Endüstri Mühendisliği / Journal of Industrial Engineering

https://dergipark.org.tr/tr/pub/endustrimuhendisligi

# A NEW ATTRIBUTE AGREEMENT ANALYSIS APPROACH FOR LONG-LENGTH PRODUCTS

İzzettin Hakan KARAÇİZMELİ[1*]

[1]Harran University, Engineering Faculty, Industrial Engineering Department, Şanlıurfa, ORCID No : https://orcid.org/0000-0002-8540-8050

**Keywords**

*Attribute agreement analysis, Qualitative data, Inspection, Quality control, Fabric quality*

**Abstract**

*To be able make right decisions in management, it is necessary to collect accurate data from processes. The precision and consistency of the measurement systems through which these data are collected affect the reliability of the data. The data divide into two categories: quantitative and qualitative. Quantitative measurements are for variables, and qualitative measurements are for attributes. Attributes are obtained in automatic or manual inspections. Manual inspections are performed by humans. Although the human eye is very sensitive, there may be problems with consistency due to many factors, ranging from psychology to environmental conditions. Attribute agreement analysis has been used for many years to understand the degree to which human inspection-based controls are consistent. The known method is mostly used in the analysis of measurement systems where piece-based pass/fail decisions are made. In the production sector, long-length products are also manufactured and cannot all be evaluated using a pass/fail decision. The present study proposes a new Attribute Agreement Analysis approach for long-length products. The proposed approach was implemented in the quality control unit of a textile fabric manufacturing business. In field tests, the agreement of the operators with each other and with the standard was found to be 45%. This finding implies that there is room for improvement.*

*Sorumlu yazar; e-posta : hkaracizmeli@harran.edu.tr

doi : https://doi.org/10.46465/endustrimuhendisligi.1685638

# UZUN METRAJLI ÜRÜNLER İÇİN YENİ BİR NİTELİKSEL TUTARLILIK ANALİZİ YAKLAŞIMI

| Anahtar Kelimeler | Öz |
|---|---|
| *Niteliksel tutarlılık analizi, Niteliksel veri, Gözlem, Kalite kontrol, Kumaş kalitesi* | *Yönetimde doğru kararlar vermek için süreçlerden alınan doğru verilere ihtiyaç vardır. Bu verilerin toplandığı ölçüm sistemlerinin hassasiyeti ve tutarlılığı verilerin güvenilirliğini etkiler. Veriler nicel ve nitel olarak iki gruba ayrılır. Nicel ölçümler değişkenler için, nitel ölçümler ise nitelikler için kullanılır. Nitelikler otomatik veya manuel gözlemlere dayalı olarak gerçekleştirilir. Manuel gözlemler insanlar tarafından yapılır. İnsan gözü çok hassastır ama psikolojiden çevresel şartlara kadar birçok faktör nedeniyle tutarlılık konusunda sorunlar yaşanabilir. İnsan gözlemine dayalı kontrollerin ne kadar tutarlı olduğunu anlayabilmek için Niteliksel Tutarlılık Analizi yıllardır kullanılmaktadır. Mevcut yöntem, daha çok bağımsız ve tekil ürünlerin geçti/geçmedi kararlarının verildiği ölçüm sistemlerinin analizinde kullanılmaktadır. Üretim sektöründe uzun metrajlı olup, tamamına geçti/kaldı kararı verilemeyen ürünler de üretilmektedir. Bu çalışmada uzun metrajlı ürünler için yeni bir Niteliksel Tutarlılık Analizi yaklaşımı önerilmektedir. Önerilen yaklaşım tekstil sektöründe kumaş üretimi yapan bir firmanın kalite kontrol biriminde uygulanmıştır. Saha uygulamasında operatörlerin birbirleriyle ve standartla olan uyumu %45 olarak belirlenmiştir. Bu bulgu, iyileştirmeye açık alan olduğunu göstermektedir.* |


| | | | |
|---|---|---|---|
| Araştırma Makalesi | | Research Article | |
| Başvuru Tarihi | : 28.04.2025 | Submission Date | : 28.04.2025 |
| Kabul Tarihi | : 02.10.2025 | Accepted Date | : 02.10.2025 |


## 1. Introduction

Managers need data to make the right decisions. Data can also be collected by measuring various processes. The science of measurement is called metrology. Czichos, Saito, and Smith (2011) divide metrology into three branches: scientific, industrial, and legal metrology. Industrial metrology ensures adequate operation of measurement instruments in production processes and testing. Good measurements increase product value, productivity, and quality.

