sighes gher gear
w number = Inspective evision processor from the result of the result	Performance Reports/ACAs timeliness 1= very unfavorable 2=unfavorable 3=neutral/non 4=favorable	ARCNet ESI Training Rep Currency (CBRNE & SARC) NA for thi period du to inadequa accoutab	Trujillo Rivers In our waivers 3= waiver(s)	Mann Time Since Last Inspection (CCIP/UEI) 1=> 2 years 2= 1 yr < time < 2 yrs 3 = 6 mo < time < 1 yr	Commander/Pro C gram Mgr off station > 4 F months or DSG Is 1=yes to above 2 S=no to above 3	Taute Cit & Sig Inspection Deficiencies from 106 GW & 144 FW for the ast 2 years Imany deficiencies Is little to no.	MICT Compliance percentag e: 95- 100%=5 90-	MICT MICT Open Critical Obs Obs IGI M 1= very unfavorable m 3=neutral / NO m 4=favorable metravorable metrav	Taute MICTT- MICTT- MICTT- 0 Obs 1 Obs 2 Obs CT Metrics Assessment incl e metrics trics	- MICT # of MICT # of overdue obs assessment Suspense s/validation dates	MICT # of missing CAPs for Obs	IGEMS # of IGE overdue ove closure CAI suspense dat suspense dat IGI 1= very unfavorable 2= unfavorable met ric- 4= favorable metric-	Taute MS # of IGEMS Indue # of overdue Dense ECDs GGEMS Metrics Asse metrics ics ics	IGEMS#of IGEMS#of Crit Sig deficiencie es s	SUM 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	RISK Highest Hink Highest Rink Highest Rink Highest Rink Highest Rink Highest Rink Highest Rink Highest Rink Highest Rink Highest Rink Highest Rink	S ghes gher ean =
w number = inspectifyligher vis glip number = no inspectifyligher vis glip number = no inspectifyligher vis gireen = recommended by 30-201 lus = complete Criteria Scaling Criteria Scaling Criteria Scaling Criteria Scaling Criteria Scaling Criteria Mits SS	Performance Reports/ACAs timeliness 1= very unfavorable 2=unfavorable 3=neutral/non 4=favorable	ARCNet ESI Training Rep Currency (CBRNE & SARC) NA for thi period du to inadequa accoutab	Trujillo Rivers In our waivers 3= waiver(s)	Mann Time Since Last Inspection (CCIP/UEI) 1=> 2 years 2= 1 yr < time < 2 yrs 3 = 6 mo < time < 1 yr	Commander/Pro C gram Mgr off station > 4 F months or DSG Is 1=yes to above 2 S=no to above 3	Taute Cit & Sig Inspection Deficiencies from 106 GW & 144 FW for the ast 2 years Imany deficiencies Is little to no.	MICT Compliance percentag e: 95- 100%=5 90-	MICT MICT Open Critical Obs Obs IGI M 1= very unfavorable m 3=neutral / NO m 4=favorable metravorable metrav	Taute MICTT- MICTT- MICTT- 0 Obs 1 Obs 2 Obs CT Metrics Assessment incl e metrics trics	- MICT # of MICT # of overdue obs assessment Suspense s/validation dates	MICT # of missing CAPs for Obs	IGEMS # of IGE overdue ove closure CAI suspense dat suspense dat IGI 1= very unfavorable 2= unfavorable met ric- 4= favorable metric-	Taute MS # of IGEMS Indue # of overdue Dense ECDs GGEMS Metrics Asse metrics ics ics	IGEMS#of IGEMS#of Crit Sig deficiencie es s	SUM 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	BISK tighest Hisk	Sighes igher ean =
w number = Inspective five vising for number in on Inspective vising for number in on Inspective vising reen = recommended by 90-201 has = complete Criteria Scaling Criteria Scaling Criteria Scaling Criteria Scaling Criteria Scaling Criteria National Criteria Scaling Criteria	Performance Reports/ACAs timeliness 1= very unfavorable 2=unfavorable 3=neutral/non 4=favorable	ARCNet ESI Training Rep Currency (CBRNE & SARC) NA for thi period du to inadequa accoutab	Trujillo Rivers In our waivers 3= waiver(s)	Mann Time Since Last Inspection (CCIP/UEI) 1=> 2 years 2= 1 yr < time < 2 yrs 3 = 6 mo < time < 1 yr	Commander/Pro C gram Mgr off station > 4 F months or DSG Is 1=yes to above 2 S=no to above 3	Taute Cit & Sig Inspection Deficiencies from 106 GW & 144 FW for the ast 2 years Imany deficiencies Is little to no.	MICT Compliance percentag e: 95- 100%=5 90-	MICT MICT Open Critical Obs Obs IGI M 1= very unfavorable m 3=neutral / NO m 4=favorable metravorable metrav	Taute MICTT- MICTT- MICTT- 0 Obs 1 Obs 2 Obs CT Metrics Assessment incl e metrics trics	- MICT # of MICT # of overdue obs assessment Suspense s/validation dates	MICT # of missing CAPs for Obs	IGEMS # of IGE overdue ove closure CAI suspense dat suspense dat IGI 1= very unfavorable 2= unfavorable met ric- 4= favorable metric-	Taute MS # of IGEMS Indue # of overdue Dense ECDs GGEMS Metrics Asse metrics ics ics	IGEMS#of IGEMS#of Crit Sig deficiencie es s	SUM 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	RISK tighest Risk	Stone State
ov number = inspecthigher vis sigh number = no inspectiover vis Green = recommended by 30-201 Cluze = complete	Performance Reports/ACAs timeliness 1= very unfavorable 2=unfavorable 3=neutral/non 4=favorable	ARCNet ESI Training Rep Currency (CBRNE & SARC) NA for thi period du to inadequa accoutab	Trujillo Rivers In our waivers 3= waiver(s)	Mann Time Since Last Inspection (CCIP/UEI) 1=> 2 years 2= 1 yr < time < 2 yrs 3 = 6 mo < time < 1 yr	Commander/Pro C gram Mgr off station > 4 F months or DSG Is 1=yes to above 2 S=no to above 3	Taute Cit & Sig Inspection Deficiencies from 106 GW & 144 FW for the ast 2 years Imany deficiencies Is little to no.	MICT Compliance percentag e: 95- 100%=5 90-	MICT MICT Open Critical Obs Obs IGI M 1= very unfavorable m 3=neutral / NO m 4=favorable metravorable metrav	Taute MICTT- MICTT- MICTT- 0 Obs 1 Obs 2 Obs CT Metrics Assessment incl e metrics trics	- MICT # of MICT # of overdue obs assessment Suspense s/validation dates	MICT # of missing CAPs for Obs	IGEMS # of IGE overdue ove closure CAI suspense dat suspense dat IGI 1= very unfavorable 2= unfavorable met ric- 4= favorable metric-	Taute MS # of IGEMS Indue # of overdue Dense ECDs GGEMS Metrics Asse metrics ics ics	IGEMS#of IGEMS#of Crit Sig deficiencie es s	SUM 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	BISK tighest Hisk	gher f ean = ess Ri:

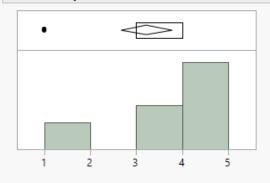
Appendix B. 129th Rescue Wing Risk-Based Sampling Strategy Criteria Statistics





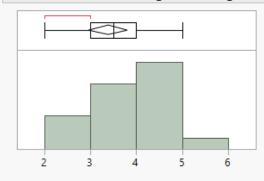
Summary Statistics Mean 3.7222222 Std Dev 0.8947925 N 18 Median 3 Interquartile Range 2

Airman/Inspections Directorate Personal Observations



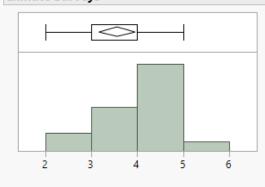
Summary Statistics					
Mean	3.2222222				
Std Dev	1.1143743				
N	18				
Median	4				
Interquartile Range	e 1				

Unit Self-Assessment Program Findings

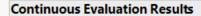


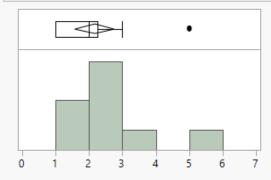
Summary Statistics				
Mean	3.3888889			
Std Dev	0.8498366			
N	18			
Median	3.5			
Interquartile Range	1			

