2019 State Damage Prevention Program Grants Progress Report
CFDA Number: 20.720

Award Number: 693JK31940021PSDP
Project Title: State Damage Prevention (SDP) Program Grants - 2019
Date Submitted: *April 6, 2020*
Submitted by: *Rick Pevarski*

Specific Objective(s) of the Agreement

Under this grant agreement, the recipient will:

> Identify a statistically valid sampling of tickets to audit, conduct a statistic audit of its current auditing process, and create an algorithm that can identify high-risk tickets that can then be audited, with the algorithm being capable of adjusting weights of variables based on data gathered from audits. (Elements 1-9)

Workscope
Under the terms of this grant agreement, the Recipient will address the following applicable elements listed in the approved application, pursuant to 49 U.S.C. §60134 (a), (b).

- Element 1 (Effective Communications): Participation by operators, excavators, and other stakeholders in the development and implementation of methods for establishing and maintaining effective communications between stakeholders from receipt of an excavation notification until successful completion of the excavation, as appropriate. (Applicable)
- Element 2 (Comprehensive Stakeholder Support): A process for fostering and ensuring the support and partnership of stakeholders, including excavators, operators, locators, designers, and local government in all phases of the program. (Applicable)
- Element 3 (Operator Internal Performance Measurement): A process for reviewing the adequacy of a pipeline operator's internal performance measures regarding persons performing locating services and quality assurance programs. (Applicable)
- Element 4 (Effective Employee Training): Participation by operators, excavators, and other stakeholders in the development and implementation of effective employee training programs to ensure that operators, the one call center, the enforcing agency, and the excavators have partnered to design and implement training for the employees of operators, excavators, and locators. (Applicable)
- Element 5 (Public Education): A process for fostering and ensuring active participation by all stakeholders in public education for damage prevention activities. (Applicable)

- Element 6 (Dispute Resolution): A process for resolving disputes that defines the State authority's role as a partner and facilitator to resolve issues. (Applicable)
- Element 7 (Enforcement): Enforcement of State damage prevention laws and regulations for all aspects of the damage prevention process, including public education, and the use of civil penalties for violations assessable by the appropriate State authority. (Applicable)
- Element 8 (Technology): A process for fostering and promoting the use, by all appropriate stakeholders, of improving technologies that may enhance communications, underground pipeline locating capability, and gathering and analyzing information about the accuracy and effectiveness of locating programs. (Applicable)
- Element 9 (Damage Prevention Program Review): A process for review and analysis of the effectiveness of each program element, including a means for implementing improvements identified by such program reviews. (Applicable)

Accomplishments for this period (Item 1 under Article IX, Section 9.01 Progress Report: "A comparison of actual accomplishments to the objectives established for the period.")

The following is a list of accomplishments thus far:

1. Vendor selected
    a. Statistical Applications & Innovations Group (SAIG), Virginia Tech
2. Contracts signed
    a. Executed Professional Services Agreement with Vendor
    b. Agreement for Disclosure and Transfer of Confidential Information and Personally Identifiable Information
3. Completed the Statement and Scope of Work with SAIG
4. Created and delivered to vendor the Data Dictionary
5. Delivered to vendor three years of web ticket entry audit data fields to be used in development of AI
6. Vendor created and delivered the VA811 Safety Risk Rating Auditor Agreement
7. In development stage

Accomplishments for this period (Item 1 under Article IX, Section 9.01 Progress Report: "A comparison of actual accomplishments to the objectives established for the period.")

The following is a list of accomplishments thus far:

1. Vendor selected
    a. Statistical Applications & Innovations Group (SAIG), Virginia Tech
2. Contracts signed
    a. Executed Professional Services Agreement with Vendor
    b. Agreement for Disclosure and Transfer of Confidential Information and Personally Identifiable Information
3. Completed the Statement and Scope of Work with SAIG

4. Created and delivered to vendor the Data Dictionary
5. Delivered to vendor three years of web ticket entry audit data fields to be used in development of AI
6. Vendor created and delivered the VA811 Safety Risk Rating Auditor Agreement
7. In development stage

Quantifiable Metrics/Measures of Effectiveness (Item 2 under Article IX, Section 9.01 Progress Report: "Where the output of the project can be quantified, a computation of the cost per unit of output.")

Quantifiable Metrics/Measures of Effectiveness (Item 2 under Article IX, Section 9.01 Progress Report: "Where the output of the project can be quantified, a computation of the cost per unit of output.")

The project is still in the development stage with Phase I almost complete and Phase II having begun. The completion and submission of the Safety Risk Rating Auditor Agreement is the only tangible item at this time, and with it being tied to Phase II, the actual creation of an algorithmic model (artificial intelligence) to assist auditing, there is no quantifiable output at this stage of the project. The Safety Risk Rating Auditor Agreement report is found in *Appendix A*. A brief overview the report is found below:

1. Purpose

Phase I of the project consists of assessing, using statistical modeling, the overall effectiveness of VA811's current auditing procedures in order to determine the current level of auditing effectiveness based on three criteria: 1) auditing repeatability (degree to which same auditors examining same tickets get the same results); 2) auditing reproducibility (degree to which different auditors examining the same tickets get the same results); 3) auditing accuracy (degree to which auditors achieve the same results auditing tickets as did the experts). A fourth component relates to analyzing whether or not the three areas related to repeatability, reproducibility, and accuracy statistically improve as the current 25 audit codes are reduced to 13 merged codes, 4 ordinal codes, and a binary assessment. Phase I is also involving SAIG identifying a statistically valid sampling for random auditing purposes. Phase II will take the Data Dictionary and the ticket audit data, along with information gleaned from the Phase I deliverables, to create a learning algorithm to ensure 100% auditing of web tickets.

2. Methodology

A total of 25 Damage Prevention Specialists (DPS) involved in the auditing process examined and, using one of 25 audit codes and "verified," indicating the ticket presented no evidence of error, scored 50 tickets. A team of 4 "experts" created the scoring key, with the key identifying the accurate code for each ticket or determining the ticket was verified. Three of the 25 DPS were randomly chosen to audit the same tickets again roughly two and a half weeks after the initial auditing of the tickets. Through this process, SAIG was able to statistically analyze the results for accuracy, repeatability, and reproducibility.

3. Value

The near completion of Phase I has provided VA811 with valuable insights into its auditing process. It is hypothesized that reducing the auditing process to a binary classification will increase overall auditing accuracy. The auditing process will involve the wider pool of auditors simply classifying tickets as verified, meaning no evidence of error, or as containing a Safety Level, meaning evidence of error exists. Tickets identified as possibly containing an error will then be turned over to a smaller and dedicated QA/QC team to assign an agreed upon audit code or determine the ticket is verified. Upon completion of the identification of a statistically valid random number of tickets to audit based on ticket volume and error rate, VA811 will adjust its current auditing practices to audit the identified number of tickets. At the conclusion of Phase II, VA811 will work with its software development software company to put in place the learning algorithm so that 100% of web-originated normal tickets will be screened using the binary classification system. Any tickets the AI engine (learning algorithm) identifies as possibly containing an error will be audited by the dedicated QA/QC team. DPS auditors will continue to also audit a statistically valid random sampling of DPS originated tickets using this binary classification system, sending tickets with potential Safety Level concerns to the QA/QC team.

Issues, Problems or Challenges (Item 3 under Article IX, Section 9.01 Progress Report: "The reasons for slippage if established objectives were not met.")

There have been no problems or challenges preventing the project from reaching goals. Phase I is almost entirely complete, and Phase II has begun.