To be able to make the right decisions in a quality check process, the collected data must be reliable. To ensure this, it is necessary to evaluate the measurement system from which the data are obtained. In an enterprise, the measurement system is a critical component that deeply affects decision making processes. Unfortunately, an ineffective measurement system can dramatically affect business performance (Montgomery, 2020). If the variance of a measurement system is low, that measurement system is considered acceptable, while it is considered unacceptable or marginal if it is high (de Oliveira et al., 2020).

Burdick, Borror, and Montgomery (2003) mentioned several objectives of the measurement system analysis. The first objective is to find out the level of variability caused by the measurement system. The second is to eliminate possible sources of variability. The third is to decide whether the measurement system is satisfactory. While it is commonly known that there is some variability in each measurement of a product or unit, due to variation in its production processes, some variability is caused by the measurement system itself. Some of this system-related variation may be due to the measurement instrument or device, while others may be due to the operator, calibration, or other possible conditions that vary over time (Montgomery, 2020).

Measurements made in enterprises can be both qualitative and quantitative. Qualitative measurements obtain attribute data, while quantitative measurements obtain variable data. Qualitative measurements are often based on human inspections. As Pendrill and Petersson (2016) note, these qualitative measurements based on human inspection are encountered not only in the service sector, but also in all sectors where all kinds of production process take place. Although numerical measurements are considered ideal, it is a fact of life that qualitative pass-fail decisions are part of the measurements in enterprises, despite advanced technology (Boyles, 2001). In other words, it may not always be possible to digitise measurements.

It should be noted that not only qualitative, but also many quantitative decisions are directly or indirectly influenced by the human factor. Qualitative inspections are often fully human-made measurements, while some quantitative inspections are made and reported fully automatically without human intervention. In some quantitative inspections made using devices or instruments, it is a human being who makes or records the measurement or does both. In fact, no device or robot is as perfect as humans, who can make many inspections and measurements in the best way. However, while it is possible to calibrate a robot or device on the basis of a level of tolerance, this is slightly different in human-made measurements. This situation leads to human-induced measurement variations. Among the main causes of these variations are perceptual differences among operators (Montgomery, 2020), disparities in training and experience levels (Automotive Industry Action Group-AIAG, 2010), fatigue and loss of attention resulting from prolonged working conditions (Diering and Kujawińska, 2020), differences in interpreting measurement criteria (Lashkari and Chenouri, 2025), as well as communication deficiencies or inadequacies in work instructions. These factors, especially in systems involving qualitative assessments, can negatively affect the consistency and accuracy of measurement results, potentially leading to errors in decision-making processes. Therefore, systematically analyzing the human factor and implementing appropriate corrective measures are critically important for enhancing the reliability of measurement systems.

### 1.1. Measurement System Analysis

Just like measurements, measurement system analysis can be divided into quantitative and qualitative categories. In quantitative analysis, the most frequently used analysis is Gage R&R (Repeatability & Reproducibility). In this analysis, the variations caused by the measurement itself and the measurement system are pinpointed (Zanobini, Sereni, Catelani, and Ciani, 2016). On the other hand, Attribute R&R analysis is reserved for qualitative analysis. This analysis is also called Attribute Agreement Analysis (AAA). There are two critical components in the measurement system. Repeatability refers to the capability of obtaining the same observed value under the same conditions. Reproducibility indicates the capability of obtaining the same observed value under different conditions, by different operators, or at different times (Montgomery, 2020). In fact, these qualifications indicate how sensitive the measurement system is. De Mast and van Wieringen (2004) defined the precision of a measurement system as the consistency observed in repeated measurements of the same object.

AIAG (2010) described AAA with reference to measurement systems analysis. The Minitab software features a menu for the AAA application, which is frequently used in industry. There are studies carried out in industrial organisations both for condition assessment and quality improvement. However, the number of AAA studies in the literature is rather limited compared to that of Gage R&R. Studies have generally focused on measurement systems in which an independent product or sample is checked, and a decision is made as to whether it passes or fails the quality test.