Climate Surveys



Summary Statistics Mean 3.5555556 Std Dev 0.7838234 N 18 Median 4 Interquartile Range 1





Summary Statistics

 Mean
 2.1666667

 Std Dev
 1.2004901

 N
 18

 Median
 2

 Interquartile Range
 1.25

Safety Reports



Summary Statistics

Mean	3
Std Dev	0.3429972
N	18
Median	3
Interquartile Range	0

QIWG, GIGC, SAIC, CII, SII, HHQ, Concerns for last year



Summary Statistics

 Mean
 2.0555556

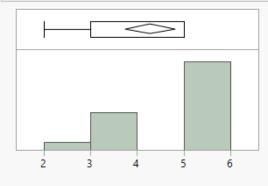
 Std Dev
 0.9375953

 N
 18

 Median
 2

 Interquartile Range
 2

Monthly End Strength



Summary Statistics

 Mean
 4.2777778

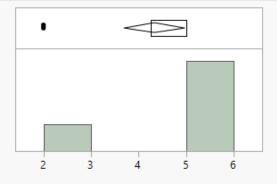
 Std Dev
 1.0740553

 N
 18

 Median
 5

 Interquartile Range
 2

Non-deployable monthly roster



Summary Statistics

Mean	4.3333333
Std Dev	1.2833779
N	18
Median	5
Interquartile Range	0.75

Performance reports/Airman Comprehensive Assessment Timelines



Summary Statistics

Mean	3.4444444
Std Dev	0.9217772
N	18
Median	4
Interquartile Range	2

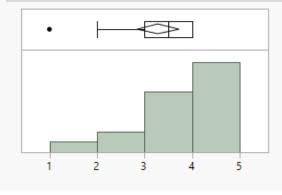
Waivers



Summary Statistics

Mean	2.3333333
Std Dev	0.9701425
N	18
Median	3
Interquartile Range	2

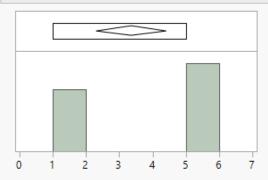
Time Since Last Inspection (CCIP/UEI)



Summary Statistics

Mean	3.2777778
Std Dev	0.8947925
N	18
Median	3.5
Interquartile Range	1

CC/PM Off Station > 4 months



Summary Statistics Mean 3.3529412

 Std Dev
 2.0291986

 N
 17

 Median
 5

 Interquartile Range
 4

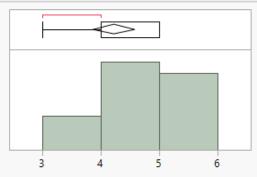
Crit/Sig Inspection Deficiencies from 106RQW/ 144FW for last 2 years



Summary Statistics

Mean	1.7777778
Std Dev	0.8084521
N	18
Median	2
Interquartile Range	1.25

MICT Compliance



Summary Statistics

 Mean
 4.2222222

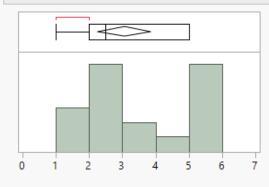
 Std Dev
 0.7320845

 N
 18

 Median
 4

 Interquartile Range
 1

IGI MICT Metrics Assessment



Summary Statistics

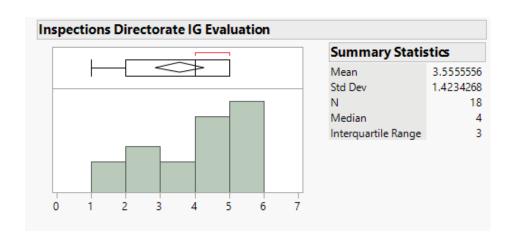
 Mean
 3.0555556

 Std Dev
 1.5893847

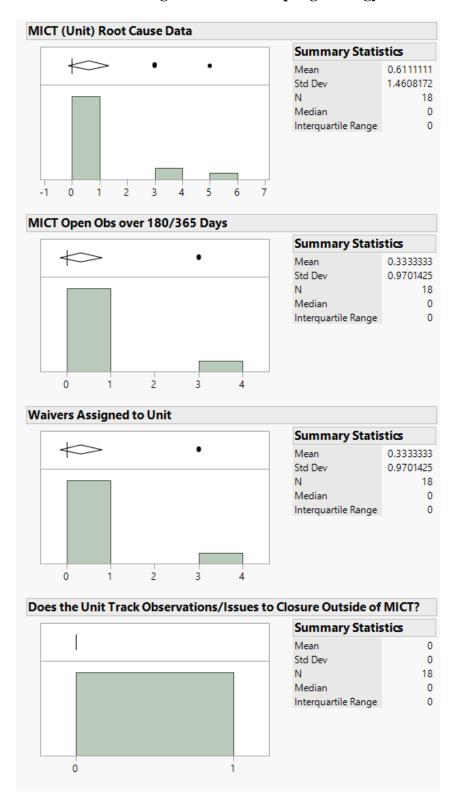
 N
 18

 Median
 2.5

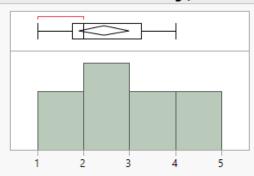
 Interquartile Range
 3



Appendix C. 163rd Attack Wing Risk-Based Sampling Strategy Criteria Statistics

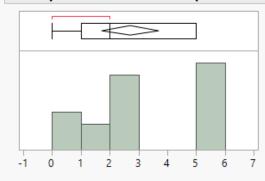


Personnel in Excessive Training (over 36 months) or awaiting training school (over 6 months)



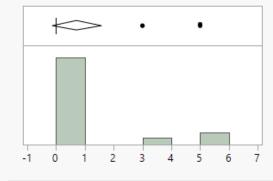
Summary Statistics				
Mean	2.444444			
Std Dev	1.0966378			
N	18			
Median	2			
Interquartile Range	1.5			

Time lapse since last vertical inspection



Summary Statistics				
Mean	2.7222222			
Std Dev	1.9942728			
N	18			
Median	2			
Interquartile Range	4			

IGEMS (Wing) Root Cause Data



Summary Statistics	
Mean	0.7222222
Std Dev	1.7083034
N	18
Median	0
Interquartile Range	0

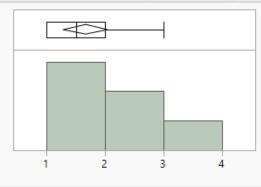
IGEMS Open Deficiencies over 180 Days



Summary Statistics	
Mean	0.3333333
Std Dev	0.9701425
N	18
Median	0
Interquartile Range	0



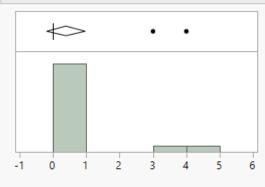
Unit Health (DEOCS/Inspection Findings)



Summary Statistics

Mean	1.6666667
Std Dev	0.766965
N	18
Median	1.5
Interquartile Range	1

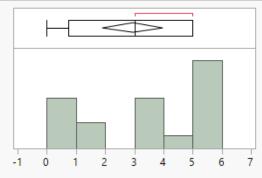
USAP/AARs



Summary Statistics

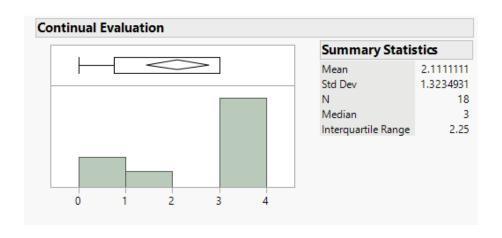
Mean	0.3888889
Std Dev	1.1447522
N	18
Median	0
Interquartile Range	0

Time Lapse since last external assessment (UAV/SAV)



