

The Subjective Judgement of Weld Quality and its Effect on Production Cost

Anna Ericson Öberg^{2,3}, Erik Åstrand^{1,3}

¹ Volvo Construction Equipment, Braås, Sweden. ² Volvo Construction Equipment, Arvika, Sweden. ³ Department of Materials and Manufacturing Technology, Chalmers University of Technology; Gothenburg, Sweden.

Abstract. The focus is often on new material, improving weld methods or increased automation when cutting production costs within the welding industry. However, the way the quality inspection is conducted can as well cause not negligible production costs.

Visual inspection is a common way to examine certain quality properties of welds. The performance of the visual inspection have been analyzed using attribute agreement analysis where experienced auditors evaluated welded parts visually for certain defects in a predefined order. The result showed a visual inspection system not capable of auditing parts the same way. The auditors evaluated differently, not only between the appraisers, but also within by assessing the same part several times contrarily.

Another test was performed where personnel from different functions within the company visually evaluated the quality of test samples. This test also showed a vast variation in the subjective evaluation. The evaluated quality score varied throughout the whole scale from “good” to “need to be repaired.”

An incapable inspection can have a severe effect on the production costs. If a defect product finds its way to the customer, without being detected in the inspection, the consequences can of course be immense. If defect free products are evaluated as defect, this creates waste in the form of unnecessary repairs and process changes. An even larger amount of waste, however not as visible, lies within the area of unused capacity and process improvements that are not executed due to improper evaluation. An example from the evaluation tests shows a possible cost reduction of up to as much as 24% for a sample that was approved by part of the appraisers.

The solution is not necessarily to automate the inspection but rather to define the actual information needed and the demands made on the evaluation method.

Keywords: weld quality, audit, judgement, visual inspection, attribute agreement analysis, type I error, type II error

1. Introduction

The question “What is considered good weld quality” could probably be answered in as many different ways as the people you ask. Depending on e.g. the load conditions the properties of the weld could have different influence. Garvin propose eight critical dimensions or categories of quality that can serve as a framework for a strategic analysis: performance, features, reliability, conformance, durability, serviceability, aesthetics and perceived quality (Garvin 1987). Since it is impossible to ask each customer about the desired quality in each of these areas, standards are used. Such standards contain e.g. limits for certain defects in respect to the desired quality decided by designers and analysers (Swedish Standard 2004; Volvo Group 2011).

No matter if a standard, or more direct customer demands, are used to set the production targets, it is necessary to assure that the products fulfill the specified requirements. Different functions hence need different types of information about the product or the process in order to make decisions (Öberg et al. 2012). It is obvious that it is necessary to assess the compliance of the product with the customer demands in order not to disappoint the customer. Equally essential is however to be able to get information about production processes and if they are run in the best possible way also in a cost perspective. To make decisions about necessary actions, improvements and process effectiveness a basis of information is needed.

Visual inspection within welding industry is commonly used. It is the cheapest method, since it is not connected to any investments in equipment or consumables. Limitations, like the sometimes subjective, variable, and generalized nature of this visual data however makes it less desirable for certain applications (Chase and Ghasemi 2008). A vast amount of research has been performed in the area of automated quality inspection in order to rule out the drawbacks connected to the human intervention. Automated solutions may as well have drawbacks necessary to consider, e.g. automation bias meaning uncritical reliance on a system without recognizing its limitations and possibilities of failure according to Bertovic et al (Bertovic et al. 2012).

According to Dickens and Bray (Dickens and Bray 1994) any “real world” situation is not ideal and results in two kinds of operational errors. These can be classified into one of two categories; Type I or Type II. Type I error means missing a defect and Type II means false calls. Ali et al (Ali et al. 2010) states that there is an optimal balance between false calls and missed defects that is industry dependent. Within non-safety critical welding industries visual inspection is due to its flexibility, speed and low cost therefore in some cases the preferred method.

