Evaluating Student Capability to Inspect Welding Quality Using a GRRS Technique

by

Rathel R. (Dick) Smith, MS
RathelRSmith@missouristate.edu
Department of Industrial Management
Missouri State University

W. H. (Bill) Drake, Ph.D., C.Mfg.E.
BillDrake@MissouriState.edu
Department of Industrial Management
(Retired)
Missouri State University

Jim Bridwell, C.W.E, C.W.I.
bridwell@otc.edu
Welding Division
Ozarks Technical Community College

Abstract

A technical curriculum often focuses on various theories in order to enhance understanding, but there is often also a need to provide the student with specific technical skills. In the welding process, it is not only important for students to gain mechanical skill, but also to gain a perception of the quality of welds. Visual inspection of weld joints, based on a defined set of standards, is a common and accepted practice for welding professionals. This project focuses on developing a model for evaluating and enhancing student capability in regard to visually inspecting welding quality by using Gauge Repeatability and Reproducibility Studies (GRRS) techniques.

Keywords: Quality Measurement, Gauge Repeatability and Reproducibility, Visual Inspection, Process Improvement, Welding Inspection.

I. Introduction

Previous research, conducted in 2005, by a member of this project team, demonstrated the application of Measurement Systems Analysis (MSA) techniques, such as Gauge Repeatability and Reproducibility Studies (GRRS), to visual and subjective inspection of quality characteristics in a wood products manufacturing facility. The project produced a measurable improvement in inspection results accuracy. The implication was (based on the results of the 2005 exploratory study) that applying the approaches utilized in that study to other environments where there is heavy reliance on visual inspection would also improve visual inspection accuracy. Researchers focused on a group of first-year welding students enrolled in a welding class at a local community college.

A technical curriculum often focuses on various theories in order to enhance understanding, but there is often also a need to provide the student with specific technical skills. In the welding process, it is not only important for students to gain mechanical welding skill, but also to acquire the capability to accurately evaluate the quality of welds. While technology-based inspection techniques, such as x-ray and ultrasound, are used to verify welding quality, there continues to be a high reliance on visual inspection. Visual inspection of weld joints is a commonly accepted practice based on defined standards. There is a certification process for visual welding inspection.

II. Purpose of Research

The primary focus of this project was to develop a model by which to evaluate student capability to visually inspect welded joints. The following outcome expectations were established for the model:

- The model would demonstrate an effective means to evaluate student capability to accurately evaluate welding quality through visual inspection.
- The model would provide feedback reports that would facilitate student learning .and group discussion.
- The model would be demonstrated in a real-world technical classroom.

A secondary, yet especially important, purpose of this project was to continue the research, in order to test the application of the GRRS system to other visual inspection environments.

III. Review of the Literature

Concerns About Visual Inspection

“Visual examination (VT) is the nondestructive examination method that is used most extensively for weldments [2].” However, Juran and Godfrey capture the limitations of this methodology, "Visual inspection remains the largest single form of inspection activity. For these characteristics, the written specifications seldom describe completely what is wanted, and often inspectors are left to make their own interpretation [3].” When verification of weld quality relies on visual inspection, there is an unavoidable reliance on human subjective judgment. This subjective judgment generally depends on human sensory qualities. “‘Sensory Qualities’ are those for which we lack technological measuring instruments and for which the senses of human beings must be used as measuring instruments. In common with other qualities, sensor qualities require: Establishment of product and process standards . . . .”[3].

Availability of Welding Standards

The American National Standards Institute (ANSI) and the American Welding Society (AWS) collaboratively provide process standards for welding. Some widely accepted process standards are AWS B1.11, Guide for the Non Destructive Examination of Welds [4]; AWS PHB-2, The Everyday Pocket Handbook for Visual Inspection and Weld Discontinuities [5]; and AWS D1.1/D1.1M: 2006, Structural Welding Code Steel [6] are primary references in defining the variables, attributes, and definitions that guide the visual welding quality inspection process. Standards and their definitions used in this project are primarily publications of the American Welding Society.