Summary Statistics

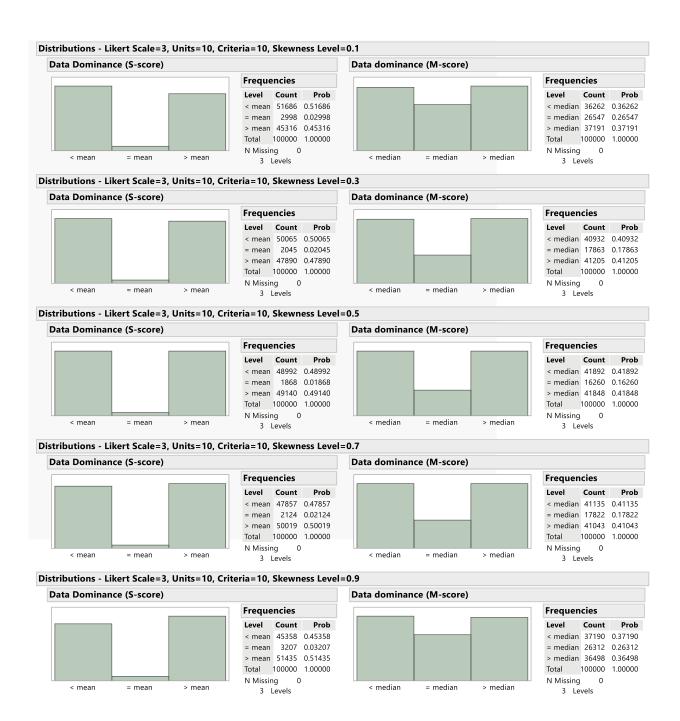
Mean	2.9444444
Std Dev	2.0714366
N	18
Median	3
Interquartile Range	4.25

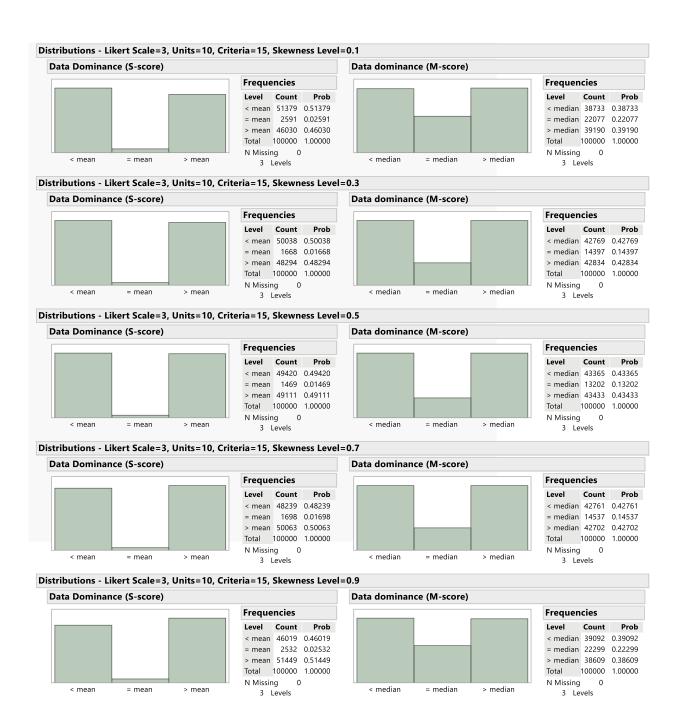


Appendix D. Simulation Graphs for Different Combinations of Likert Scale, Criteria, Skewness Level and Units

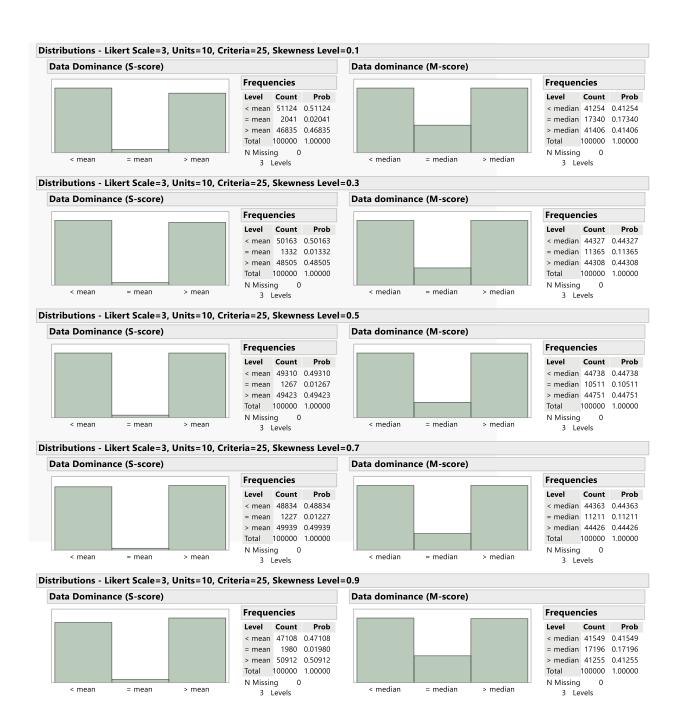


Note: In the graphs with 10 units, five criteria, and a 3-point Likert scale, there are missing sample sizes or N Missing for skewness levels of 0.1 and 0.9. The standard deviation is 0 for those data points (the simulation generated the exact value), causing the respective S-score and M-score to be undefined. We expect smaller sample sizes to result in more variability and anomalies.