### 1.2. Related Works

For example, Farías-Rodríguez, Monge-Moreno, Rosales-Vargas, and Guzmán-Arroyo (2015) used AAA to inspect the products in a foundry factory. Marques, Lopes, Santos, Delgado, and Delgado (2018) used this analysis to help operators classify circuits as good/bad in an electronic circuit manufacturing process. In this study, operators were asked to distinguish possible faulty circuits by looking at the circuit images. The results revealed that operators should receive additional training. Simion (2018) used AAA to perform visual inspection in a chain manufacturing company. He noted that these systems are the most problematic measurement systems, since subjective decisions are made, particularly during visual inspections by operators.

In the measurement phase of 6 Sigma studies, it is essential to ensure the consistency of the measurement systems. Therefore, when measurement data is needed, measurement systems are analysed first. Aust and Pons (2022) carried out AAA in a process in which the blades of the aircraft engines were controlled. The purpose of the study was to test the degree to which the controllers were consistent. If the operator evaluates a perfectly functioning wing as if it were defective, the maintenance cost increases. On the other hand, if you call a defective wing reliable, the risk of an accident increases. Yadav, Mathiyazhagan, and Kumar (2019) applied the 6 Sigma method in the automobile windscreen manufacturing process. They also performed AAA to determine the agreement of the operators. For this study, 20 parts were selected. Of these 20 parts, five were good, five were defective, five were marginal good, and five were marginal defective. The results indicated that the agreement of the operators was sufficient. Similarly, Simion (2019) used AAA for visual inspection in the production process of exterior lighting projectors in an automotive sub-industry business.

Unnikrishnan, Donovan, Macpherson, and Tormey (2019) carried out a study on machine learning in the pharmaceutical industry. In this study, micrographs of samples taken from the production of emulsion cream prepared for drug production were produced using a microscope. The micrographs were classified as "target, acceptable, marginal and unacceptable" both by the machine learning method and by analysts. Globally considered, the study indicated that the results obtained using machine learning were more consistent than those of AAA when agreement with reference was considered.

Qualitative measurements can also be used by digitising them. Lyu and Chen (2008a) noted that there are few studies on attribute data in the literature. They proposed a generalised linear model for systems that include attribute data. In their study, the data was quantified by calculating the mean and variance for the attribute data collected by inspection. In another study, Lyu and Chen (2008b) proposed a model based on the calculation of variance using the Poisson distribution for bivariate attribute data. Performing a different version of their study, the authors introduced an exponential distribution into their model (Lyu and Chen, 2010). It may be useful to design functional interfaces so that such models can be used comfortably by businesses.

Qualitative agreement analysis, on the contrary, is also used when numerical measurements are classified into discrete groups. For example, de Almeida et al. (2021) used AAA to investigate the consistency of the eight different connection methods used in their study to analyse the transformers they identified in Brazil for voltage sags. This technique was found to be very useful in reducing subjectivity in controls.

AAA is a useful tool to check the agreement among evaluators. It can be used in many different areas for this purpose. For example, Murphy, Moeller, Page, Cerqua, and Boarman (2009) conducted an agreement analysis of the responses given to questions received by email at Ohio State University. The results indicated that this analysis helped to develop a common language among employees and improve the quality of the service provided. Mejdell et al. (2010) conducted an agreement analysis with 43 agricultural students using a scoring system intended for equine wounds. They categorised the possible wounds in horses into 5 groups and showed the relevant photographs to the students. Then they gave the students a test with 40 photographs on a CD-ROM. The rating of an experienced veterinarian was accepted as the standard. The students then scored the photographs twice at 10-day intervals. In another study, AAA was used to determine soil colours (Marques-Mateu, Moreno-Ramon, Balasch, and Ibanez-Asensio, 2018). Petrović et al. (2024) employed AAA to assess the consistency of visual inspection for the presence of parasites in frozen fish. Hryb, Pawlaczyk, Zhou, and Majchrzak (2025) used AAA to evaluate color consistency in decorative papers utilized in furniture applications. In their study, they proposed a method involving the combination of the reference sample with the part to be inspected, which they referred to as a multi-variant reference standard. In another study, Xiang, Yuan, Zhang, and Zhang (2025) applied AAA in the context of smart agriculture, specifically for labeling biological images.

Another study addressed the issue of consistency from a very different perspective. Xenarios, Kakumanu, Nagothu, and Kotapati (2017) studied the gender-different effects of weather extremes in rural South India. These researchers also used AAA in their study. They developed a questionnaire for the participants from different villages in the study. Participants responded to the survey questions with binary responses. Using the collected data, they calculated village-based and gender-based agreement rates. Their purpose was to understand people's perceptions and determine what can be done to eliminate gender inequality. In another study, conducted by Cotter and Yesilbas (2018), AAA was used for the Human Factors Analysis and Classification System (USA, Department of Defence) among three instructor pilots, who had experience analysing the impact of human factors on air force accidents.