Mid-term Financial Status Report

See *Appendix B – Mid Term Financial Statement*

Plans for Next Period (Remainder of Grant)

VA811 has provided SAIG with all data needed to create the AI auditing engine. SAIG is working on creating the learning algorithm, which should be ready for installation by the end of the grant. At that point, VA811 will work with its ticket entry software developer in order to identify the best way to integrate the AI engine into the auditing process through Ticket Entry.

Appendix A

VIRGINIA TECH

SAIG

Data • Science • Decisions

Statistical Applications & Innovations Group

VA811 Safety Risk Rating Auditor Agreement :

Accuracy, Reproducibility & Repeatability

Jennifer H Van Mullekom, PhD

Eric Bae, MS

VIRGINIA TECH | STATISTICAL APPLICATIONS & INNOVATIONS GROUP | MARCH 25, 2020

VA811 Safety Risk Rating Auditor Agreement: Accuracy, Reproducibility & Repeatability

Executive Summary

This report assesses the initial screening measurement system of web excavation ticket entry at VA811 for safety level risk. Accuracy, repeatability, and reproducibility of the auditors on an 25 audit code measurement scale is characterized via attribute agreement analysis, a form of measurement systems analysis for discrete data. Three additional scales which collapse the 25 code scale into smaller numbers of categories were also assessed. These analyses were performed for two purposes: 1) providing information on operations improvement for VA811 and 2) determining which scale should be used as the dependent variable for a predictive model to inform safety level risk audits.

Statistical analyses indicate the measurement system with 25 three digit safety risk codes has poor accuracy (auditors have poor agreement with experts), poor repeatability (auditors have poor agreement with themselves), and poor reproducibility (auditors have poor agreement with each another). When the scales are collapsed reducing the number of codes, the quality of the measurement system improves. However, additional steps should be taken to ensure a fully validated measurement system. This report as well as a wealth of human factors rating research supports reducing the number of codes in the initial screening audit for safety level risk. In addition, it supports the use of a two level safety risk scale (safety violation/no safety violation) for predictive modeling.

Introduction

As more customers employ VA811's web ticket systems to request utility identification prior to excavation, detection of potential errors in tickets with high accuracy is essential. The web ticketing system contains safety level risks in a higher proportion than direct calls to VA811 representatives. VA811 has contracted with VT SAIG to ultimately provide a predictive model that will aid in the identification of high risk tickets. The predictive model will be implemented as part of a new risk-based audit plan. Prior to modeling, it is important to understand the accuracy and precision of the safety level violation system that will be used in the modeling. This report details the measurement system analyses (MSA) for the audit safety level violation system on various scales derived from the original 25 category scale.

Current methods of detecting errors involve multiple auditors assigning safety level codes. Auditors assigned the status as "accurate" if no issue is found. After this initial assessment, the ticket is forwarded to an expert for assignment of a final audit code and resolution of the safety risk issue. For this MSA, 15 different auditors and a consensus panel of experts audited 50 tickets. All 15 auditors assigned one of the 25 safety level codes specified in current procedure to each of the 50 tickets on the 10th of December 2019, whereas three of the 15 auditors performed yet another audit on the 31st of December 2019.

This report will answer four questions regarding the measurement system:

1. Is the measurement system repeatable? Do the ***same*** auditors reviewing the ***same*** tickets, get the same results on multiple trials?
2. Is this measurement system reproducible? Do ***different*** auditors reviewing the ***same*** tickets get the same results?
3. Is this measurement system accurate? Do ***auditors*** get the same results as reached in the consensus session by ***experts***?
4. Do the quantities which evaluate repeatability, reproducibility, and accuracy improve as we reduce the 25 safety level codes to 13 merged codes, to 4 ordinal codes, and finally to a binary assessment (accurate versus violation)?

Details of the methodology are seen in the Methods section. Statistical tools used are explained in detail in the Analysis section. Output of the analysis is summarized in the Results section. The Appendix contains detailed output from each analysis with an explanation of interpretation in each section of the Appendix.

Methods

The MSA will be performed using attribute agreement analysis. In this attribute agreement analysis, we start by asking ourselves the following questions:

1. Is the outcome consistent over different trials for each auditor? (Repeatability)
2. Is the outcome consistent across all auditors? (Reproducibility)
3. Is the outcome consistent for both auditors and experts? (Accuracy/Bias)
4. Is there a functional difference between leaving all categories as possible choices versus merging similar categories together?

The first question represents repeatability. Poor repeatability indicates inconsistency in individual auditors. The second question represents reproducibility. Poor reproducibility suggests that there is high variability among safety level codes assigned to the same ticket by multiple auditors. Accuracy is assessed by comparing auditors rating to the panel of experts' consensus rating. The first three questions are part of what is considered "Attribute Agreement Analysis," and involves statistical tools such as Fleiss' Kappa and Kendall's W and Tau. These tools will be further explained in the *Analysis* section. Finally, the fourth and final question asks if reducing total category options leads to improved measurement systems.

To answer the fourth question, four different scenarios of category options will be analyzed using Attribute Agreement Analysis menu option in Minitab™. The results of this analysis will be compared to each other and to benchmarks common to MSA. The scenarios of category options considered – all codes, reduced codes, risk level, and violation status—are shown in Table 1.

Table 1. Original Scale and Modified Scales

Original Scale	Modified Scales		
Full Codes	Similar Merged	Risk Level	Violation Status
Accurate	*Accurate*	*1*	*No*
110*	110	1	No
140	140	1	No
150*	150	1	No
151		1	No
152		1	No
161*	161	1	No
162*		1	No
170	*170*	*1*	*No*
171		1	No
172*		1	No
174		1	No
181*	**181**	**2**	**Yes**
190	***190***	***2***	***Yes***
191		**2**	**Yes**
240	*240*	*4*	*Yes*
241		4	Yes
260	260	4	Yes
261*	261	4	Yes
262*		4	Yes
270	*270*	*4*	*Yes*
271		**3**	**Yes**
273		**3**	**Yes**
290	*290*	4	Yes
291		*4*	*Yes*

* Codes that were not used by any of the auditors were starred

* Codes that the experts used are underlined and in italics

Table 1 illustrates how the categories were merged from one scenario to the next. There were a total of 25 different codes including "Accurate;" however, only 18 were used by any of the 50 auditors.

Repeatability and reproducibility analyses were performed directly on the data set with the original codes. After that, the analyses were performed on merged codes (second column of *Table 1*), risk levels (third column), and violation status (fourth column). Note that on the risk level scenario, the option "accurate" has been merged into codes colored green, representing "minimal risk" category. On the violation status, all options formally belonging to 110 to 174 were collapse into the "No Violation Status" or, alternatively, "Accurate" as the potential risk associated with tickets assigned to those codes are very low. The remaining codes are considered a violation.

There are two types of agreements: absolute and relative agreements. Absolute agreement requires an exact match and is most commonly used in measurement systems analysis with nominal or named categories that have no ordering. Absolute agreement percentage is the total number of tickets with agreement divided by the total number of tickets. Absolute agreement is calculated for all four scenarios. Relative agreement, however, does put emphasis on the scale, and is used for cases with ordinal variables. As the name implies, ordinal variables have an order to them but there is not a defined numerical interpretation for the distance between categories. Low, medium, high or even 1-5 on a survey rating scale are examples of ordinal scales. Relative agreement is calculated for the risk level scenario scaled 1-4 that has an order of severity in column 3 of Table 1.