Gauge Repeatability and Reproducibility Studies (GRRS)

While there are widely accepted standards and nomenclature to guide welding visual inspection, there is still concern about the capability of individuals to accurately evaluate welding quality. Just as with evaluations using measuring and monitoring devices, there is concern about the repeatability and reproducibility of evaluations by visual inspection. Although the problems associated with visual inspection loom large, the industry recognized that MSA techniques exist that can improve overall visual inspection accuracy [7]. A recent project revealed that 76.7 % of the quality professionals surveyed recognized the importance of using the GRRS technique to evaluate accuracy of visual inspections; however, in actual practice only 16.6% report using it on an often-to-always basis [8]. GRRS, which has become a standard term to describe this process [7, 9], provides an estimate of how much of the observed process variation is due to measurement system variation. This is typically done by a methodical procedure of measuring or inspecting and re-measuring or re-inspection of the same parts by different participants. According to Raffaldi and Kappele [10], "If measurement variation can be reduced and gauge repeatability and reproducibility ratios improved, it is easier to differentiate between parts that are in or out of specification, allowing parts to be accepted or rejected with greater confidence." In a recent study where GRRS was applied to multiple visual inspection data, significant improvement in accuracy was realized through an application of review and discussion of GRRS results [1].

Recognition of a Need to Evaluate Skills Learning

Throughout a variety of professions, there appears to be recognition of a need to develop and use practical means to evaluate and improve skills. For example, in the medical community bench top tying and suturing skills are being evaluated to determine capability among surgeons of various experience levels [11]. GRRS techniques are being used to evaluate and improve visual inspection results in a wood products manufacturing organization [1]. Occupational Safety and Health requires training for employees operating a fork truck. A required component of this training is the use of a process that includes the evaluation of knowledge through either verbal or written questions and actual observation of employees operating such equipment [12].

Absence of Literature on Welding Inspection Evaluation

While the literature review did provide standards and guidelines for the welding inspection process, members of this research team found no reference to methodology used to evaluate repeatability and reproducibility of welding inspection results. As a result of literature searches that included welding-specific journals and related literature, the research team concluded that the development of a system to evaluate the welding practitioner's and welding inspection professional's capability to make accurate judgments in evaluation of welds based on these standards is a missing component of welding and welding inspection instruction.

IV. Project Methodology

The model proposed by Smith, Callahan, and Strong [1] utilized a two-trial GRRS evaluation in which participants evaluated selected samples for multiple characteristics, followed by an analysis to define participant capability in terms of their repeatability (can individual evaluators replicate their own evaluations?) and reproducibility (can the group of evaluators get the same results?). Reports from this analysis became information for group discussion relative to characteristic definitions. The group discussion was followed by a repetition of the two trials, and the data analysis. This allowed a pre- and post-comparison to determine the level of improvement gained by using the technique. Unlike the 2005 study there was insufficient time in a regular semester to run both pre- and post-review studies; therefore, the team modified the model as shown in Figure 1 [1].