### 1.3. Contribution of The Study

In earlier studies in the literature, AAA studies conducted in the manufacturing or service sector, researchers have always used piece-based samples (products) that can be considered discrete. However, some industrial products have hardly a discrete structure. A good example of this could be textile fabric production. In fabric manufacturing, depending on customer orders, a product may be 50 meters, 1,000 meters, or even 10,000 meters in length. Although they may vary in size relative to each other, all are considered long-length products compared to piece-based products, and each production run exhibits continuity within itself. In the present study, the AAA approach is proposed for production systems where qualitative data is obtained by human inspections and products are continuous rather than discrete. The proposed AAA approach was tested in the quality control process in a fabric manufacturing business. The results revealed that there were differences between the measurements made by the operators who participated in the study. Some suggestions were made to eliminate these differences. Thus, a new AAA approach was created and tested in the field that can be used in similar production systems. In fact, the proposed AAA aims to radically improve quality control processes in such production systems.

With this aim in mind, studies on measurement system analysis in the textile sector were also reviewed. In the textile industry, studies are carried out to investigate the level of consistency or bias, particularly in measurements using instruments. For example, Clonts et al. (2006) analysed the measurement of the colour factor by using spectrophotometers. Their analysis revealed that the variance was relatively high in repeated measurements. Kulsum, Islam, Sayed, Shahin, and Islam (2024); Sybilska, Walawska, and Matusiak (2017); Wang, Li, and Wang (2021), also compared the Data Color spectrophotometer with the Digi Eye colourimeter to evaluate the colour factor. These are measurement system analyses for quantitative data. Research focusing on the analysis of measurement systems based on qualitative data remains scarce. AAA was used in a study on touching, which is the feeling a human gets when he/she touches a fabric. As touching is subjective to some extent, depending on the people touching the fabric, interpersonal agreement may be a problem. Sülar and Okur (2007) investigated the consistency of repeated measurements of fabric samples they prepared in their research on this subject.

The rest of the paper is as follows. The next section introduces the new AAA approach and elaborates on the design of the implementation study. The third section presents the results of the analyses. The last section draws conclusions and makes suggestions for prospective studies.

## 2. The New "AAA" Approach

Various decisions are constantly made in companies. In quality control units, authorised employees decide whether the quality of the product is satisfactory or not. Inspections are essential for making such decisions. These inspections are used for data collection. As seen in Figure 1, these data are divided into two groups, variable or attribute. Variable data can be obtained automatically using measurement devices or manually by operators using these devices. On the other hand, attribute data are the results of inspections and decisions made by operators. In fact, systems have also been developed in which attribute data can be automatically obtained. The development of sensor and camera technologies, in conjunction with advancements in artificial intelligence and machine learning, has enabled the automation of manual inspections. However, since the suggested systems are not as capable as the human eye and are still developing, additional time is needed for them to replace humans. However, they can be used as a tool for preliminary inspection before operator control begins.

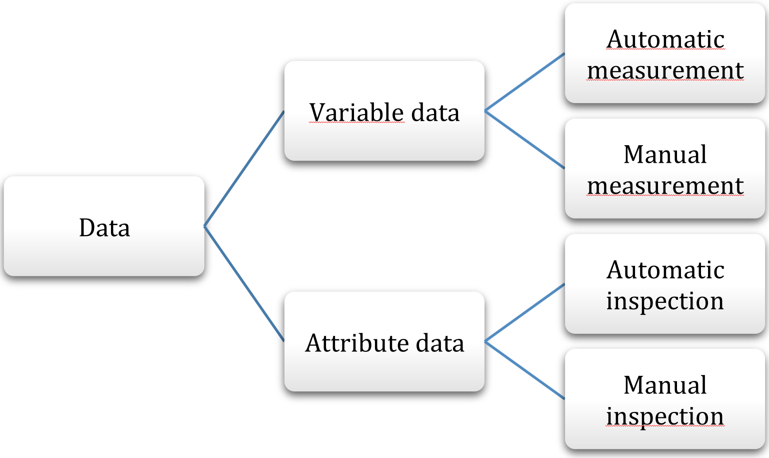


Figure 1. Data and Measurement Types

During the quality control stage, all or some of the products are checked. In the textile industry, during quality control processes, the type and quantity of defects in the fabric are determined and recorded in the system. If there is a defect above the tolerance level previously negotiated with the customer, the defective part is cut off. In such long-length products, it is highly unlikely that a whole product to be entirely faulty or faultless. It is critical to determine which part of the product is defective.