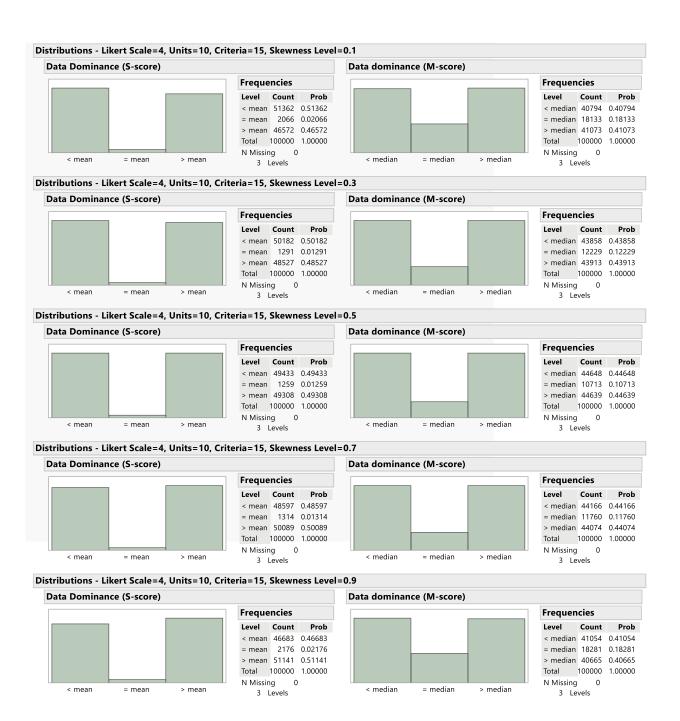




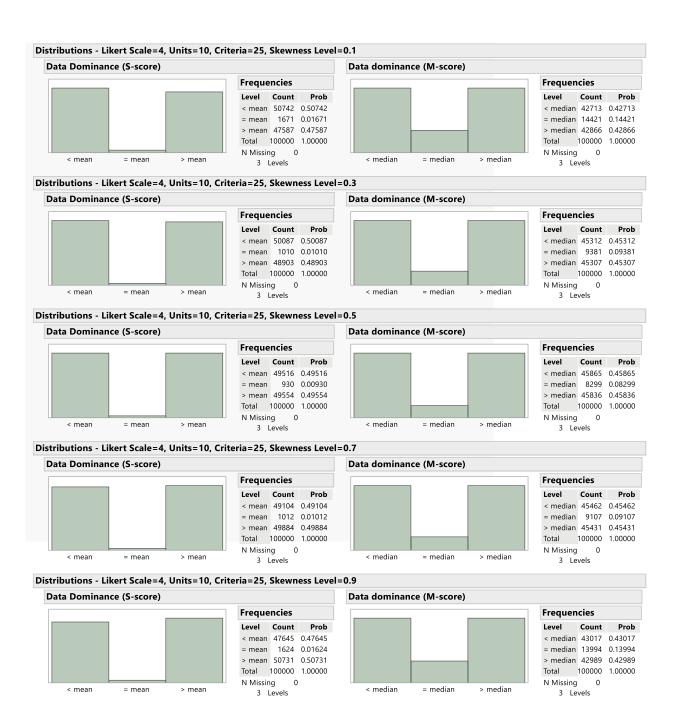


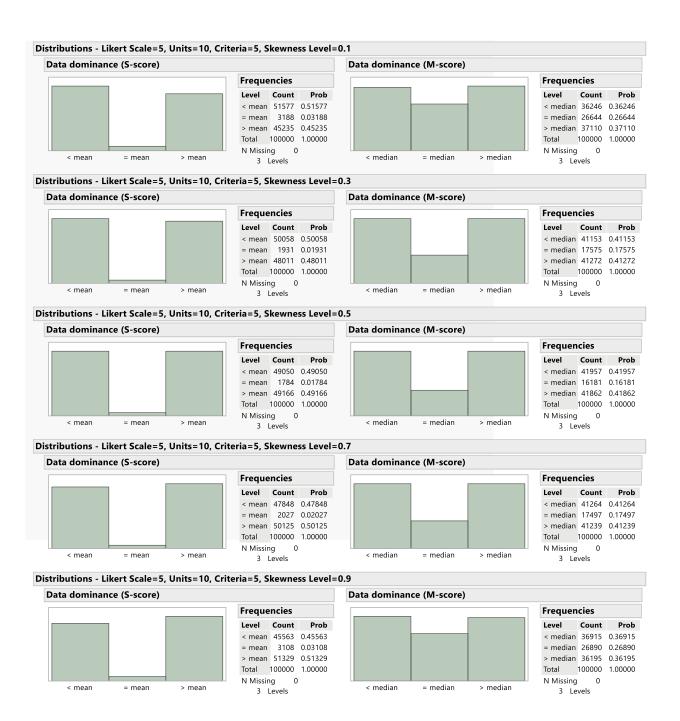


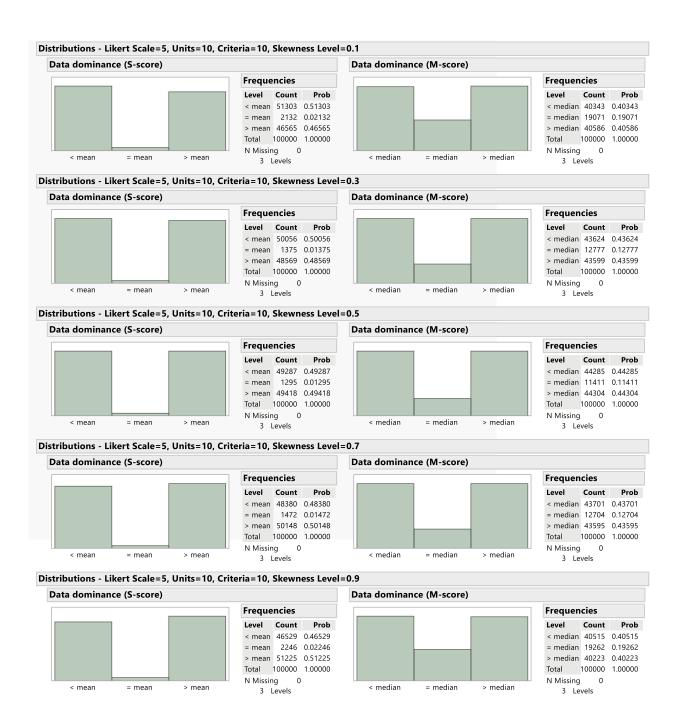








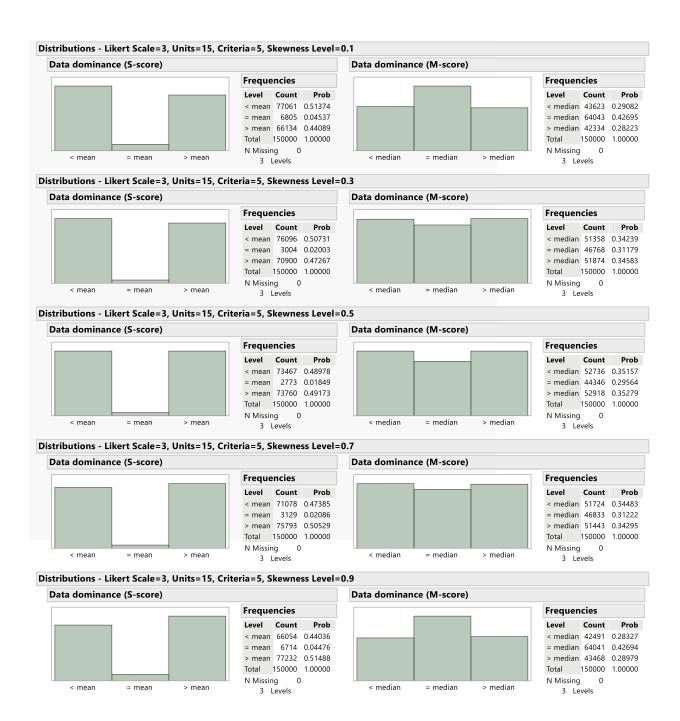


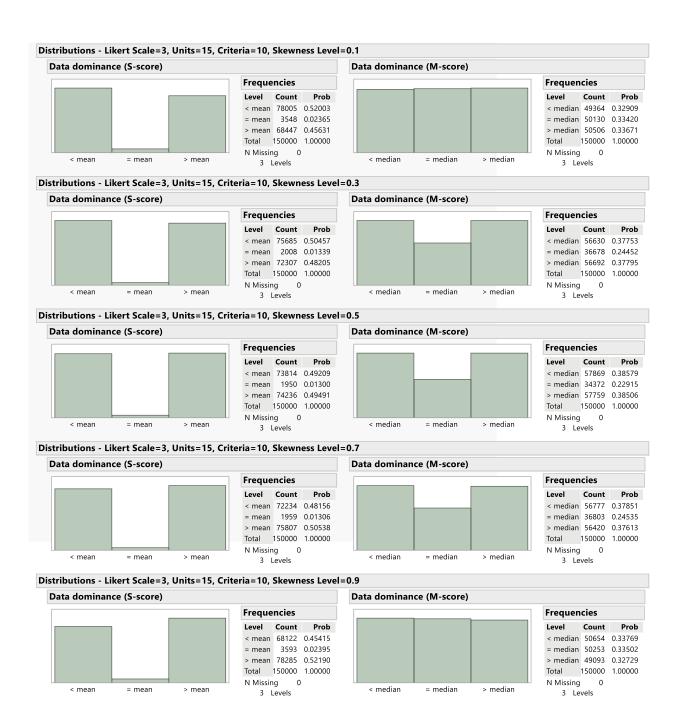


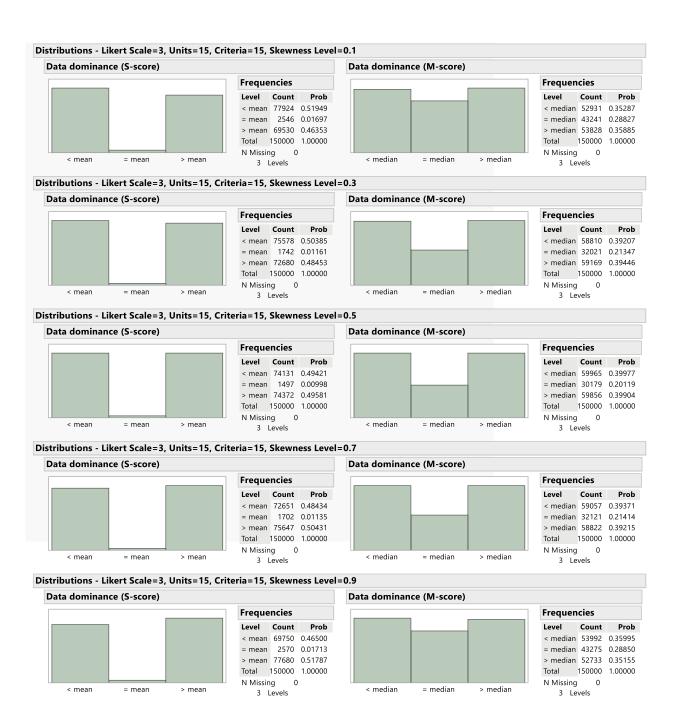


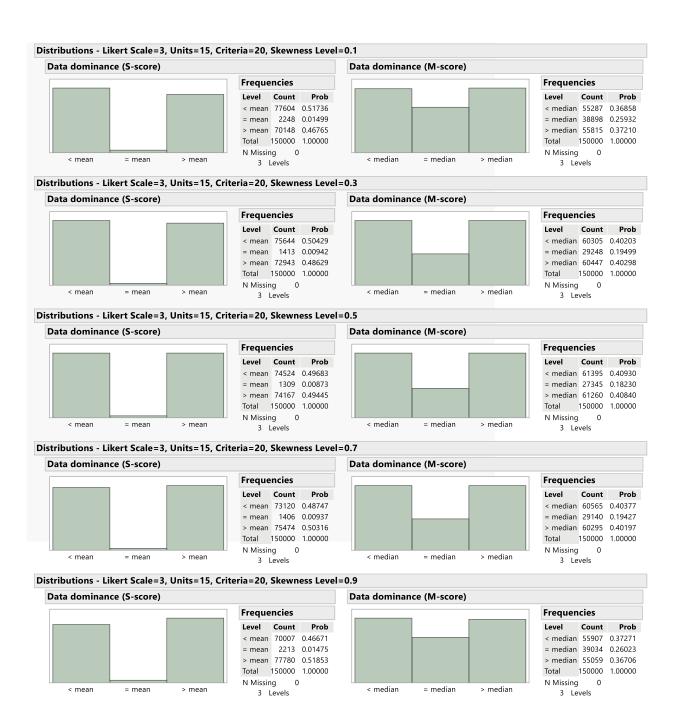


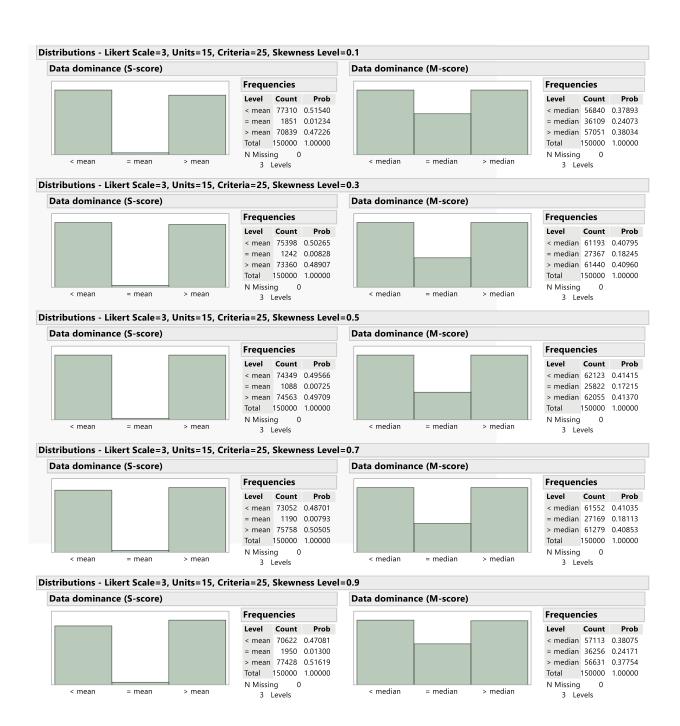


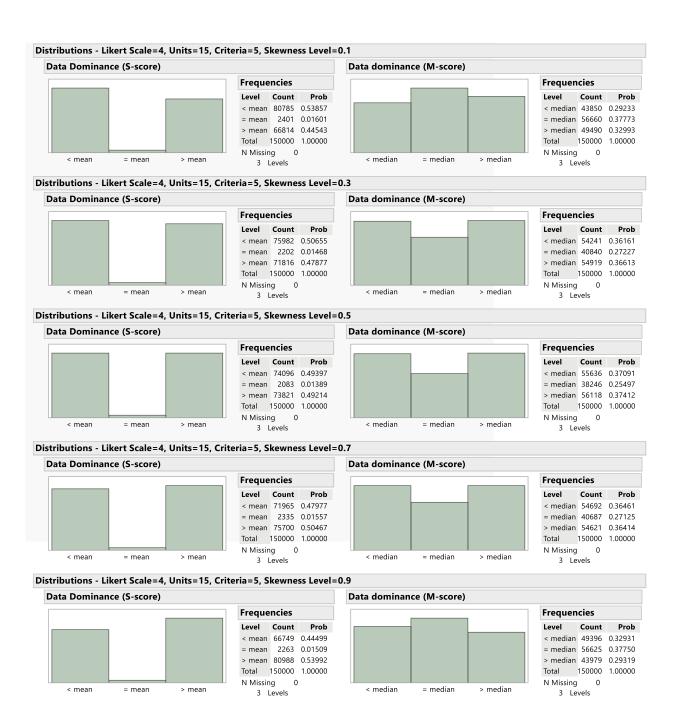


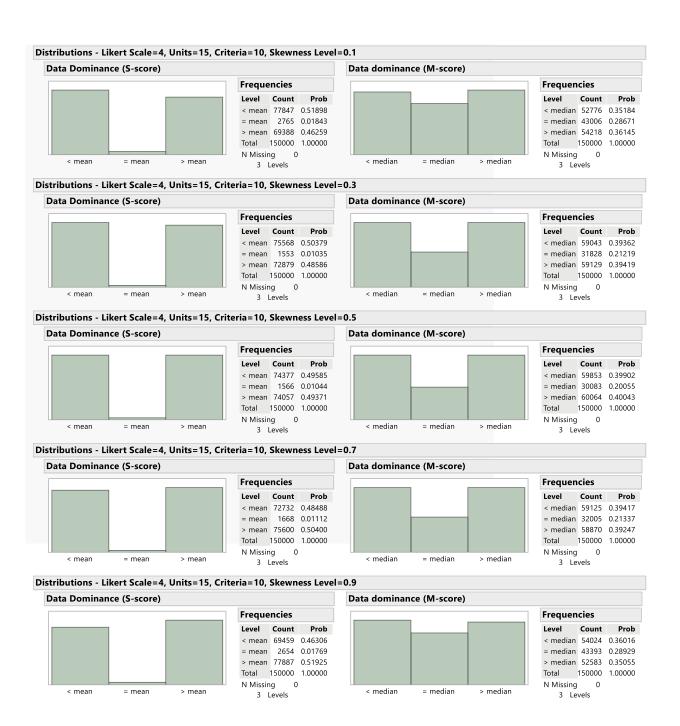


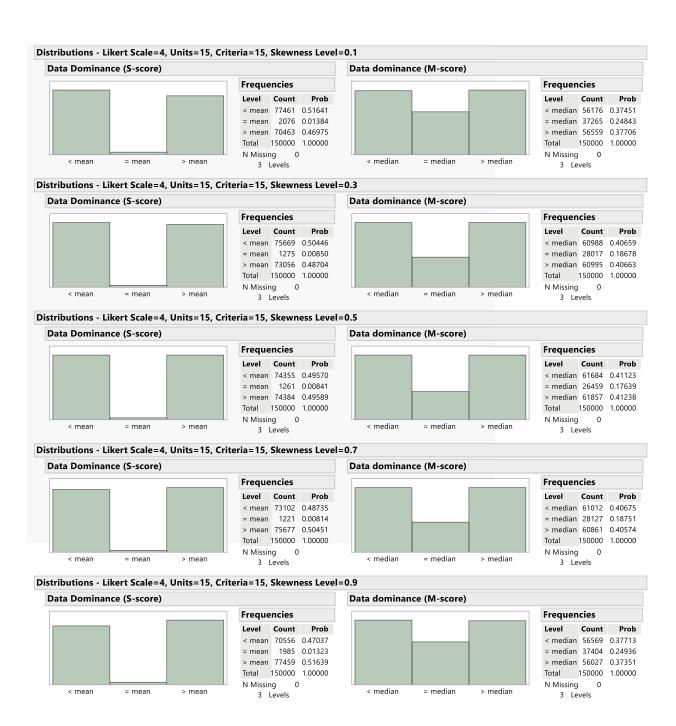


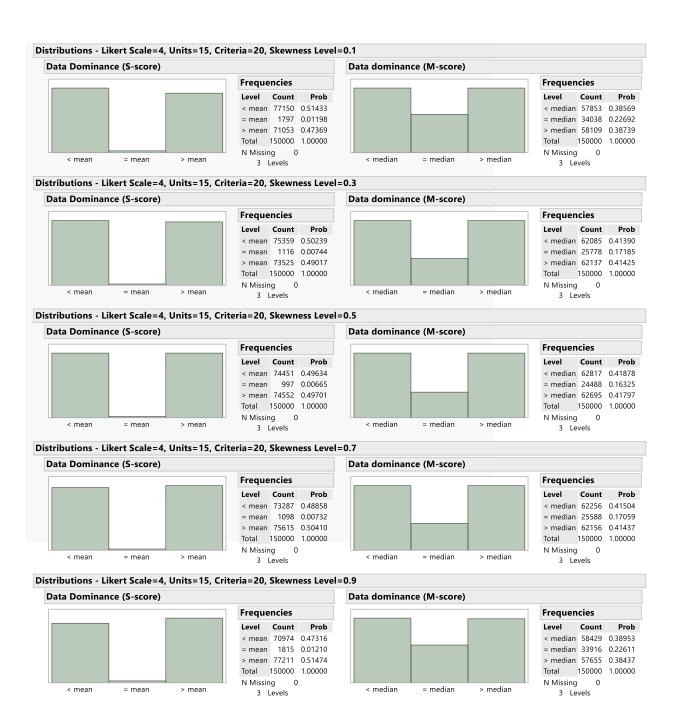


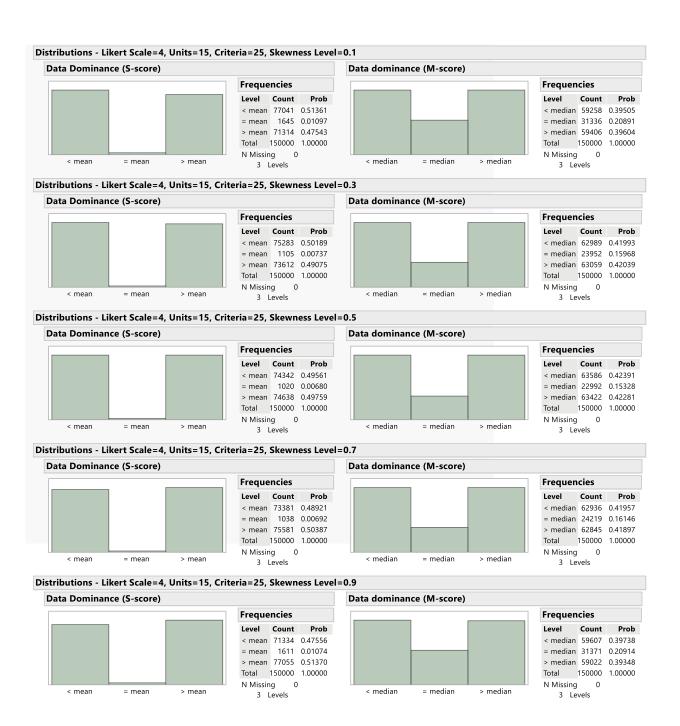




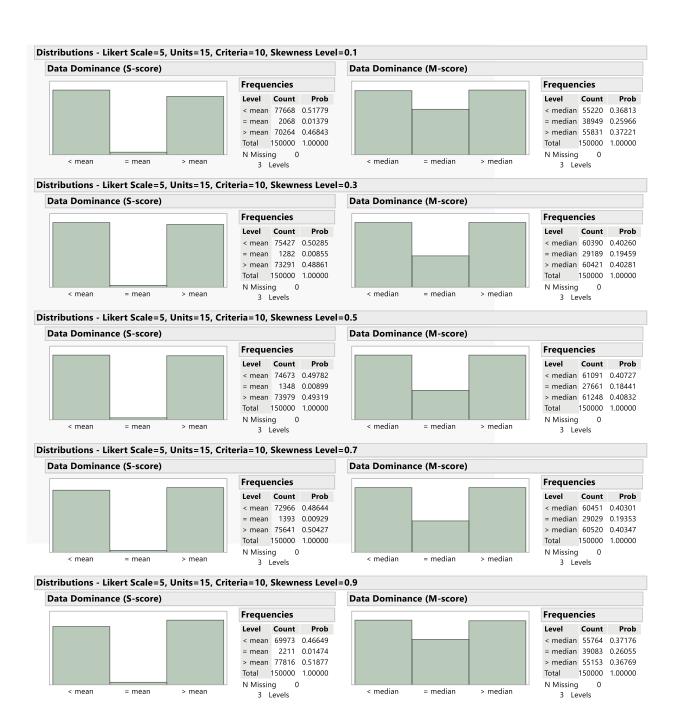




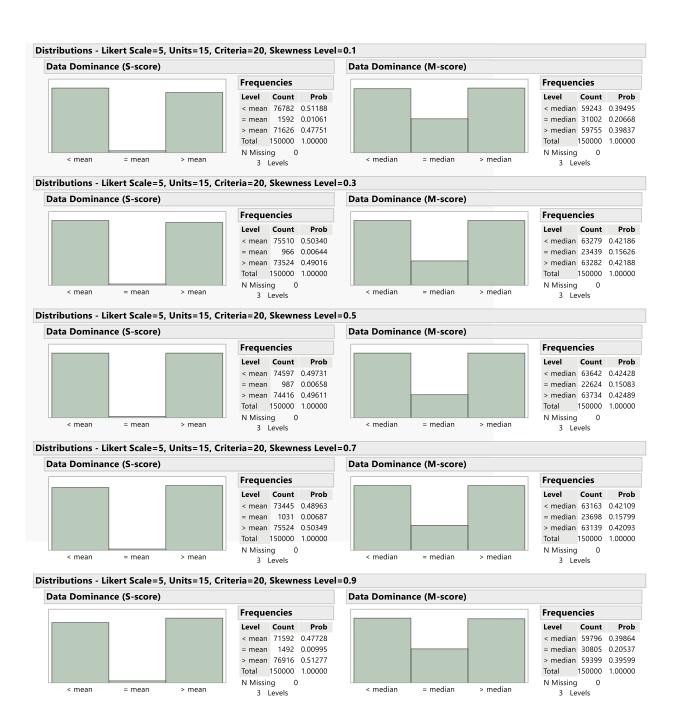


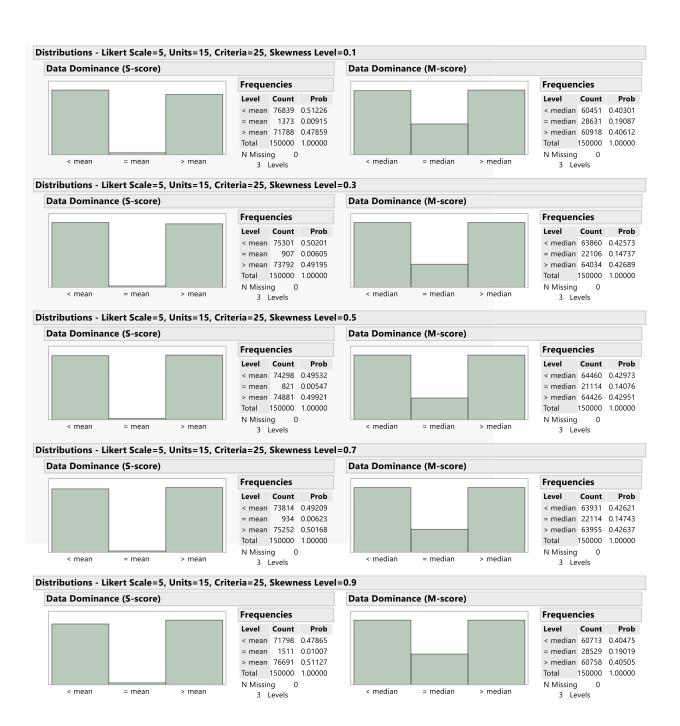


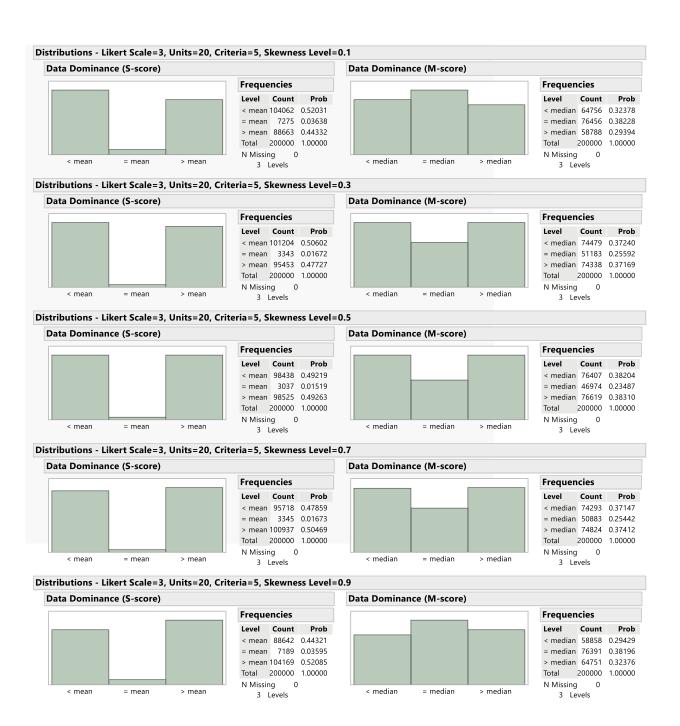


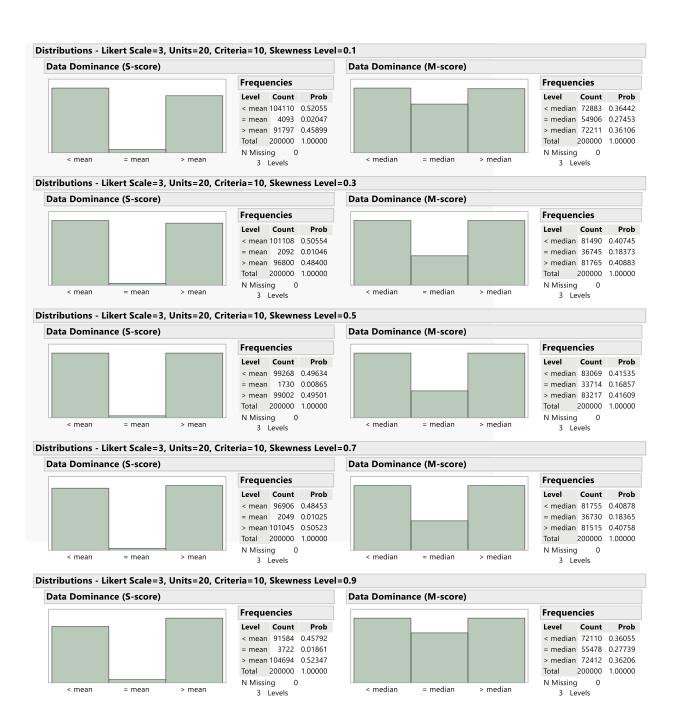








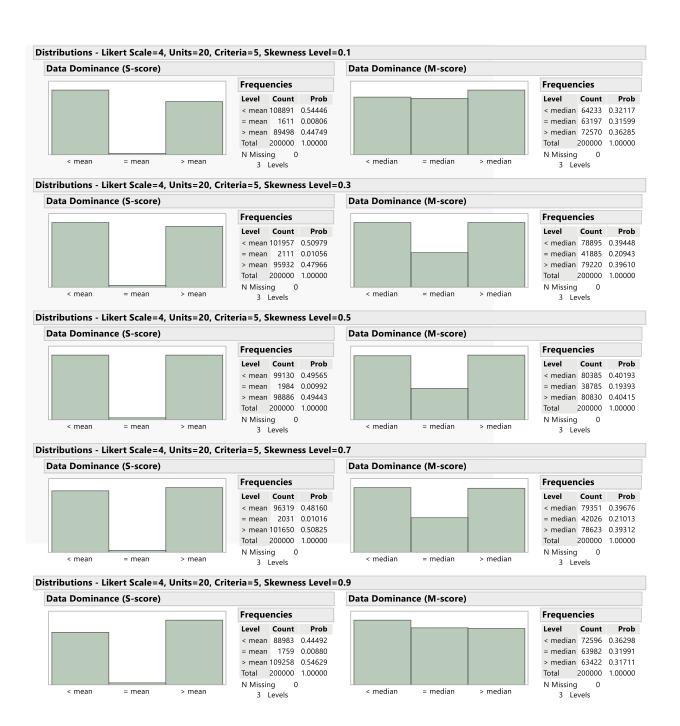










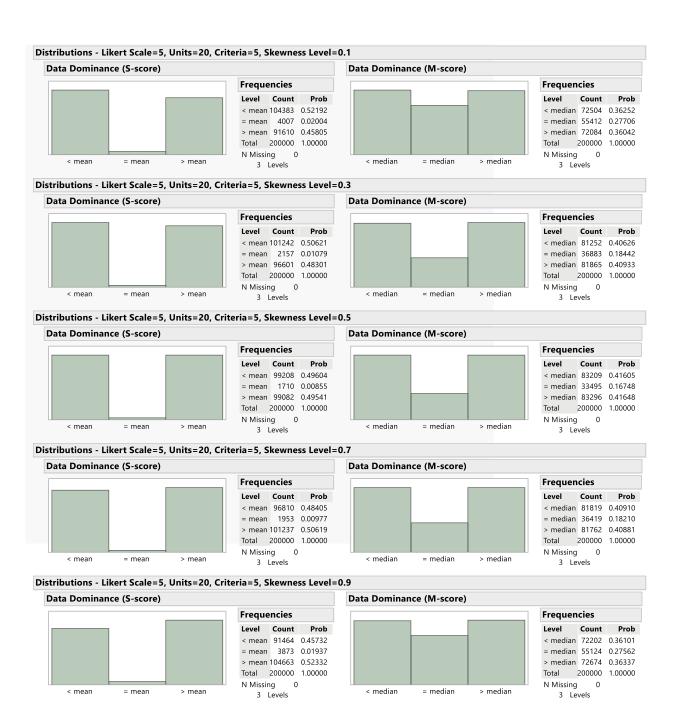


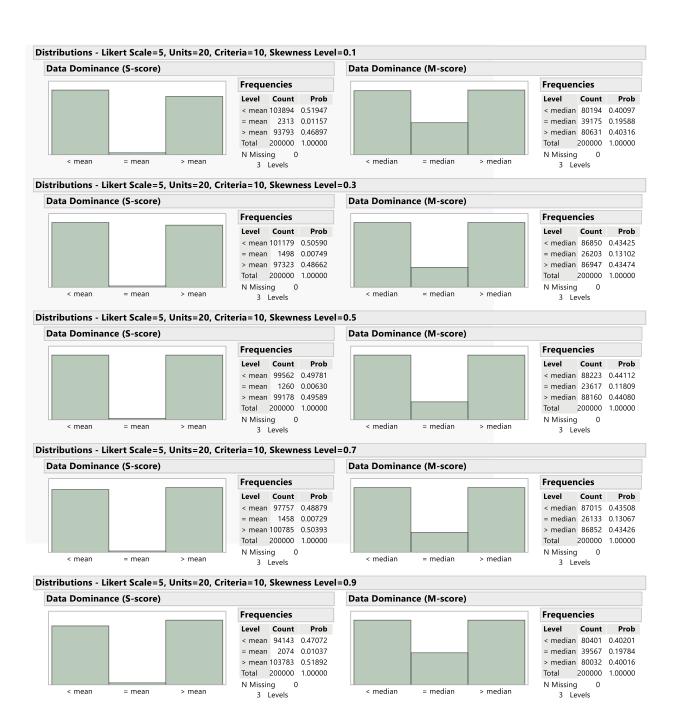


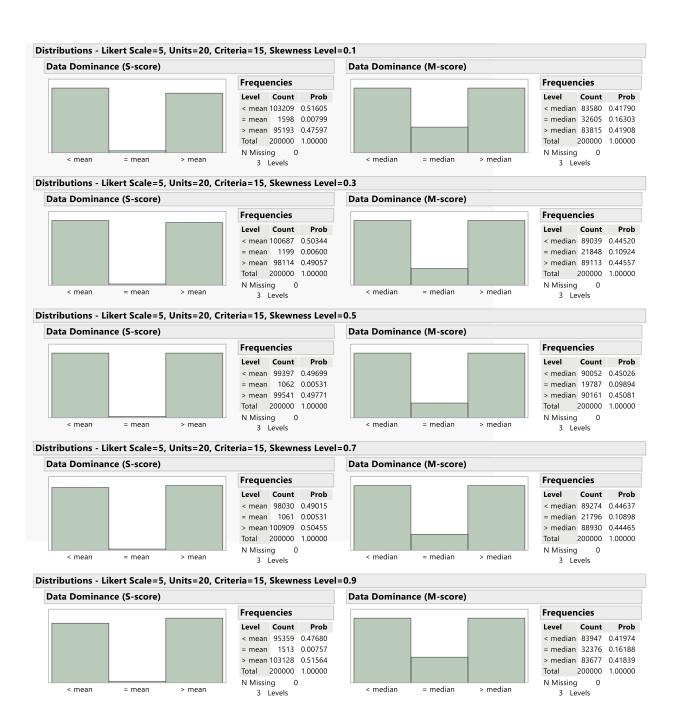




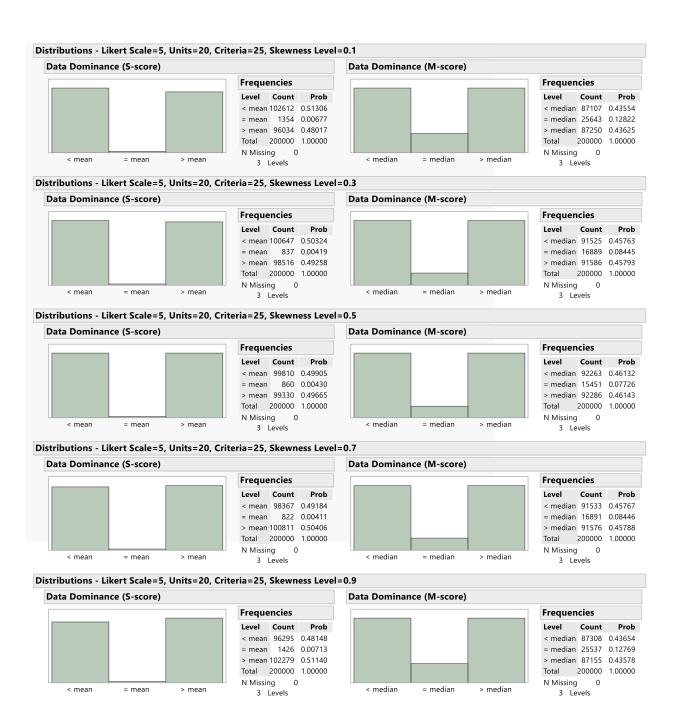












Bibliography

- Air Force Instruction (AFI) 33-360. (2015, December 1). *Publications and Forms Management*.

 Retrieved from https://static.e-publishing.af.mil/production/1/saf_aa/publication/dafi33-360/dafi33-360.pdf