To further characterize absolute and relative agreement, consider two different auditors rating the same ticket with 151 and 152. This will still count as the same "disagreement" as rating it with 151 and 290, despite the fact that the latter appears to be a more serious disagreement. Fleiss's Kappa and the agreement percentage both measure absolute agreement. For the risk level scenario, additional statistics called Kendall's W and Kendall's Tau were calculated. Output of these statistics can be seen on *Table 2* on the Results section. The Kendall statistics measure relative agreement for ordinal data.

For the repeatability section, in addition to those on the reproducibility, agreement percentage and the kappa values were calculated for each auditor's two assignments and each auditor against the expert. *Table 3* shows the output of this.

Analysis

In this analysis, Fleiss' Kappa, Kendall's W, and Kendall's Tau were utilized to develop attribute agreement measurement systems. Fleiss' Kappa measures absolute agreement and the Kendall's W and Tau measure relative agreement for ordinal data. Minitab™ was used to calculate these statistics, and the Minitab™ output can be seen in the Appendix section.

Fleiss' Kappa

Fleiss' Kappa measures the degree of agreement over and above the amount of agreement by chance. The Kappa can take the value between -1 and 1, where 1 represents complete agreement, and -1 represents complete disagreement, and 0 represents agreement level that is equal to the level that would have been obtained completely by chance. As a rule of thumb for measurement system analysis, the Kappa value of above 0.9 qualifies as acceptable.

Kendall's W

Both Kendall's W and Kendall's Tau are applied when the outcome is ordinal. Of the four scenarios – all codes, reduced codes, risk levels, and violation status – only the risk levels involve ordinal measurements of between 1 to 4. Therefore, both W and Tau are applied only when analyzing risk levels.

Kendall's W measures the degree of association of ordinal assessments made by multiple auditors when assessing the same samples. The W can take any value between 0 and 1, where 0 represents no concordance and 1 represents perfect concordance.

Kendall's Tau

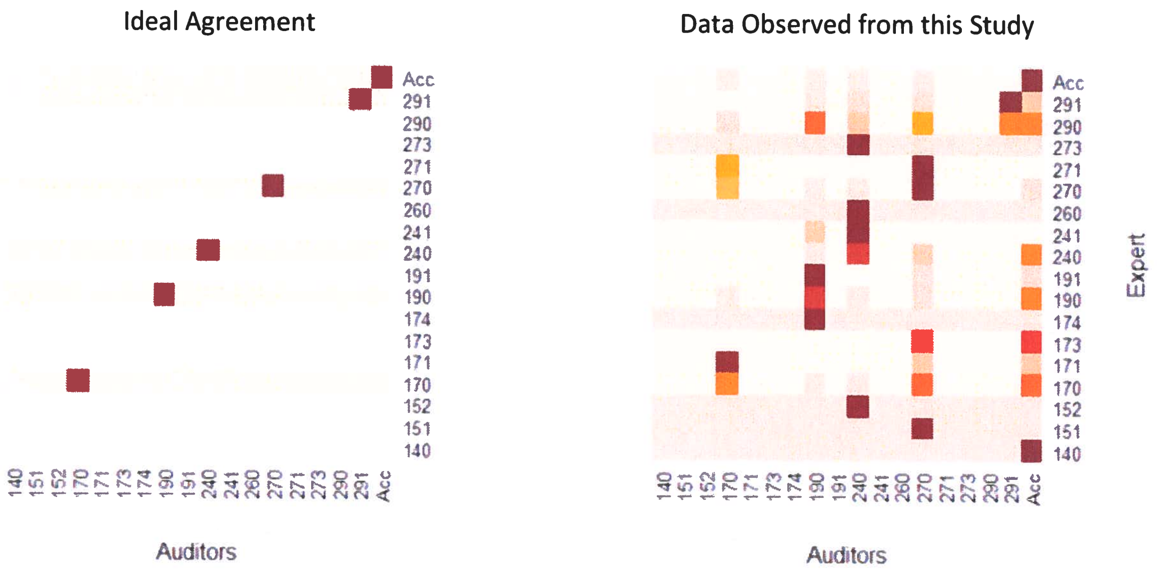
Kendall's Tau, also known as Kendall's correlation coefficient, is a correlation coefficient specifically for ordinal variables and, therefore, follow values between -1 and 1, where -1 represents complete opposite and 1 represents complete match.

For both Kendall's W and Tau, the same rule of thumb of above 0.9 as an acceptable outcome applies.

Results

This section contains a brief summary of the results obtained from the assessment agreement analysis as laid out on the Introduction section. Repeatability and reproducibility outputs will be mentioned separately.

A heat map will be used to compare an ideal system with the observed data from this study. *Figure 1* illustrates the heat map of the assessment of the auditors against the expert. In an ideal situation, where all auditors agreed with the expert 100% of the time, the left heat map would be produced. The ideal map shows all points occupying cells in the along the diagonal from the bottom left to the top right. The right side of the figure is the heat map that was obtained from the study. While some patterns of diagonals are identifiable, it is clear that there is a distinct visible difference from the "ideal" heat map.



Ideal Agreement

Data Observed from this Study

Figure 1

Left: An ideal scenario where all auditors' predictions matches with the expert's
Right: Current scenario

Table 2 is a summary of the reproducibility and accuracy. The measures of agreement increase moving from the left column to the right column of the table. The more the scale (codes) are collapsed, the higher the average percentage agreement and the higher the average Kappa value for both reproducibility and accuracy. Likewise, the percentage of unanimous agreement for both among the 15 auditors and with the expert increased as more codes were merged.

However, both the percent of agreement and the overall Kappa appears to remain far below the acceptable values of 90% and 0.90 respectively, as do the Kendall statistics. Even the last scenario considered, violation status as a binary outcome, representing the least complicated scale does not achieve this benchmark. This prompts consideration for a measurement system improvement project.

Table 2 Reproducibility and Accuracy summary table of the relevant output of the four scenarios

		Full Codes	Similar Merged	Risk Level	Violation Status
Reproducibility Between-auditors	% Agreement Range (% Average)	36 – 64 (52.80)	42 – 70 (58.40)	58 – 76 (67.87)	62 – 82 (73.60)
	Kappa Range (Overall Kappa)	<0 – 0.47 (0.29)	<0 – 0.47 (0.36)	0.17 – 0.43 (0.40)	0.43 (0.43)
	Unanimous % Agreement	6	8	24	28
	Kendall's Tau	-	-	0.52	-
Accuracy Auditor vs. Expert	Kappa Range vs. Expert (Overall Kappa)	0.20 – 0.58 (0.40)	0.22 – 0.57 (0.45)	0.47 – 0.48 (0.46)	0.47 (0.47)
	Unanimous % Agreement vs. Expert	6	8	24	28
	Kendall's W (Overall)	-	-	0.24 – 0.61 (0.47)	-

Table 3 is the summary table for repeatability measurements. There were three auditors – auditors 9, 10, and 11 – who evaluated the sample of 50 tickets twice – once at the 10th of December, 2019 and another at the 31st of December, 2019. The same metrics described in Table 2 are reported in Table 3. Once again, with fewer categories, we observe higher agreement range among the auditors themselves, each auditor against the expert, and all three auditors and the expert.

Note that the percentage of agreement differs when it is calculated among auditors and when the auditors were compared to the experts. This is because all auditors agreed unanimously on certain tickets but did not agree with the experts.

