

*Journal of*

---

# **INDUSTRIAL TECHNOLOGY**

---

*Volume 21, Number 3 - July 2005 through September 2005*

---

## ***Developing a Practical Quality System to Improve Visual Inspections***

*by Mr. Rathel R. Smith, Dr. Neal Callahan and Dr. Shawn Strong*

*Peer-Refereed Article*

**KEYWORD SEARCH**

**Manufacturing  
Metrology  
Quality  
Research  
Research Methods  
Statistical Methods**

---

*The Official Electronic Publication of the National Association of Industrial Technology • [www.nait.org](http://www.nait.org)*

© 2005

---



Mr. Rathel R. (Dick) Smith is an assistant professor of Industrial Management at Southwest Missouri State University. Mr. Smith is also an active consultant to the manufacturing community, specializing in quality systems and ISO 9001.



Dr. R. Neal Callahan is an Assistant Professor in the Department of Industrial Management at Southwest Missouri State University. His research and teaching interests include computer integrated manufacturing and quality systems. Dr. Callahan received his Ph.D. in Engineering Management from the University of Missouri-Rolla.



Dr. Shawn Strong is an associate professor and department head of Industrial Management at Southwest Missouri State University. Dr. Strong teaches courses in project management, fluid power, electronics, and mechanical systems. Dr. Strong received his PhD from Iowa State University.

# Developing a Practical Quality System to Improve Visual Inspections

by Mr. Rathel R. Smith, Dr. Neal Callahan and Dr. Shawn Strong

## Introduction

While quality assurance programs place considerable emphasis on using variable data to control process and product conformance, there is often unavoidable reliance on visual inspections of attribute characteristics, which is often subjective in nature. Attribute characteristics include aspects such as physical appearance, cleanliness, shades of color, etc. While in some instances these characteristics can actually be measured, the cost of suitable equipment is often excessive or the evaluation process too time consuming when compared to the results. Thus, there needs to be an effective evaluation system to improve visual inspections to fill this gap. Reliance on subjective human judgment, and the variation in those judgments when compared from person to person, creates quality evaluation problems inhibiting the effective use of visual inspection.

The importance of gauge control in traditional variable data systems is well established (Besterfield 2004). Evaluating the effectiveness of the measuring system is an important part of quality control and process improvement activities. For any measurement being reported: Observed value = True value + Measurement Error (Montgomery 2005). When applying gauge control methods, measurement error is split into two categories: Repeatability, which is caused by equipment variation, and Reproducibility, which is a result of appraiser variation (Besterfield 2004). Little research has been done in the area of visual inspection gauge control, even though this method is commonly relied upon in many areas of industry including manufacturing and construction. For the purposes of this article, repeatability in visual inspection is the capability of an individual

inspector to be consistent in their own evaluations while reproducibility is the capability of a group of inspectors to be consistent with each other.

## Purpose and Scope

The purpose of this article is to demonstrate the practical application of a quality system process that can be used to improve consistency in visual inspections where the use of attribute measuring equipment is either too cost-prohibitive or too time-consuming. Reducing variation in the shop floor environment leads to remarkable productivity improvements (Emiliani 1998). Although the material used in this study (wood) has several interrelated attributes, this article focuses on those characteristics that directly influence the effective utilization of wood for manufactured components. The scope of this article includes an overview of the proposed quality system, including development, methodology of the study, analysis of collected data, and a summary of findings.

## Visual Inspection Improvement Quality System Overview

Any system developed should have mechanisms built in to allow for continuous monitoring and improvement (Eldar and Ronen 1995; Donnabedian 1986; Williamson 1978). Developing and implementing a quality system to improve visual inspections created several challenging areas for this study. Of particular concern was the development of a model for the proposed system while considering documented and accepted quality systems practices and the application of those principles to a manufacturing environment. The quality system model developed through this study (see Figure 1) illustrates the structure and systematic process used.