There are two important decisions in quality control; the first is to decide whether there is a defect in the product or not, and the second is to identify the type of defect. In some manufacturing systems, there may be too many types of defects. Textile production is a clear example of this, as many types of production defects are seen in textile products. Therefore, quality control operators need to decide correctly both whether there is a defect and, if there is a defect, what type of defect it is. For this reason, operators should have advanced inspection skills. Moreover, it is essential that they receive good training in their work and gain enough experience with the types of defects. Since the pattern and colour of fabrics also vary a lot, the appearance of the same defect may be different in different fabrics. Figure 2 shows a fabric inspection machine used in textile companies. Fabrics are passed over such machines, and necessary checks are made. During this process, the attention and experience of the operator and the environmental conditions are very important.

Figure 2. A Fabric Inspection Machine

In this paper, we propose an AAA approach for long-length or continuous products that cannot be controlled individually. To develop the approach, a small AAA experiment was first conducted (Karaçizmeli, 2023). In this experiment, a product with 10 defects was checked twice by two different operators. The results indicated that the agreement with the standard was 60%. This result suggested that a different AAA approach is needed. However, in this small-scale experiment, only faulty region decisions were checked. However, when checking products, consistent decisions should be made not only for faulty regions but also for those without faults. Therefore, it is essential to follow a different approach in the selection of products. The extent to which operators do their work properly can be determined more accurately if there are defective regions on the product, as well as regions that are within tolerances and should be accepted as defect-free.

As shown in Figure 3, the first and most important step in the proposed approach is the selection of products. This step also highlights the key distinction between the proposed method and the known AAA method. In current AAA applications, piece-based samples are not suitable for long-length production processes. In the proposed approach, a product that reflects the continuous nature of long-length manufacturing is selected. Each decision region on the product is treated as an individual sample. In this way, multiple samples—each potentially containing different types of defects—can be evaluated using a single product.

Another important consideration in product selection is the appropriate length of the product to be used. This decision should be made by taking into account both the cost of analysis and the adequacy of the results. If the product is too long, it increases the cost of fabric used as well as labor and machine costs, since it extends the inspection time for operators. On the other hand, if the product is too short, the analysis may not be sufficiently reliable. Therefore, a fabric length that includes representative defects from production and is determined by specialist judgment should be selected.

It is decided that the results will be more effective if the product is a product of average difficulty in terms of decision making. This is because the risk of defects increases in more difficult products, whereas the opposite is true for easier products. There should be some defective areas, some of which are different types of defects, in the sample product. There should also be decision regions in the product that should be accepted as defect-free. It is stated in the Minitab software that the higher the number of samples to be checked, the more accurate the results get (Minitab, 2023). For the proposed approach, the number of samples to be checked is the number of regions for each of which a decision is made.

Later, in the second step, a specialist checks the product and keeps a record of the defective regions, along with the types of defects. Thus, the standard is established. The next step is the first inspection stage. The fact that the operators are not fully aware of the research objectives ensures that the inspection procedure provided reliable results. The operators who participated in the study are given the same product at different times independently of each other, and they are asked to check it, identify defects, and record their regions. All operators are asked to perform the same process in turn. After a certain period, such as one week, has passed since the first inspection, the second inspection stage begins. Leaving an intervening time between the two inspections is to ensure that operators forget the fabric. The same fabric is checked again by the same operators. It is essential that this process be carried out unannounced, as in the first inspection.

The results obtained in the fifth step are then analysed by the process manager. These analyses can be carried out in many ways. The agreement between the standard and the operators, the agreement of the operators with each other (inter-operator agreement), and the agreement of the operators with themselves (intra-operator agreement) are analysed.

Cohen's Kappa statistic is used for statistical analysis. Analyses help identify the needs of operators. In the last step (the improvement step), the necessary actions are planned to meet these needs.