Table 3. Repeatability summary table of the relevant output of the four scenarios

		Full Codes	Similar Merged	Risk Level	Violation Status
Repeatability Auditor vs. Self	% Agreement Range (% Average)	54 – 72	60 – 86	68 – 82	78 – 90
	Kappa Range (Overall Kappa)	<0 – 1* (0.34 – 0.64)	<0 – 1* (0.44 – 0.81)	<0 – 0.82 (0.33 – 0.78)	0.48 – 0.80
	Kendall's Tau	-	-	0.76-0.89	-

In every scenario, for both reproducibility and repeatability, the goal of over 90 % average prediction accuracy rate and 0.90 Fleiss' Kappa, Kendall's W, and Kendall's Tau was not met.

DISCUSSION

Accuracy

The ability of the auditors to align their classifications to the expert consensus classification falls below the desired thresholds of 90% for absolute agreement and 0.9 for Kappa or Kendall's statistics.

Reproducibility

The reproducibility analysis shows that as more codes were merged, agreement increased among the auditors. Fleiss' Kappa statistics also improved. However, in all of the four scenarios, the improvement in accuracy fell short of our standard of 90% for absolute agreement and 0.90 in Kappa and Kendall's statistics. This is an indication that reproducibility should be improved significantly.

Repeatability

The repeatability analysis also shows that as more codes were merged, agreement between each auditor's assessments on multiple trials improved. However, similarly to the reproducibility result, improvements as the scale was collapsed usually fell well short of the aforementioned standard. There were a few exceptions with one auditor into the ordinal risk level 1-4 measurement and the binary violation status measurement.

Conclusions and Recommendations

The analyses in this report were performed for two purposes: 1) providing information on operations improvement for VA811 and 2) determining which scale should be used as the dependent variable for a predictive model to inform safety level risk audits.

With respect to operations improvement, this study shows that the number and complexity of audit code descriptions impair accuracy, repeatability, and reproducibility. While collapsing categories suggests improvement in the system, this was done via computer during data analysis. The analysis indicates a trend but does not necessarily characterize the full potential of such a shift to a scale with fewer categories. It is hypothesized that a measurement system with as possible would provide a marked improvement over the derived 4 level risk scale and 2 category binary scale analyzed in this study. Such a transition must be accompanied by thorough operational definitions and appropriate training with a follow up MSA. This transition is supported by human factors research summarized in the following quote: "In general, inspection performance is degraded as the number and types of defects increases, primarily as a result of limitations of human memory. " (Dalton & Drury, 2004). For a thorough consideration of all factors in the design of a human visual inspection system, see https://prod-ng.sandia.gov/techlib-noauth/access-control.cgi/2012/128590.pdf

An initial ticket audit is used to send the excavation ticket through final audit by supervisors and experts. The final audit codes recorded for tickets reflect expert opinion. All 25 detailed codes are eligible to be recorded for a final classification by the expert auditor. This final classification will be modeled in the second phase of the project. Based on this study, we propose converting the 25 code scale to a binary classification of (no violation, violation) for modeling phase of the project. This decision is based on both the quality of the measurement system and types of models planned for subsequent phases of the project.

References

Sandia Literature Review of Visual Inspection: https://prod-ng.sandia.gov/techlib-noauth/access-control.cgi/2012/128590.pdf

Dalton, J., & Drury, C.G. (2004). Inspectors' performance and understanding in sheet steel inspection. Occupational Ergonomics, 4, 51-65.

Appendix

This section contains an overview of a general interpretation of each item of the analysis on *Tables 2* and *3* (first and second columns). Most output has been generated using Minitab™. For each part, an example output is provided. Unless otherwise stated, all examples provided are extracted from output produced using the full 3-digit codes. The Minitab™ File with raw data will be provided as part of the documentation package.

Reproducibility

For the reproducibility, there were five key parts – assessment agreement percentage, Kappa statistic within auditors, unanimous agreement amongst auditors, Kappa statistic between auditors and the experts, and the unanimous agreement between the auditors and the experts.

Between-auditors, Assessment Agreement

The assessment agreement is obtained by comparing each auditors' assessment to that of the experts' and tallying up the percentage of assessments that matched. Figure 2 illustrates an example of the assessment agreement of each of the 15 auditors, along with a 95 % confidence interval. This figure is generated from when all three-digit codes were used.

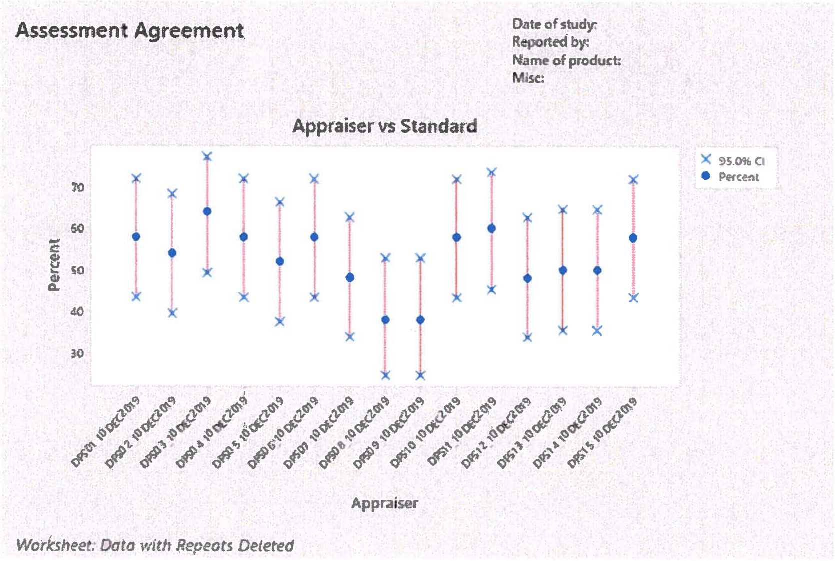
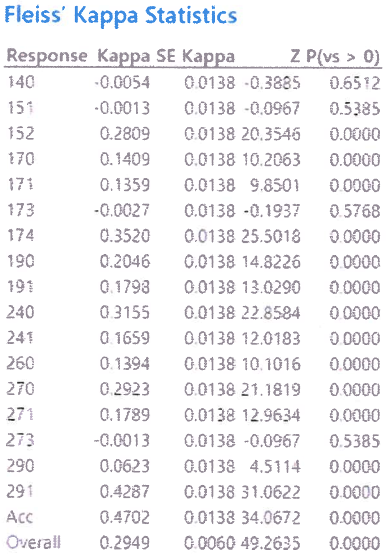


Figure 2 Example of an assessment agreement plot

Between-auditors, Kappa Range

Fleiss' Kappa statistic has been described in the Analysis section. This Kappa statistic calculates the auditors' degree of agreement on each of the responses. A sample output is in Figure 3, which illustrates the Kappa statistics of all 3-digit codes that were used by any of the auditors.