**Methodology**

**Background**

Working on a quality improvement project with a wood products manufacturer presented an ideal opportunity to conduct this exploratory study to test the feasibility and effectiveness of the proposed quality system model in a real world setting. Two limitations to this study were the fact that only three individuals were designated as quality inspectors and the time constraints associated with removing those individuals from the production floor for participation in the study. According to Besterfield (2004), the number of parts, appraisers, and trials often vary in a gauge control study. However, two to three trials and two to three appraisers are optimum.

Effective utilization of raw materials (wood) is a major profitability issue for this wood product manufacturer. They use a general term “proof” as one measure of effective utilization. In its simplest definition, proof is leaving the appropriate amount of white wood, knots, saw marks and other imperfections while processing component boards. During preliminary rough-cut and plane operations, complete removal of these imperfections causes excessive wood waste while leaving the appropriate amount of imperfections has no negative impact on the appearance and functionality of the product. This evaluation of proof (imperfections/characteristics) is based on visual inspection and subjective evaluation, thus enabling the testing of this quality system.

**Identify Critical Characteristics**

In most instances, critical characteristics (measurable) data are defined during product and/or process development. Control of these characteristics is achieved through the assignment of a target value, allowable tolerances, and implementing a capable measurement system. When critical characteristics are more subjective-based, the desired attributes are more difficult to identify, define, standardize, measure, and control. For this study, production and quality managers provided input to identify nine distinct critical characteristics deemed appropriate to control

wood utilization. These include sap and heart side saw marks, knots, rotten and want wood, and sap side white wood.

**Define Characteristics and Assign Numeric Values**

Once critical characteristics were identified, definitions were developed to enhance inspector understanding. Word descriptions, actual samples, and pictures served as the descriptive means for characteristic classification. For analysis purposes, it was also important to assign numeric values to identified subjective characteristics. In some instances, the use of a zero or one represents yes/no evaluations. For example, the board either did or did not have saw marks. The use of a sliding scale also provided an opportunity to weight the range (such as one being minimal, five being average, and 10 being perfect).

**Training Inspectors on Characteristics**

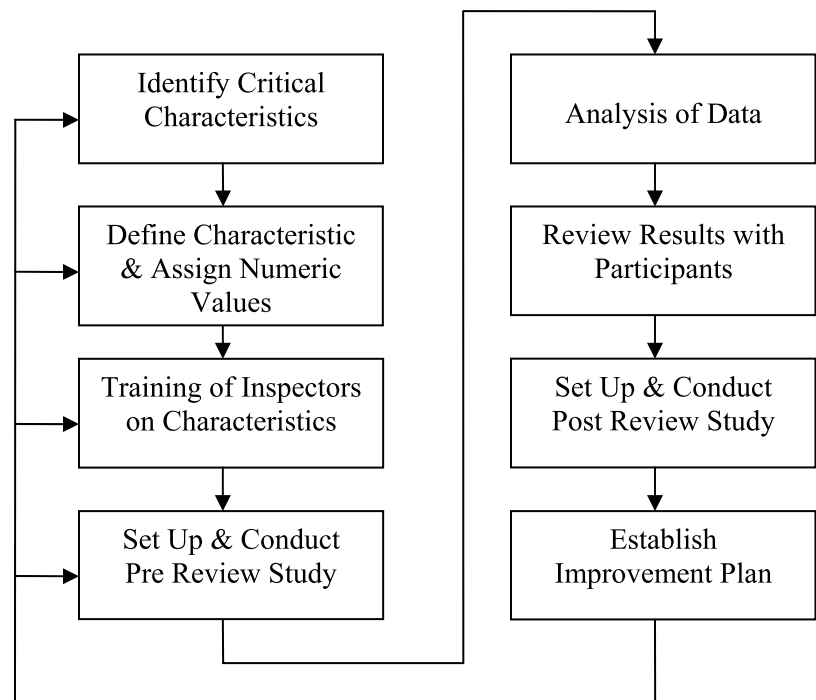
As with any new system, it is important to provide training. However, according to Motwani, Frahm, and Kathawala (1994, p. 8), “Training needs to be preceded by a well articulated strategy that employees can understand.”