- Air Force Instruction (AFI) 90-201. (2018, November 20). *The Air Force Inspection System*. Retrieved from https://static.e-publishing.af.mil/production/1/saf_ig/publication/afi90-201/afi90-201.pdf.
- Camm, F., Werber, L., Kim, J., Wilke, E., & Rudavsky, R. (2013). *Charting the Course for a New Air Force Inspection System*. rand.org. Retrieved from https://www.rand.org/content/dam/rand/pubs/technical_reports/TR1200/TR1291z1/RAND _TR1291z1.pdf.
- Casella, G., & Berger, R. L. (2002). *Statistical inference* (Second edition). Brooks/Cole Cengage Learning.
- Craft, B. A. (2016, December). Changing Air Force Culture: An Analysis of the New Air Force
 Inspection System Implementation Strategy. Retrieved from
 https://apps.dtic.mil/sti/pdfs/AD1054623.pdf
- Eschenbach, T. G. (2011). *Engineering Economy: Applying Theory to Practice* (3rd ed.). Oxford University Press.
- Hyde, R. D. A New Inspection System: Improving Warfighting Capability. *The Inspector General Brief 65*, no. 1 (January 2013)
- Johns, R. (2010, March). *Likert items and scales*. Survey Question Bank: Methods Fact Sheet.

 Retrieved May 25, 2022, from

 https://www.sheffield.ac.uk/polopoly_fs/1.597637!/file/likertfactsheet.pdf

- McClave, J. T., Benson, P. G., & Sincich, T. (2018). Statistics for Business and Economics.

 Pearson.