Fleiss' Kappa Statistics

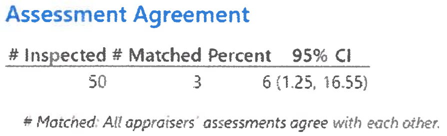
Response	Kappa	SE Kappa	Z	P(vs > 0)
140	-0.0054	0.0138	-0.3885	0.6512
151	-0.0013	0.0138	-0.0967	0.5385
152	0.2809	0.0138	20.3546	0.0000
170	0.1409	0.0138	10.2063	0.0000
171	0.1359	0.0138	9.8501	0.0000
173	-0.0027	0.0138	-0.1937	0.5768
174	0.3520	0.0138	25.5018	0.0000
190	0.2046	0.0138	14.8226	0.0000
191	0.1798	0.0138	13.0290	0.0000
240	0.3155	0.0138	22.8584	0.0000
241	0.1659	0.0138	12.0183	0.0000
260	0.1394	0.0138	10.1016	0.0000
270	0.2923	0.0138	21.1819	0.0000
271	0.1789	0.0138	12.9634	0.0000
273	-0.0013	0.0138	-0.0967	0.5385
290	0.0623	0.0138	4.5114	0.0000
291	0.4287	0.0138	31.0622	0.0000
Acc	0.4702	0.0138	34.0672	0.0000
Overall	0.2949	0.0060	49.2635	0.0000

Figure 3 Fleiss' Kappa statistic by each response

This example shows that none of the responses reached the Kappa value of 0.5, let alone 0.9, which indicate poor level of agreement among auditors. There are, likewise, a few with Fleiss' Kappa smaller than zero, though none of them appeared to have P-value (far right column) small enough (under 0.05) to be considered significantly different from zero.

Between-auditors, unanimous agreement

This value represents the percentage of tickets on which all 15 auditors agreed to a particular code value. The figure below indicates that, when none of the 3-digit codes were merged into common categories, of the 50 tickets analyzed, all auditors agreed on 3, which corresponds to 6 percent of all tickets.



Assessment Agreement

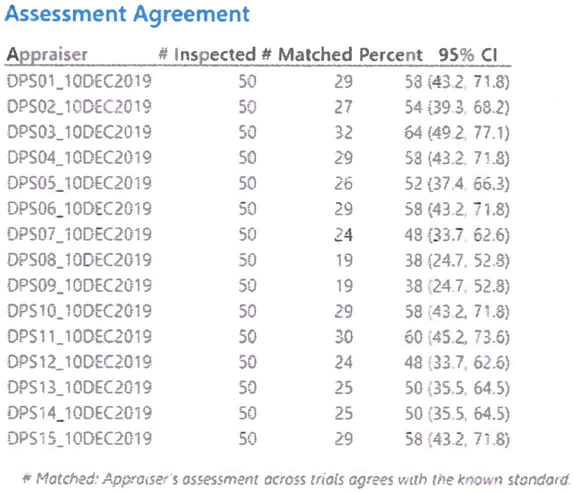
# Inspected	# Matched	Percent	95% CI
50	3	6	(1.25, 16.55)

Matched: All appraisers' assessments agree with each other.

Figure 4 Total number and percentage of tickets with unanimous agreement amongst all auditors

Each auditor vs. Experts, assessment agreement

The assessment agreement is obtained by comparing each auditors' assessment to that of the experts' and tallying up the percentage of assessments that matched. Figure 2 illustrates an example of the assessment agreement of each of the 15 auditors, along with a 95 % confidence interval. This figure is generated from when all three-digit codes were used.



Assessment Agreement

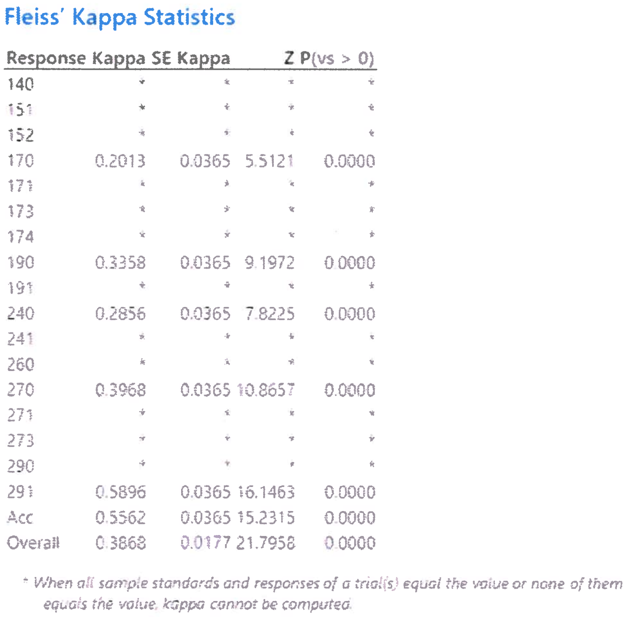
Appraiser	# Inspected	# Matched	Percent	95% CI
DPS01_10DEC2019	50	29	58	(43.2, 71.8)
DPS02_10DEC2019	50	27	54	(39.3, 68.2)
DPS03_10DEC2019	50	32	64	(49.2, 77.1)
DPS04_10DEC2019	50	29	58	(43.2, 71.8)
DPS05_10DEC2019	50	26	52	(37.4, 66.3)
DPS06_10DEC2019	50	29	58	(43.2, 71.8)
DPS07_10DEC2019	50	24	48	(33.7, 62.6)
DPS08_10DEC2019	50	19	38	(24.7, 52.8)
DPS09_10DEC2019	50	19	38	(24.7, 52.8)
DPS10_10DEC2019	50	29	58	(43.2, 71.8)
DPS11_10DEC2019	50	30	60	(45.2, 73.6)
DPS12_10DEC2019	50	24	48	(33.7, 62.6)
DPS13_10DEC2019	50	25	50	(35.5, 64.5)
DPS14_10DEC2019	50	25	50	(35.5, 64.5)
DPS15_10DEC2019	50	29	58	(43.2, 71.8)

Matched: Appraiser's assessment across trials agrees with the known standard.

Figure 5 Total percentage of tickets for which each auditor's decision matched with that of the experts'

All auditors vs. Experts, Kappa statistic

This represents the Kappa statistics for the codes that were used by the experts. As the figure below illustrates, there were only six codes used by the experts; hence, most responses do not have Kappa statistics assigned. Also noticeable is that the Kappa values are higher for "Acc" and "291" and lower for codes in between. This is an indication that the auditors were, in general, more likely to assign codes to tickets with either no detectable risk (Acc) or with very high risk (291) but were more likely to disagree on tickets with medium-level risks.



Fleiss' Kappa Statistics

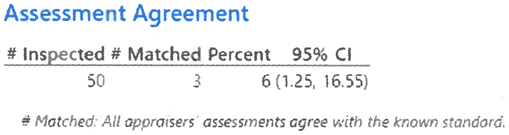
Response	Kappa	SE Kappa	Z	P(vs > 0)
140	*	*	*	*
151	*	*	*	*
152	*	*	*	*
170	0.2013	0.0365	5.5121	0.0000
171	*	*	*	*
173	*	*	*	*
174	*	*	*	*
190	0.3358	0.0365	9.1972	0.0000
191	*	*	*	*
240	0.2856	0.0365	7.8225	0.0000
241	*	*	*	*
260	*	*	*	*
270	0.3968	0.0365	10.8657	0.0000
271	*	*	*	*
273	*	*	*	*
290	*	*	*	*
291	0.5896	0.0365	16.1463	0.0000
Acc	0.5562	0.0365	15.2315	0.0000
Overall	0.3868	0.0177	21.7958	0.0000

* When all sample standards and responses of a trial(s) equal the value or none of them equals the value, kappa cannot be computed.

Figure 6 Fleiss' Kappa statistics of experts' choices against those of the 15 auditors

All auditors vs. Experts, assessment agreement

This item represents the total number of tickets on which all auditors AND the experts agreed. There were a total of three tickets (6 % of all tickets) on which all auditors and the experts agreed. This number matches with that obtained between-auditors, indicating that there were no tickets on which the auditors unanimously agreed but the experts did not.