Introducing inspectors to an expanded definition and numbering system required a well-planned training program. The two primary goals for this training were to give inspectors the necessary skills and tools to complete inspection evaluations consistently. During the training process for this study, sample boards, written definitions, and pictures were used to stimulate discussion and enhance learning on what specific value to assign for each characteristic type.

**Set Up and Conduct Pre Review Study**

The process used for data collection in this study was to have each inspector do an evaluation of each of fifteen selected and numbered sample boards. The numbering of the boards simplified identification by the individual coordinating the study throughout the various trials of the Pre and Post studies. Instructions were given to each participant (inspector) to visually inspect the boards and record their evaluation on a tally sheet until assessment of all characteristics on all fifteen boards was completed. The data collection process included three separate trials using the same boards. Allowing one

Figure 1: Quality System Model



to two weeks between trials minimized participants' ability to recall previous trial evaluations.

**Analysis of Data**

Analysis included comparison of each participant's results by sample and characteristic. During the three trials, each of the three inspectors evaluated fifteen boards resulting in 135 data points for each characteristic. There were nine total characteristics bringing the total to 1215 data points for both the Pre and Post studies. This number of participants and trials is comparable to that suggested by Besterfield (2004) and mentioned in the Methodology section of this paper. Three types of summary reports were created from the analysis of data as follows:

1. The Participant Evaluation by Sample and Characteristic Summary (example Table 1) was a summary report of the actual evaluations posted on the tally sheet by the three participants over the three trials. This simple report provided insight into both the repeatability and reproducibility capability of participants.
2. The By-Characteristic Summary (example Table 2 and Table 4) provided a brief but useful data analysis for each of the nine characteristics. Average and standard deviation of participant evaluations provided insight into which characteristics appeared more problematic.
3. The Participant Comparison by Characteristic (example Table 3 and Table 5) provided overall and by characteristic participant-to-participant comparison. The average of trials, range between trails, trial standard deviation and difference from group average provided insight into individual participant repeatability and group reproducibility. In addition, it provided useful information for determining which characteristics needed to be focused on in training.

**Review with All Participants**

Once the participants completed their three trials with all fifteen boards, the reports described in the previous section were used for a group review of study results. The Participant Evalu-

ation by Sample and Characteristic Summary (Table 1) for "heart side saw marks" is just one example of the nine total characteristics studied and review in this study. From the Participant Evaluation by Sample and Characteristic Summary report each participant was able to see differences in their actual evaluations from trial to trial (repeatability). For example, Participant 1 on sample board 3 did not see the heart side saw marks during Trial 1 but did during Trials 2 and 3 (see Table 1). However, Participant 3 made the same evaluation during each trail for sample boards 1, 2, 4, 5, 6 and 8. Participant 2 was inconsistent on all three trials for sample boards 6 and 12. This review process obviously caused the participant to wonder why their evaluation was not consistent from trial to trial.

Also incorporated into this review was the opportunity for each participant, to look at his or her evaluations in comparison to other participants. For example, on board 4 Participants 1 and 3 gave 0 points for all trials while Participant 2 gave scores of 8, 10, and 10 respectively. In addition, the

evaluations of board 8 were consistent across all participants while board 10 indicated confusion within and between the participants. This review certainly resulted in a group discussion to arrive at consensus as to what the evaluations should have been.