- Mueller, S. P. From The Inspector General. *The Inspector General Brief 64*, no. 5 (November 2012)
- Mueller, S. P. From The Inspector General. *The Inspector General Brief 66*, no. 4 (July 2014)
- Office of the Secretary of the Air Force, Inspector General, Inspections Directorate (SAF/IGI).

 August 2016. *Commander's Inspection Program Handbook Ver 4.*0.
- Office of the Secretary of the Air Force, Inspector General, Inspections Directorate (SAF/IGI). 1

 July 2013. Commander's Inspection Program Wing/FOA/DRU Implementation Guide

 Ver 1.0.
- Rogers, M. E. From TIG: Tenets of The New Inspection System. *The Inspector General Brief* 64, no. 1 (March 2012)
- Saaty, T. L. (1994). How to Make a Decision: The Analytic Hierarchy Process. *Interfaces*, 19–43.
- Stuart, T. (2019, June 4). *How to write a waiver during "The summer of waivers"*. Edwards Air Force Base News. Retrieved March 1, 2022, from https://www.edwards.af.mil/News/Article/1865702/how-to-write-a-waiver-during-the-summer-of-waivers/
- Willits, F., Theodori, G., & Luloff, A. E. (2016). Another Look at Likert Scales. *Journal of Rural Social Sciences*, 31(3). https://egrove.olemiss.edu/jrss/vol31/iss3/6/

REPORT DOCUMENTATION PAGE Form Approved OMB No. 074-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-	2. REPORT TYPE		3. DATES COVERED (From – To)	
YYYY)	Master's Thesis		September 2020 – June 2022	