The variation in evaluation result affects the production cost in different ways. Indirect effects are connected to decisions about changing the *process* or not whereas direct effects are more connected to decisions concerning the *product* itself. The direct effects are connected to the consequences if the appraiser either misses an actual defect or convicts a part without defects. The cost for not detecting, and stopping, defects from reaching the customer can be huge, such as compensation and replacement. Repairs, replacement, and additional operations that are unnecessary would be the result in the “false call” case. Since the production is optimized for production and not repairs, there is also an additional risk that the repairing itself introduces defects or reduced quality.

The indirect effects are more hidden in comparison to the direct effects. The indirect effect is more connected to when the variation in judgement initiates changes of the process (or in some cases leaves it unchanged when it should have been changed). Higher production costs could be the consequence and among others Stenbacka (Stenbacka 2009) and Ericson Oberg et al (Ericson Öberg et al. 2012) have previously described how substantial financial impact for example a deviation of weld throat size could have.

This leads to the question how does the subjective judgement influence the cost within a non-safety regulated welding company?

2. Empirical Study: Visual Judgement

Two investigations have been conducted; one investigation studying visual inspection performed by auditors and a second investigation studying the experienced quality impression.

2.1. Investigation I – Visual Inspection

The participants in the investigation were 16 auditors from 12 factories in seven countries within the same company. 16 welds were selected to be evaluated. The area to evaluate was marked and the weld demand was presented next to the weld. For each weld 15 defect types were going to be investigated. All auditors evaluated the same weld three times during the trial. The inspection order was fully randomized. For each weld the auditor should decide if the weld was defect free or not and in that case how severe the defect was (four levels). In total the investigation contained 11 520 answers. The test was planned by the same person that analysed the result but was facilitated by two other persons.

2.2. Investigation II – Experienced Quality Impression

To study the phenomena of the subjective weld quality culture within a large organization and its effect of the fabrication cost, an investigation were 43 persons were asked to give their own opinion of the welds appearance of different weld

specimens were conducted. 43 persons participated in the study evaluating 10 welded samples. The samples had been manufactured using different weld parameters which created a different appearance. The participants could choose between four different levels when assessing the weld quality: Good, Acceptable, Not acceptable, Not acceptable and needs to be repaired.

3. Result

3.1. Investigation I – Visual Inspection

The auditors were divided into experienced and less experienced, before the analysis of the result. Only auditors whose primary task is to perform weld audits were included in the analysis.

The visual inspection was analysed using an Attribute Agreement Analysis since the data was considered to be attribute data. An example of evaluation method to use for ordinal data is Kendall's coefficient of concordance which measures the associations among ratings and do not treat misclassification equally. The coefficient ranges from 0 (no agreement) to 1 (complete agreement) where coefficients above 0.9 are considered very good. The result shows very low values on the Kendall's coefficient of concordance, 0.42, meaning that the auditors had problems to classify the welds with high agreement. The p-value is less than alpha (0.05) for all ratings. That means that the null hypothesis - that the agreement between appraisers is due to chance - can be rejected.

To see if the result was connected to difficulties dividing the defects into four defect levels, the data was separated into ok/not ok without levels. Fleiss' Kappa was calculated for the divided data. Fleiss' Kappa ranges from < 0 (no agreement) to 1 (complete agreement) where values above 0.9 are considered very good. The result from the attribute agreement analysis of the judgements, only divided into ok/not ok, also shows very low values on Fleiss' Kappa, 0.39, meaning that the auditors had problems to classify the welds with high agreement. The p-value is less than alpha (0.05) for all ratings. That means that the null hypothesis - that the agreement between appraisers is due to chance - can be rejected.

3.2. Investigation II – Experienced Quality Impression

The participants in the investigation were divided into four categories depending on their profession and experience in welding; welders, quality personnel, design and customers. The category "customers" was represented by people working with final inspection of the assembled machine. The result showed a vast variation in judgement between the participants, both between the categories as well as within. The answers covered the whole range of judgement selections for most samples. The investigation indicates that the closer to the production the appraiser

is, the more demanding. The customer representatives as well as the designers were more satisfied with the quality than e.g. the welders.

4. Discussion

As the investigations indicated, the result of the visual judgements varies significantly in the current system. The investigations show a risk for both Type I and Type II errors which can lead to cost for break downs as well as unnecessary repairs. Safety critical industries are driving the development of non-destructive testing forward to decrease the risk of missing defects. However, the largest potential for other industries is probably within the indirect area, when the judgement is used for process development.