While the review discussed above allowed for individual-by-individual, characteristic-by-characteristic, and board-by-board review, participants and management were also interested in a more detailed analysis. The data shown in Table 2 provided by-characteristic average evaluation scores and standard deviations over three trials for all participants. This average also provided useful information relative to the quality of wood and capability of production operators to leave appropriate proof. The standard deviation reflected the overall repeatability capability by characteristic and thus provided insight into which characteristic presented problems in consistent evaluation for participants. These results presented opportunity for management to identify those specific characteristics that should be focused upon in future

**Table 1: Participant Evaluation by Sample and Characteristic Summary**

Characteristic: Heart Side Saw Marks									
Sample	Participant 1			Participant 2			Participant 3		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
1	10	10	10	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	0	10	10	8	10	10	0	10	10
4	0	0	0	10	8	8	0	0	0
5	10	10	10	0	0	0	8	8	8
6	0	0	0	8	5	10	0	0	0
7	8	8	8	0	0	0	8	10	10
8	10	10	10	8	10	10	10	10	10
9	10	10	10	0	0	0	8	10	0
10	0	8	8	10	5	10	0	0	10
11	8	8	8	0	0	0	10	8	0
12	10	10	8	5	10	0	10	8	0
13	0	0	0	0	8	8	0	0	0
14	10	10	8	10	10	10	8	8	10
15	0	0	0	0	0	0	10	10	8

definition improvement and participant training. One side benefit with this report is that it also gave management insight into what characteristics should be focused upon in operator training to improve overall utilization. In addition, this data established a reference point to compare effectiveness of this Review with All Participants phase of the defined quality system after the Post Review Study. This report was also shared with participants (inspectors) in an effort to keep them totally involved in the entire process.

A third report, the Participant Comparison by Characteristic (Table 3) was prepared to evaluate the consistency within each participant and allow for comparison. The average of trials and difference from group average were used to evaluate the relative evaluation range (high- low) of participants compared to the group. Participant 3, at 18.6222 evaluated lower than Participant 2, at 23.7111 base on the overall evaluation score for all boards. This report was also prepared for each of the nine characteristics such that participants could identify which specific characteristics contributed to overall participant differences. Additionally, the range between trials and trial standard deviation demonstrated the individual capability of each participant to repeat their evaluations. Participant 3 with a trial standard deviation of 3.9202 as compared to Participant 2 with a 5.5964 trial standard deviation. However, Participant 3 with a difference from group average of -0.4074 and a trial standard deviation of 5.1448 demonstrated better evaluation capability. The range between participants was included in this report to remind participants that the primary objective was to improve consistency and that the lower this range, the better.

**Set Up and Conduct Post Review Study**

After the review with all participants, and revisions of the various definitions of characteristics were completed, participants underwent retraining. The purpose for additional training was to review the revised definitions and reinforce lessons previously learned. Once retrained, participants repeated the characteristic evaluations exercise after a two-month period. This allowed for conducting of a Post Review study in the same manner and using the same sample boards and data collection techniques as the Pre Review study. Data analysis for the Post study utilized the same processes as the Pre Study.

After completion of an analysis of the data captured in the Post set of trials, a determination relative to the level of improvement participants experienced between the Pre Review and Post Review studies was made. The By-Characteristic Summary and Participant Comparison by Characteristic reports from the Pre Review study provided a base line to evaluate the level of progress made by participants. On an individual characteristic basis, looking at the average standard deviation across

three trials between the Pre and Post Review studies presented additional information useful in evaluating the effectiveness of training and for discovering new improvement opportunities.

Table 4 presents summaries of individual characteristic data. There was a reduction in the average participant standard deviation between the Pre and Post Review study in six of the nine characteristics indicating improvement in repeatability. In addition, the average reduction in average standard deviation for the six was .419 as compared to .230 for the other three. Table 4 also aided in identifying problematic characteristics such as Sap Side White Wood which experienced a .400 increase in average participant standard deviation between the Pre and Post Study; thus indicating that either an improved definition needed to be developed or participants required additional training.