16-06-2022				
TITLE AND SUBTITLE			5a. CONTRACT NUMBER	
		5b.	GRANT NUMBER	
Evaluating a Statistical-Based Assessment Tool for Stratifying Risk				
Among U.S. Air Force Organizations		5c.	5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)		5d.	PROJECT NUMBER	
Low, Tiffany A., Major, USAF		5e.	TASK NUMBER	
		5f.	WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)			8. PERFORMING ORGANIZATION	
Air Force Institute of Technology			REPORT NUMBER	
Graduate School of Engineering and Management (AFIT/EN)				
2950 Hobson Way, Building 640			AFIT-ENV-MS-22-J-066	
WPAFB OH 45433-7765				
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR'S ACRONYM(S)	
23 rd Wing Inspector General Inspections			23 WG/IGI	
5090 Gardner St, Building 115, Suite 105			11. SPONSOR/MONITOR'S REPORT	
Moody AFB, GA 31699			NUMBER(S)	
229-257-9352			. ,	
dustin.mccutcheon@us.af.mil or 23wg.igi@us.af.mil				
ATTN: TSgt Dustin McCutcheon				

12. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRUBTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT

The Air Force Inspection System is a proponent of utilizing a risk-based sampling strategy (RBSS) for conducting inspections from major command levels down to the unit level. The strategy identifies areas deemed most important or risky by commanders