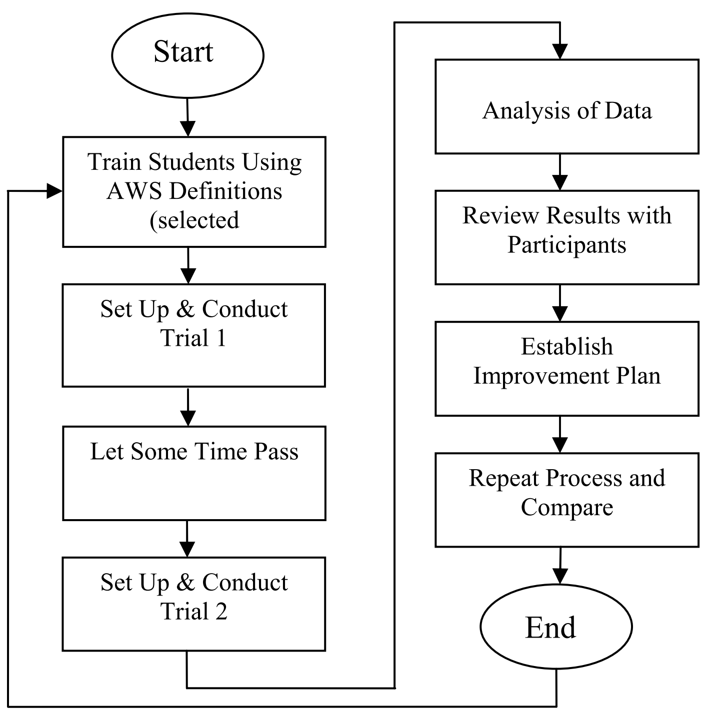

***Figure 1*: Evaluation System Model for Welding**

The first step in the process, Train Students Using AWS Definitions, was completed in both classroom and laboratory settings. Except for the introduction of the Welding Quality Evaluation project, classroom and laboratory activity followed the routine teaching model commonly utilized at the Technical Community College. The teaching model presentation in the classroom instruction began with a review of the acceptance criteria from the ANSI/AWS D1.1–2004 [6]. Discussions focused on the categories for structures and how they applied to different types of structures. The discussion of weld discontinuities included; cracks, surface porosity, incomplete joint penetration, incomplete fusion, undercut, under fill, excessive root reinforcement, excessive face reinforcement, and splatter. As part of the discussion, students looked at examples of each of these different discontinuities and applied the acceptance criteria from ANSI/AWS D1.1-2004 [6] to determine how much was acceptable, what caused the discontinuity, and how the defect might be repaired or prevented. During laboratory time, students examined some of the common discontinuities and worked on ways to prevent them.

The second step in the process, Set Up and Conduct Trial 1 Study, was handled as follows: Sixteen weld-joined samples were created for use in this project. Each of the created samples contained one or more of the identified discontinuity types presented in the classroom. In addition, discontinuities were also at various levels in different samples; as a result, for many of the samples, the existence or nonexistence of a specific discontinuity was truly a matter of subjective judgment. In other words, there was not a predetermined set of answers. A tally sheet for recording inspections results was developed to facilitate data collection. In some instances, there were discontinuities in the samples not included on the tally sheet. In this project, participants were asked to make a personal judgment regarding the sample having or not having discontinuities. This personal judgment required participants to apply sensory judgments in the visual inspection of samples, following the ANSI/AWS definition for the discontinuity. The research team selected twelve samples from the total of sixteen prepared samples for the study. Basically, participants were asked to do a visual inspection, of each of the samples and record their inspection results on a tally sheet.

The third step in the process, Let Some Time Pass, was critical to the GRRS process. The primary purpose of this step was to allow time for participants to forget their evaluation results from Trial 1. To further minimize the memory effect, samples were assigned random numbers and were mixed-up between trials. The study coordinator maintained the sample's true identity for data analysis purposes.

The fourth step in the process was Set Up and Conduct Trial 2. After a few days the same samples were presented again as a Trial 2, and participants recorded their inspection results. This step was completed using the same samples and methods used in Trial 1.

The fifth step in the process was Data Analysis. Once the GRRS project team had data from both Trial 1 and 2, an analysis was completed to evaluate repeatability (within participant) and reproducibility (between participants) agreement. An Excel spreadsheet was developed for analysis purposes. Further detail about the types of reports produced, through the spreadsheet, can be found the Project Outcomes section below.

The sixth step in the process, Review Results with Participants, was completed in the classroom. Participants were provided the reports mentioned above. In addition, the trial samples were provided to participants for further review and discussion relative to the existence or nonexistence of discontinuity types.

The seventh step, Establish Improvement Plan, resulted from each participant being able to identify specific discontinuity types that were consistently inaccurate or inconsistent.

The eighth and last step in the process was Repeat the Process and Compare, results. As mentioned previously there was not sufficient time in the semester to repeat the process. The research team strongly believes, however, that this step should be included in the model and used as time permits.

V. Project Outcomes

The project was first introduced to the student participants approximately six weeks into the semester. The project and its purpose were presented to the student participants in a classroom setting. From an original student class population of thirteen, six students actually completed both trials. Class, student, and instructor time constraints prohibited the making up of missed sessions; therefore, data on the other seven students was incomplete.

The basic analysis model proposed by Smith, Callahan, and Strong [1] was modified and simplified for use in an educational environment by technical instructors. Three summary reports were prepared for review with participants as follows:

- The Participant Repeatability Report was prepared for each participant. The report documented the individual participant's evaluations during the two trials by sample and discontinuity.
- The All-Participant Repeatability Report was prepared and shared with the group. The report was simply a summary of the instances in which the participant's evaluation in Trial 1 did not match the evaluation in Trial 2.
- The By Discontinuity Reproducibility Report was prepared and shared with the group. A report was prepared for each discontinuity included in the study. This report allowed participants to see the degree to which they agreed or disagreed with other participants.