In addition to the Pre/ Post differences noticed in both average and standard deviation, there was a reduction for all characteristics as shown in Table 5. A reduction of 4.2222 in the range between participants relative to the average of trials suggests that there

**Table 2**

<b>By Characteristic Summary (Pre Review)</b>			
<b>Sap Side</b>		<b>Averages</b>	<b>Std Deviation</b>
	White Wood	5.67	0.36
	Missing Wood	1.30	0.98
	Knots	0.67	0.32
	Saw Marks	1.63	0.38
	Rotten	0.59	0.26
<b>Heart Side</b>			
	Missing Wood	1.10	0.83
	Knots	2.99	0.44
	Saw Marks	4.52	2.05
	Rotten	0.73	0.72

**Table 3**

<b>Participant Comparison by Characteristic (Pre Review all characteristics)</b>				
	<b>Participant 1</b>	<b>Participant 2</b>	<b>Participant 3</b>	<b>Range</b>
Average of Trials	20.5556	23.7111	18.6222	5.0889
Range Between Trials	9.4667	10.2000	7.0667	3.1333
Trial Std Dev	5.1448	5.5964	3.9202	1.6762
Difference From Group Average	-0.4074	2.7481	-2.3407	

was improvement in participant ability to reproduce their own evaluations. It is equally important, to notice the reduction, in the averages between Pre and Post studies for participants relative to both their range between trails and their trial standard deviations. Specifically, there was a reduction of 4.2889 in range between trials and 2.2957 for trial standard deviation further indicating that inspector capability to repeat their evaluations was improved.

**Establish Improvement Plan**

As with all closed-looped systems, it is important to encourage continuous improvement. After reports from the Pre and Post studies were analyzed, the following became primary comparison possibilities and potential opportunities for improvement:

1. Identification of participants whose evaluation averages did not improve in comparison to the group. This was a comparison by characteristic and all characteristics evaluation scores.
2. Identification of participants whose evaluation ranges and standard deviations did not improve in comparison to the group. This comparison was

**Table 4**

<b>By Characteristics Summary (Pre Post Review Comparison)</b>						
	Averages			Standard Deviations		
	Pre	Post	Difference	Pre	Post	Difference
<b>Sap Side</b>						
White Wood	5.67	5.10	0.58	0.36	0.76	-0.40
Missing Wood	1.30	0.37	0.93	0.98	0.64	0.33
Knots	0.67	0.71	-0.04	0.32	0.00	0.32
Saw Marks	1.63	1.78	-0.15	0.38	0.05	0.33
Rotten	0.59	0.00	0.59	0.26	0.00	0.26
<b>Heart Side</b>						
Missing Wood	1.10	2.81	-1.72	0.83	0.96	-0.14
Knots	2.99	0.72	2.27	0.44	0.59	-0.15
Saw Marks	4.52	3.56	0.96	2.05	1.50	0.55
Rotten	0.73	0.00	0.73	0.72	0.00	0.72
Average Decrease When Decreased						-0.419
Average Increased When Increased						0.230

- by characteristic and all characteristics evaluation scores. .
3. Identification of any characteristic that presented difficulty for most participants as indicated by: characteristic averages, standard deviations and a review of individual sample scores.
4. Identification of changes in the group averages between studies.

**Summary of Results**

Results from this exploratory study indicate that the proposed quality system model is useful in improving repeatability and reproducibility in visual inspections. While this system presents several steps, perhaps the most important for improving visual inspection is the ability to review Pre and Post study results with participants. The proposed