Assessment Agreement

# Inspected	# Matched	Percent	95% CI
50	3	6	(1.25, 16.55)

Matched: All appraisers' assessments agree with the known standard.

Figure 7 Assessment agreement of all auditors and the experts

Repeatability

There were four types of assessment agreement percentage and four types of Fleiss' Kappa statistics – auditors vs. self, among auditors, each auditor vs. experts, and all auditors vs. experts. In addition, for the risk level, two Kendall's statistics were calculated as well. There were three auditors with separate assessments on the same 50 tickets on two different dates.

Auditors vs. self & each auditor vs experts, assessment agreement

The figure below shows two types of assessment agreements – within appraisers and each appraiser against the experts. The former is represented by the first plot, while the latter is represented by the second plot.

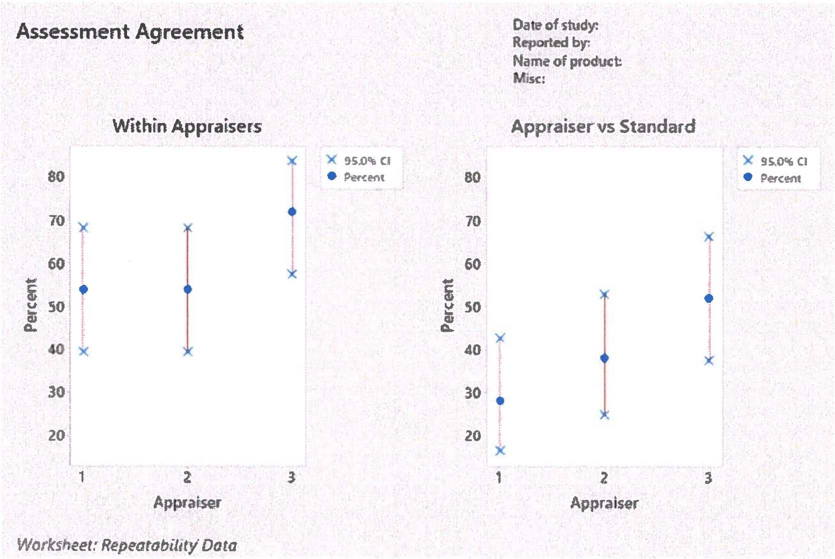
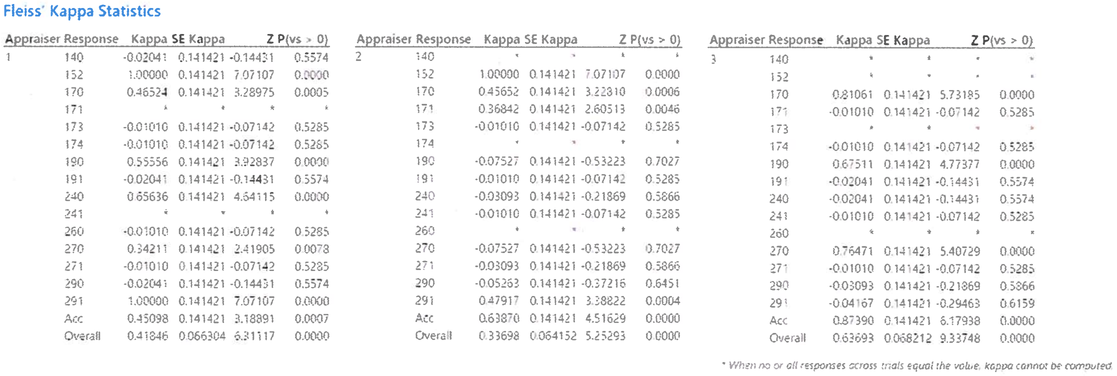


Figure 8 Assessment agreement within appraisers and each appraiser against the experts

Auditors vs. self, Fleiss' Kappa

This large table shows the Fleiss' Kappa statistics of the codes used by each of the three auditors against oneself. Ideally all responses should have Kappa statistics of 1, meaning all auditors selected the same codes for both times. In this example, there were two 1's – 152 for the auditor 1 (9) and 152 for the auditor 2 (10). However, this is mostly due to small sample size. Most others did not come near the benchmark of 0.9; this is an indication that the auditors did not have good assessment agreement level with themselves when assessed at two different times.



Fleiss' Kappa Statistics

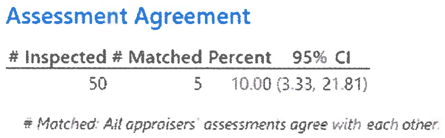
Appraiser	Response	Kappa	SE Kappa	Z	P(vs > 0)
1	140	-0.02041	0.141421	-0.14431	0.5574
	152	1.00000	0.141421	7.07107	0.0000
	170	0.46524	0.141421	3.28975	0.0005
	171	*	*	*	*
	173	-0.01010	0.141421	-0.07142	0.5285
	174	-0.01010	0.141421	-0.07142	0.5285
	190	0.55556	0.141421	3.92837	0.0000
	191	-0.02041	0.141421	-0.14431	0.5574
	240	0.65636	0.141421	4.64115	0.0000
	241	*	*	*	*
	260	-0.01010	0.141421	-0.07142	0.5285
	270	0.34211	0.141421	2.41905	0.0078
	271	-0.01010	0.141421	-0.07142	0.5285
	290	-0.02041	0.141421	-0.14431	0.5574
	291	1.00000	0.141421	7.07107	0.0000
	Acc	0.45098	0.141421	3.18891	0.0007
	Overall	0.41846	0.066304	6.31117	0.0000
2	140	*	*	*	*
	152	1.00000	0.141421	7.07107	0.0000
	170	0.45652	0.141421	3.22810	0.0006
	171	0.36842	0.141421	2.60513	0.0046
	173	-0.01010	0.141421	-0.07142	0.5285
	174	*	*	*	*
	190	-0.07527	0.141421	-0.53223	0.7027
	191	-0.01010	0.141421	-0.07142	0.5285
	240	-0.03093	0.141421	-0.21869	0.5866
	241	-0.01010	0.141421	-0.07142	0.5285
	260	*	*	*	*
	270	-0.07527	0.141421	-0.53223	0.7027
	271	-0.03093	0.141421	-0.21869	0.5866
	290	-0.05263	0.141421	-0.37216	0.6451
	291	0.47917	0.141421	3.38822	0.0004
	Acc	0.63870	0.141421	4.51629	0.0000
	Overall	0.33698	0.064152	5.25293	0.0000
3	140	*	*	*	*
	152	*	*	*	*
	170	0.81061	0.141421	5.73185	0.0000
	171	-0.01010	0.141421	-0.07142	0.5285
	173	*	*	*	*
	174	-0.01010	0.141421	-0.07142	0.5285
	190	0.67511	0.141421	4.77377	0.0000
	191	-0.02041	0.141421	-0.14431	0.5574
	240	-0.02041	0.141421	-0.14431	0.5574
	241	-0.01010	0.141421	-0.07142	0.5285
	260	*	*	*	*
	270	0.76471	0.141421	5.40729	0.0000
	271	-0.01010	0.141421	-0.07142	0.5285
	290	-0.03093	0.141421	-0.21869	0.5866
	291	-0.04167	0.141421	-0.29463	0.6159
	Acc	0.87390	0.141421	6.17938	0.0000
	Overall	0.63693	0.068212	9.33748	0.0000

* When no or all responses across trials equal the value, kappa cannot be computed.