Collectively, these reports served to allow both the participants and the instructor to evaluate participant capability. In addition, because there was not an answer key, the reports stimulated discussion about the existence or nonexistence of discontinuities in marginal samples.

The research team also completed statistical analysis of the results using both correlation and analysis of variance between the two trials. While these techniques indicated the magnitude of overall inaccuracy, they were not an effective tool to enhance learning.

Participant Repeatability Report

The data presented in Table 1 below is based on the actual subjective evaluations provided by Participant 2. The report provided insight for both the participant and the instructor into the participant's capability to evaluate welding samples consistently.

Table 1 ***Participant 2 Repeatability Report***

Sample	Cracks		Surface Porosity		Incomplete Joint Penetration		Incomplete Fusion		Undercut		Under Fill		Excess Root Reinforcement		Excess Face Reinforcement		Spatter		T1 ≠ T2
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	
A	1	1	1	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	3
B	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0
C	1	1	1	0	1	1	1	0	0	0	0	0	0	0	1	0	1	1	3
D	0	0	1	1	1	1	0	0	1	1	0	1	0	0	1	0	1	1	2
E	0	0	1	0	1	1	0	1	0	1	1	0	1	0	0	0	1	1	5
G	1	1	1	0	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1
J	1	0	1	0	1	1	0	0	1	1	1	1	0	0	0	0	0	0	2
K	0	1	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1	6
L	0	0	1	0	1	1	0	0	1	1	0	0	0	0	1	0	1	1	2
M	0	1	0	1	1	1	1	1	0	0	0	0	0	0	1	0	1	0	4
O	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	1	1	1
P	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1	0	0	1	3
T1 ≠ T2	3		7		3		4		3		3		1		5		3		32

In the report a *0* indicated no discontinuity reported, and a *1* indicated that the participant reported a discontinuity. *T1* and *T2* indicated Trial 1 and Trial 2 respectively. Summary data, listed in the right column and bottom row, assisted both the participant and instructor in interpretation. Both data were the sum of incidents in which the reported evaluation in Trial 1 did not match Trial 2. For example, Participant 2 on sample M had a total of four instances in which evaluations in Trial 1 and Trial 2 did not match; therefore, special attention was given to that sample by Participant 2 in order to better understand what was seen and/or not seen. Perhaps more important were the totals of mismatches across the bottom row, which helped both the instructor and participant identify those more problematic discontinuities for the participant. For example, Surface Porosity, with 7 mismatched incidents appeared to be a discontinuity that would require further study as a part of the improvement plan. The grand total of 32 mismatched incidents, was used to compare Participant 2 to other participants and could have been used for further comparison purposes had the process been repeated.

All-Participant Repeatability Report

The data presented in Table 2 below is based on the actual subjective evaluations provided by participants in the two trials. The report provided insight into how each participant's repeatability compared other participant's and which samples appeared to be more problematic.

Table 2 ***All-Participant Repeatability Report***

Sample	Participant #						All-Participant Average
	1	2	3	4	5	6	
A	1	3	4	3	2	3	2.67
B	3	0	2	1	2	2	1.67
C	2	3	3	0	0	2	1.67
D	1	2	0	1	2	2	1.33
E	3	5	3	0	1	3	2.50
G	1	1	2	0	1	3	1.33
J	3	2	3	0	2	3	2.17
K	2	6	3	1	0	3	2.50
L	2	2	3	2	1	3	2.17
M	4	4	2	0	0	3	2.17
O	3	1	3	1	0	0	1.33
P	1	3	3	0	2	6	2.50
Total	26	32	31	9	13	33	24.00

The participants were guided in a discussion of the results found in the bottom row of the report. Participant 6, with 33, and Participants 2 and 3, with 32 and 31, respectively, obviously had more instances of mismatch than Participants 4 and 5, with 9 and 13, respectively. In addition, this report was beneficial to the instructor in determining problematic samples for future studies. For example, samples K, E, and P, with an average mismatch of 2.50, appeared to be the most difficult samples; while samples D, G, and O, at 1.33, were perhaps not difficult enough.

Undercut Discontinuity Reproducibility Report

The data presented in Table 3 below is based on the actual subjective evaluations provided by all participants. The report was prepared for each discontinuity and each trial. It provided both participants and instructor insight into the reproducibility capability of the group.