**Table 5**

<b>Participant Comparison by Characteristic (Pre Post Comparison all characteristics)</b>					
	Pre Review			Range	Average
	Participant 1	Participant 2	Participant 3		
Average of Trials	20.5556	23.7111	18.6222	5.0889	20.9630
Range Between Trials	9.4667	10.2000	7.0667	3.1333	8.9111
Trial Std Dev	5.1448	5.5964	3.9202	1.6762	4.8871
Difference From Group Average	-0.4074	2.7481	-2.3407		
	Post Review			Range	Average
	Participant 1	Participant 2	Participant 3		
Average of Trials	19.1111	18.2444	18.8000	0.8667	18.7185
Range Between Trials	4.5333	5.3333	4.0000	1.3333	4.6222
Trial Std Dev	2.5175	2.9472	2.3094	0.6378	2.5914
Difference From Group Average	0.3926	-0.4741	0.0815		
				Pre-Post Range	Pre-Post Average
				Difference	Difference
Average of Trials				4.2222	2.2444
Range Between Trials				1.8000	4.2889
Trial Std Dev				1.0384	2.2957

quality system provides vital information towards the development of an affective and individualized inspector improvement plan during the Pre and Post Reviews.

Specific components of the review process and potential for improving visual inspection results include the following: first, the individual inspector's review of their evaluations by board and by trial provides a foundation for enhancing repeatability. Second, the individual inspector's review by board and by trial of their evaluation as compared to other inspectors provides a foundation to enhance reproducibility. Third, the group discussion involving all participant inspectors provides an open forum to clarify understanding of critical characteristic and identify those characteristics that appear to be more problematic. Fourth, the group discussion provides opportunity for the individual conducting the study to target those characteristics that appear to be more problematic for improved definition or additional inspector training. Fifth, the data analysis and group discussion are beneficial to establish an improvement plan for each individual inspector involved in the study.

There is evidence that the proposed system is an effective tool for improving visual inspections. This support comes first, and foremost, from the positive feedback participants provided during the various stages of the study

and their strong belief that the entire process enhanced their understanding of what evaluations were appropriate for the various levels of characteristics. In addition, a t-test to determine whether the differences found between Pre and Post studies were statistically significant. The results are listed in Table 6. The All Characteristic Totals presented in Table 5 are the summation, by participant, for evaluations made for all characteristics. The t-test resulted in an overall p-value of 0.0045, which indicates statistical significance for the reduction of total scores from Pre to Post training.

### Recommendations for Further Study

Since all variation in this study came from human input, training methodology should be evaluated and future data analyzed to determine:

1. If statistically, acceptable levels of repeatability can, be maintained by individual inspectors.
2. If statistically, acceptable levels of reproducibility can, be maintained between groups of inspectors.
3. If this methodology, can be replicated in other settings where visual inspection is relied upon.

### Conclusion

Many production enterprises utilize visual inspection subjective data. Most manufacturers, service industries, other

business sectors, and educational entities use such data. Organizations today are better understanding the value of converting data into useful information. At the same time, managers fear making decisions from information that is not reliable due to error associated with measurement repeatability and/or reproducibility. The studied quality system has demonstrated potential to remove or substantially reduce fears relative to visual inspection, thereby improving decision-making ability; and, improved decision-making improves the potential for success.

### References

- Besterfield, D. H. (2004), *Quality Control* (7<sup>th</sup> ed.), Englewood Cliffs, New Jersey: Prentice Hall
- Eldar, R. and Ronen, R. (1995), Implementation and evaluation of a quality assurance program, *International Journal of Health Care Quality Assurance*, 8 (1), 28
- Emiliani, M. L. (1998), Continuous personal improvement, *Journal of Workplace Learning*, 10 (1), 32
- Montgomery, D. C., (2005), *Introduction to Statistical Quality Control*, New York: John Wiley and Sons
- Motwani, J. G.; Frahm, M. L.; Kathawala, Y., (1994), The key to quality improvement, *Training for Quality*, 2 (2), 8

Table 6

Pre Post Review All Characteristic Evaluation Totals and t-Test Results

	Participant 1			Participant 2			Participant 3		
PRE	352	310	263	359	341	367	280	270	288
POST	300	312	282	251	241	233	283	230	240

	p-value
All Participants	0.0045
Participant 1	0.3614
Participant 2	0.0001
Participant 3	0.0862