Figure 9 Fleiss' Kappa statistics of the codes by each auditor

Among auditors, assessment agreement

The figure below is a cross-auditor comparison of assessment agreement. Out of 50 tickets inspected, only five (10 % of all tickets) were unanimously agreed upon by the three auditors in BOTH dates.



Assessment Agreement

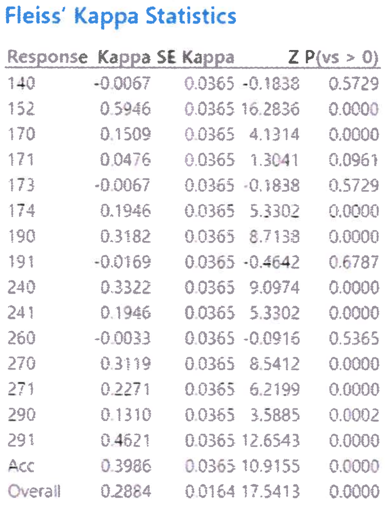
# Inspected	# Matched	Percent	95% CI
50	5	10.00	(3.33, 21.81)

# Matched: All appraisers' assessments agree with each other.

Figure 10 Assessment agreement among auditors

Among auditors, Fleiss' Kappa

This is the Fleiss' Kappa statistics of all three auditors for both dates. Therefore, this Kappa statistic would be comparing a total of six different input for each ticket.



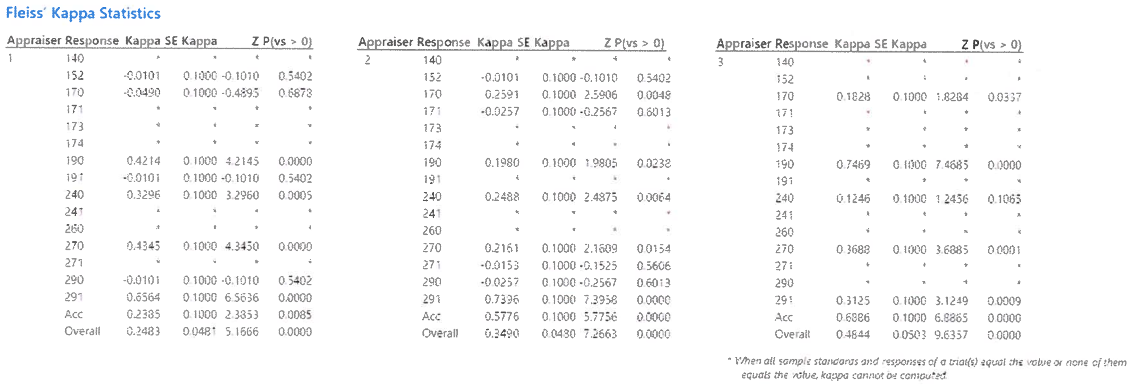
Fleiss' Kappa Statistics

Response	Kappa	SE Kappa	Z	P(vs > 0)
140	-0.0067	0.0365	-0.1838	0.5729
152	0.5946	0.0365	16.2836	0.0000
170	0.1509	0.0365	4.1314	0.0000
171	0.0476	0.0365	1.3041	0.0961
173	-0.0067	0.0365	-0.1838	0.5729
174	0.1946	0.0365	5.3302	0.0000
190	0.3182	0.0365	8.7138	0.0000
191	-0.0169	0.0365	-0.4642	0.6787
240	0.3322	0.0365	9.0974	0.0000
241	0.1946	0.0365	5.3302	0.0000
260	-0.0033	0.0365	-0.0916	0.5365
270	0.3119	0.0365	8.5412	0.0000
271	0.2271	0.0365	6.2199	0.0000
290	0.1310	0.0365	3.5885	0.0002
291	0.4621	0.0365	12.6543	0.0000
Acc	0.3986	0.0365	10.9155	0.0000
Overall	0.2884	0.0164	17.5413	0.0000

Figure 11 Fleiss' Kappa statistics of the codes among the three auditors

Each auditor vs. experts, Fleiss' Kappa

This compares the two output from each auditor to that of the experts. There are several empty values because none of the auditors assigned those codes to any of the tickets.



Fleiss' Kappa Statistics

Appraiser	Response	Kappa	SE Kappa	Z	P(vs > 0)
1	140	*	*	*	*
	152	-0.0101	0.1000	-0.1010	0.5402
	170	-0.0490	0.1000	-0.4895	0.6878
	171	*	*	*	*
	173	*	*	*	*
	174	*	*	*	*
	190	0.4214	0.1000	4.2145	0.0000
	191	-0.0101	0.1000	-0.1010	0.5402
	240	0.3296	0.1000	3.2960	0.0005
	241	*	*	*	*
	260	*	*	*	*
	270	0.4345	0.1000	4.3450	0.0000
	271	*	*	*	*
	290	-0.0101	0.1000	-0.1010	0.5402
	291	0.6564	0.1000	6.5636	0.0000
	Acc	0.2385	0.1000	2.3853	0.0085
	Overall	0.2483	0.0481	5.1666	0.0000

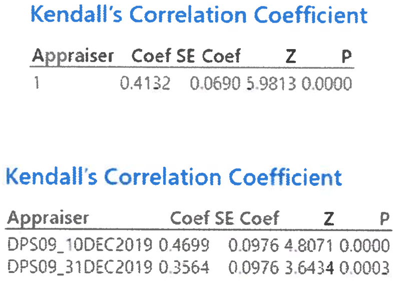
Appraiser	Response	Kappa	SE Kappa	Z	P(vs > 0)
2	140	*	*	*	*
	152	-0.0101	0.1000	-0.1010	0.5402
	170	0.2591	0.1000	2.5906	0.0048
	171	-0.0257	0.1000	-0.2567	0.6013
	173	*	*	*	*
	174	*	*	*	*
	190	0.1980	0.1000	1.9805	0.0238
	191	*	*	*	*
	240	0.2488	0.1000	2.4875	0.0064
	241	*	*	*	*
	260	*	*	*	*
	270	0.2161	0.1000	2.1609	0.0154
	271	-0.0153	0.1000	-0.1525	0.5606
	290	-0.0257	0.1000	-0.2567	0.6013
	291	0.7396	0.1000	7.3958	0.0000
	Acc	0.5776	0.1000	5.7756	0.0000
	Overall	0.3490	0.0480	7.2663	0.0000

Appraiser	Response	Kappa	SE Kappa	Z	P(vs > 0)
3	140	*	*	*	*
	152	*	*	*	*
	170	0.1828	0.1000	1.8284	0.0337
	171	*	*	*	*
	173	*	*	*	*
	174	*	*	*	*
	190	0.7469	0.1000	7.4685	0.0000
	191	*	*	*	*
	240	0.1246	0.1000	1.2456	0.1065
	241	*	*	*	*
	260	*	*	*	*
	270	0.3688	0.1000	3.6885	0.0001
	271	*	*	*	*
	290	*	*	*	*
	291	0.3125	0.1000	3.1249	0.0009
	Acc	0.6886	0.1000	6.8865	0.0000
	Overall	0.4844	0.0503	9.6357	0.0000

** When all sample standards and responses of a trial(s) equal the value or none of them equals the value, kappa cannot be computed.*

Each auditor vs. experts, Kendall's statistics (Risk Level Only)

This is the Kendall's correlation coefficients. The coefficient of the top represents the ordered correlation between the two output of auditor 1 (or 9), whereas the two coefficients of the bottom represents the ordered correlation between the two assessments of auditor 1 (9) to the experts'. Note that, unlike other examples listed on the Appendix, this example is generated from the scenario where all risk codes were grouped by risk levels ranging from 1 to 4, with higher number representing higher risk level. This is because the main advantage of the Kendall's coefficient, as opposed to a regular Pearson's and the Fleiss' Kappa, is that it adjusts its coefficients of ordinal variables by how "close" or "far" the two paired values are.