Table 3 Undercut Reproducibility Report Trial 1

Sample	Participant #						All-Participant Average
	1	2	3	4	5	6	
A	1	0	0	1	1	1	0.67
B	0	0	1	0	1	0	0.33
C	0	0	0	0	0	0	0.00
D	1	1	1	1	1	1	1.00
E	1	0	1	1	1	1	0.83
G	0	1	0	0	0	1	0.33
J	1	1	1	1	1	1	1.00
K	0	1	0	0	0	0	0.17
L	1	1	1	0	1	1	0.83
M	0	0	0	0	0	0	0.00
O	0	0	0	0	0	0	0.00
P	0	1	0	0	0	0	0.17
Total	5	6	5	4	6	6	5.33

The actual evaluation by all participants by sample was included in the report. The all-participant average was difficult to explain, but still valuable. Notice that samples C, M, and O averaged *00.0,* and samples D and J averaged *1.00,* which indicates perfect agreement (reproducibility) between participants. Any average not a perfect *0.00* or *1.00* represents disagreement, and the closer the average to *.50*, the higher the level of disagreement. The instructor reinforced agreement and used the samples to guide students to a better understanding of disagreement.

While not readily obvious, the all-participant average did suggest what the answer might be. In those few instances in which the group agreed on the existence or nonexistence of a discontinuity, this agreement may in fact have been the true answer. It is important to understand that there was not a perfect answer otherwise this report would have compared participant evaluations to the known standard.

VI. Conclusion and Recommendations

Given the stated outcome expectations for the project, the following outcomes were achieved:

- The project demonstrated an effective means by which to evaluate student capability to accurately evaluate welding quality through visual inspections. The system presented here and the obtained results suggest that the expectation was achieved. In addition, instructor and participant feedback reinforced the accomplishment.
- The project provided feedback reports that facilitated student learning and group discussion. The three feedback reports not only provided outcomes, but stimulated extensive discussion and focused learning.

- The project was conducted in a real-world technical classroom at a community college with six fully-participating students.

The next step for the project will be to repeat and further prove the effectiveness of the process. The additional study should be conducted in a similar environment, but with sufficient time to complete two additional trials. Further, an Excel spreadsheet should be utilized for completing trials in a self-directed mode at a laboratory computer workstation, which would conserve valuable instructor and classroom time.

References:

1. Smith, R., Callahan, R., & Strong, S. (2005), Developing a Practical Quality System to Improve Visual Inspection. *Journal of Industrial Technology, 21*(3), July-September, 1-7.

2. Welding Handbook, Vol. 1 (2001), *Welding Science & Technology*, (9th ed); Miami, Florida, American Welding Society.

3. Juran, J., & Godfrey, A. (1999). *Juran's Quality Handbook*, 5th edition. New York: McGraw-Hill.

4. *AWS B1.11: 2000, Guide for the Non Destructive Examination of Welds*, (2000), Miami, Florida, American Welding Society. .

5. *AWS PHB-2, The Everyday Pocket Handbook for Visual Inspection and Weld Discontinuities*, (2000), Miami, Florida, American Welding Society.

6. *AWS D1.1/D1.1M: 2006 Structural Welding Code Steel*, (2006), Miami, Florida, American Welding Society.

7. *AIAG, Measurement Systems Analysis, Reference Manual*, Third Edition. Southfield, Michigan: Automotive Industry Action Group.

8. Smith, R., McCrary, S., & Callahan, R. (2007), Gauge Repeatability and Reproducibility Studies (GRRS) and Measurement System Analysis: A Multimethod Exploration of the State of Practice, *Journal of Industrial Technology, 23*(1), 2-12.

9. Besterfield, D. H. (2004), *Quality Control*, 7th edition. Englewood Cliffs, New Jersey: Prentice Hall.

10. Raffaldi, J., & Kappele, W. (2004). Improve Gauge R&R Ratios. *Quality*, January, 48-54.

11. M.B., B.Sc. Simon D. Bann, M.B., Ch.B. Mansoor S. Khan, M.D. Ara W. Darzi, (2003), Measurement of Surgical Dexterity Using Motion Analysis of Simple Bench Tasks, *World Journal of Surgery, 27*(3), April, 390-394.

12. Powered Industrial Truck Operator Training; *Final Rule - 63*:66237-66274 (1998, December 1), Retrieved November 13, 2006, from http://www.osha.gov/index.html.