When people in different functions evaluate the quality differently it is clear that there will be a variation in the control of the process. The quality culture, which in many ways should be considered as positive, can easily contribute to large safety margins creating a more expensive process. The organization often reacts stronger on information suggesting that the quality is too low than the opposite. In certain industries where the consequences of a Type I error could be severe, this is of course a sound reaction. However in an industry with less safety issues, the cost perspective also needs to be considered. This results in unnecessary safety margins consisting of e.g. lower weld speed than possible which in turn creates longer lead time and increased cost.

As an example of the consequence of the scatter in the weld quality judgement in respect to weld cost two samples can be compared –a weld manufactured according to the current norm of the weld shop compared to the sample 9 from the judgement test previously described. Sample 9 deviates from the norm in its appearance. This norm is set by the people within the organization rather than weld demands. The sample is however accepted according to the weld class by all “Customers and Designers” in the test. The sample is welded with a 30% higher welding speed, giving a production cost only 76% of the current sample. A change in the welding process would also make it possible to produce 19% more in the current welding robot. For a company acting in a high competitive environment this is really good news.

Each company has to consider not only the consequences of type I and type II errors for its business but also what information is influencing its process control. Basing process development on subjective judgement containing the amount of variation showed in the investigations reported is foredoomed to failure. Being aware of the situation makes it however possible to influence the sources of variation as well as to reduce the effect of the variation. In such a process it is however important to reflect about the actual information need that should be extracted in order to take the right actions.

5. Conclusions

The subjective judgement of weld quality can influence the cost within a welding company in several aspects. Two investigations have shown examples of existing variation in judgement. The variation can have a severe effect on the production cost leading to type I or type II errors. A less obvious but nevertheless serious effect is that the variation in judgement might affect the manufacturing process e.g. through unused capacity and not executed process improvements.

Acknowledgement

The research work is partly funded by Vinnova (WIQ project) and the participating companies and universities. A great contribution has been made by the staff at Volvo CE Arvika and Braås.

References

- Ali A, Balint D, Temple J, Leever P Quantifying and improving the reliability of NDE through modeling manual ultrasonic inspections. In: Review of Quantitative Nondestructive Evaluation, 2010. vol 1211. American Institute of Physics, pp 1949-1956
- Bertovic M, Fahlbruch B, Müller C, Pitkänen J, Ronneteg U, Gaal M, Kanzler D, Ewert U, Schombach D (2012) Human Factors Approach to the Acquisition and Evaluation of NDT Data. Paper presented at the 18th World Conference on Nondestructive Testing, Durban, South Africa, 16-20 April 2012.
- Chase S, Ghasemi H (2008) Implications of the long term bridge performance program for life cycle costing in the United States. *Structure and Infrastructure Engineering* 5 (1):3-10. doi:10.1080/15732470701322719
- Dickens JR, Bray DE (1994) Human Performance Considerations in Nondestructive Testing. *Materials Evaluation* (September):1033-1041
- Ericson Öberg A, Johansson M, Holm EJ, Hammersberg P, Svensson L-E (2012) The Influence of Correct Transfer of Weld Information on Production Cost. Paper presented at the Swedish Production Symposium, Linköping Sweden, 6-8 November 2012
- Garvin DA (1987) Competing on the eight dimensions of quality. *Harvard Business Review* 65 (6):101-109
- Volvo Group (2011) STD 181-0004 Fusion welding - Weld classes and requirements life optimized welded structures steel, thickness ≥ 3 mm. Volvo Group,
- Swedish Standard (2004) SS-EN ISO 5817 Welding-fusion-welded joints in steel, quality levels for imperfections.
- Stenbacka N (2009) *Svetsekonomi och produktivitet*. 1 edn. Svetskommissionen, Stockholm, Sweden
- Öberg A, Hammersberg P, Svensson L-E Selection of Evaluation Methods for New Weld Demands: Pitfalls and Possible Solutions. In: 18th World Conference on Nondestructive Testing, Durban, 16-20 April 2012 2012.