Kendall's Correlation Coefficient

Appraiser	Coef	SE Coef	Z	P
1	0.4132	0.0690	5.9813	0.0000

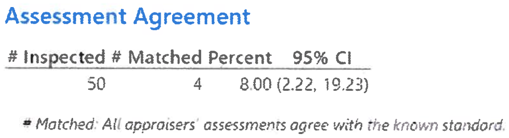
Kendall's Correlation Coefficient

Appraiser	Coef	SE Coef	Z	P
DPS09_10DEC2019	0.4699	0.0976	4.8071	0.0000
DPS09_31DEC2019	0.3564	0.0976	3.6434	0.0003

Figure 12 Kendall's correlation coefficients

All auditors vs. experts, assessment agreement

The figure below represents the assessment agreement among all auditors AND the experts. Out of 50 tickets inspected, four (8 % of all tickets) were unanimously agreed upon by the three auditors, as well as the experts, in both dates. Note that this percentage is smaller than that when assessment agreement was calculated only amongst the auditors. This is because in one ticket, the experts did not agree with the decision that was unanimous amongst auditors.



Assessment Agreement

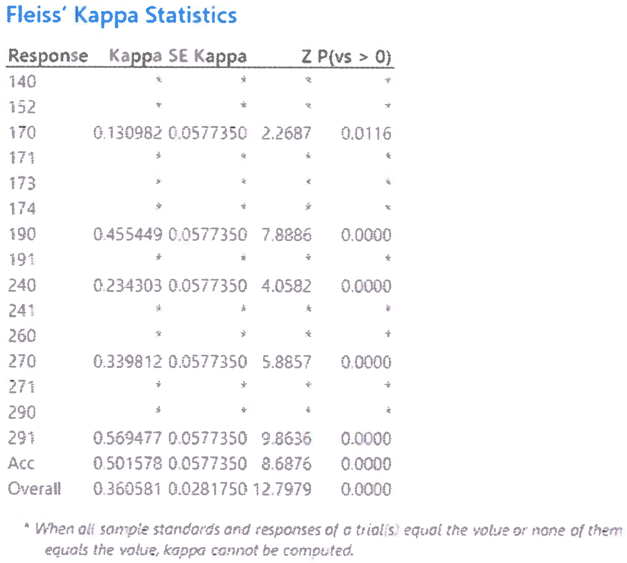
# Inspected	# Matched	Percent	95% CI
50	4	8.00	(2.22, 19.23)

# Matched: All appraisers' assessments agree with the known standard.

Figure 13 Assessment agreement of each auditor vs. experts

<u>All auditors vs. experts, Fleiss' Kappa</u>

Lastly, this is the Fleiss' Kappa of the three auditors, both times, against the experts. Kappa values never reach above 0.90, indicating poor agreement between the auditors and the experts.



Fleiss' Kappa Statistics

Response	Kappa	SE Kappa	Z	P(vs > 0)
140	*	*	*	*
152	*	*	*	*
170	0.130982	0.0577350	2.2687	0.0116
171	*	*	*	*
173	*	*	*	*
174	*	*	*	*
190	0.455449	0.0577350	7.8886	0.0000
191	*	*	*	*
240	0.234303	0.0577350	4.0582	0.0000
241	*	*	*	*
260	*	*	*	*
270	0.339812	0.0577350	5.8857	0.0000
271	*	*	*	*
290	*	*	*	*
291	0.569477	0.0577350	9.8636	0.0000
Acc	0.501578	0.0577350	8.6876	0.0000
Overall	0.360581	0.0281750	12.7979	0.0000

** When all sample standards and responses of a trial(s) equal the value or none of them equals the value, kappa cannot be computed.*

Figure 14 Fleiss' Kappa of the three auditors against the experts

Appendix B – Mid-term Financial Status Report

FEDERAL FINANCIAL REPORT

(Follow form instructions)

1. Federal Agency and Organizational Element to Which Report is Submitted US Department of Transportation Pipeline and Hazardous Materials Administration	2. Federal Grant or Other Identifying Number Assigned by Federal Agency (To report multiple grants, use FFR Attachment) 693KK31940021PSDP	Page 1	of 1 pages

3. Recipient Organization (Name and complete address including Zip code)
Virginia Utility Protection Service, Inc.
1830 Blue Hills Circle NE Roanoke, VA 24012

4a. DUNS Number 146011619	4b. EIN 55-0859075	5. Recipient Account Number or Identifying Number (To report multiple grants, use FFR Attachment)	6. Report Type □ Quarterly □ Semi-Annual □ Annual □ Final	7. Basis of Accounting □ Cash □ Accrual

8. Project/Grant Period From: (Month, Day, Year) September 18, 2019	To: (Month, Day, Year) September 27, 2020	9. Reporting Period End Date (Month, Day, Year) April 15, 2020

10. **Transactions**	Cumulative
(Use lines a-c for single or multiple grant reporting)	
Federal Cash (To report multiple grants, also use FFR Attachment):	
a. Cash Receipts	0.00
b. Cash Disbursements	0.00
c. Cash on Hand (line a minus b)	0.00
(Use lines d-o for single grant reporting)	
Federal Expenditures and Unobligated Balance:	
d. Total Federal funds authorized	100,000.00
e. Federal share of expenditures	0.00
f. Federal share of unliquidated obligations	0.00
g. Total Federal share (sum of lines e and f)	0.00
h. Unobligated balance of Federal funds (line d minus g)	100,000.00
Recipient Share:	
i. Total recipient share required	0.00
j. Recipient share of expenditures	0.00
k. Remaining recipient share to be provided (line i minus j)	0.00
Program Income:	
l. Total Federal program income earned	0.00
m. Program income expended in accordance with the deduction alternative	0.00
n. Program income expended in accordance with the addition alternative	0.00
o. Unexpended program income (line l minus line m or line n)	0.00

11. Indirect Expense	a. Type	b. Rate	c. Period From	Period To	d. Base	e. Amount Charged	f. Federal Share
				g. Totals:			

12. *Remarks: Attach any explanations deemed necessary or information required by Federal sponsoring agency in compliance with governing legislation:*

13. **Certification: By signing this report, I certify that it is true, complete, and accurate to the best of my knowledge. I am aware that any false, fictitious, or fraudulent information may subject me to criminal, civil, or administrative penalities. (U.S. Code, Title 18, Section 1001)**

a. Typed or Printed Name and Title of Authorized Certifying Official Rick F. Pevarski President & CEO	c. Telephone (Area code, number and extension) (540) 283-2520
	d. Email address rpevarski@va811.com
b. Signature of Authorized Certifying Official	e. Date Report Submitted (Month, Day, Year) April 7, 2011
	14. Agency use only:

Paperwork Burden Statement
According to the Paperwork Reduction Act, as amended, no persons are required to respond to a collection of information unless it displays a valid OMB Control Number. The valid OMB control number for this information collection is 0348-0061. Public reporting burden for this collection of information is estimated to average 1.5 hours per response, including 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 the burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to the Office of Management and Budget, Paperwork Reduction Project (0348-0060), Washington, DC 20503.