and prioritizes them accordingly for an independent assessment by the Inspector General. While Air Force regulation specifies the need to use a RBSS for inspection, the implementation process is delegated to individual commands and, subsequently, wings. The 23rd Wing, the sponsor for this research, directed us to analyze a RBSS tool highlighted as an example from which to adopt for those units within Air Combat Command, the major command for the 23rd. Our analysis entailed both descriptive and inferential measures. The results identified some potential shortfalls for which solutions were proposed. The recommended measures include using a median-based metric instead of a mean-based metric to score risk for organizations, a 3-point Likert scale to evaluate criteria, as well as a scoring system for dichotomous criteria when mixing with either a 3-, 4-, or 5-point Likert scale criteria.

15. SUBJECT TERMS

Risk analysis, mean centering, median centering, Likert scales

16. SECURITY CLASSIFICATION 17. OF:		υ,		19a. NAME OF RESPONSIBLE PERSON Edward White, AFIT/ENC	
a. REPORT	b. ABSTRACT	c. THIS PAGE	OF ABSTRACT	OF PAGES	19b. TELEPHONE NUMBER (Include area code) (937) 255-3636, ext. 4540 (Edward.white@afit.edu)
U	U	U	UU	140	(Edward.winte@art.cdu)

Standard Form 298 (Rev. 8-98) Prescribed by ANSI Std. Z39-18