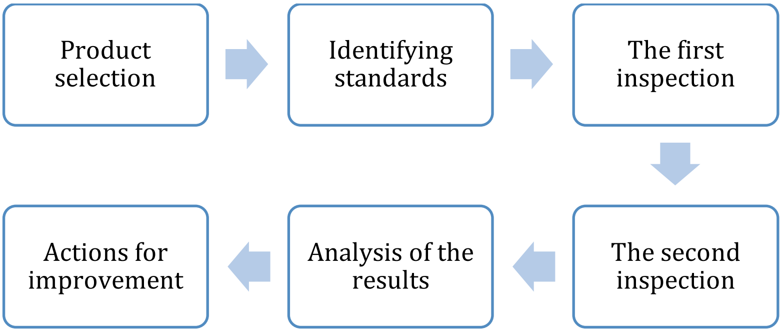


Figure 3. Steps of Implementation

$$Level\ of\ Agreement\ (\%) = \frac{Number\ of\ Matched\ Regions}{Number\ of\ Inspected\ Regions} \times 100 \quad (1)$$

The original version of the Kappa statistic, introduced by Cohen in 1960 and its modified versions have been used for many years in medical statistics, quality engineering and industrial statistics (de Mast and van Wieringen, 2007). The Kappa statistic shows a quantitative measure of the agreement between observers beyond what might be expected by chance (Vetter and Schober, 2018). It can be used for only two independent observers (Cohen, 1960; Viera and Garrett, 2005; McHugh, 2012), but Fleiss Kappa can be used for three or more people (Fleiss, 1971; Vetter and Schober, 2018). The Kappa statistic can vary between -1 and +1. -1 shows complete incompatibility, 0 means that chance might be involved, and +1 means that decisions are perfectly compatible (Vetter and Schober, 2018). Kappa values greater than 0.70 are considered adequate, and those greater than 0.90 indicate perfect agreement (Marques et al., 2018). Kappa statistics are calculated based on Equation (2). In this equation, "$P_o$" represents the proportion of observed agreement among operators, while "$P_e$" denotes the proportion of agreement expected by chance. In this study, Minitab was used to perform the calculations.

$$Kappa = \frac{P_o - P_e}{1 - P_e} \quad (2)$$

## 2.1. Implementation Design

For the test study, a fabric with a medium difficulty level and 20 regions to be identified was selected. The length of the fabric was 50 metres. Firstly, the department specialist, who was accepted as a point reference, determined the standard by making the necessary checks. According to the established standard, there was no defect in seven of the 20 regions, while there was a defect in 13 of them. Then, the operators were asked to check the fabric and prepare a defect map of it. An operator from each of the three shifts participated in the study at different times. They were not informed about the purpose of the study to prevent them from communicating with each other. One week later, the same fabric was inspected again by the same operators. Thus, in addition to a reference measurement, three operators obtained two repeated measurements. The collected data were analysed using Minitab.

This study complied with research and publication ethics.

## 3. Results and Discussion

Table 1 presents the inspection results in the AAA application. In the table, "Ok" means there is a defect and "Nok" means that there is no defect. O1C1 refers to the first check made by the first operator, and O1C2 refers to the second by the same operator.

Table 1

Inspection Results

| Part | Standard | O1C1 | O1C2 | O2C1 | O2C2 | O3C1 | O3C2 |
|---|---|---|---|---|---|---|---|
| 1 | Ok | Nok | Ok | Ok | Ok | Ok | Ok |
| 2 | Ok | Ok | Ok | Ok | Ok | Nok | Ok |
| 3 | Ok | Ok | Ok | Ok | Ok | Ok | Ok |
| 4 | Nok | Nok | Nok | Nok | Nok | Nok | Nok |
| 5 | Ok | Ok | Ok | Ok | Ok | Ok | Nok |
| 6 | Ok | Ok | Ok | Ok | Ok | Ok | Ok |
| 7 | Nok | Ok | Ok | Nok | Nok | Ok | Nok |
| 8 | Nok | Nok | Nok | Nok | Nok | Nok | Nok |
| 9 | Ok | Ok | Nok | Nok | Nok | Ok | Ok |
| 10 | Ok | Ok | Ok | Ok | Ok | Nok | Nok |
| 11 | Ok | Ok | Ok | Ok | Ok | Ok | Ok |
| 12 | Ok | Ok | Ok | Ok | Ok | Ok | Ok |
| 13 | Ok | Nok | Ok | Ok | Nok | Ok | Ok |
| 14 | Nok | Nok | Nok | Nok | Nok | Ok | Nok |
| 15 | Ok | Ok | Ok | Ok | Ok | Ok | Ok |
| 16 | Nok | Nok | Nok | Nok | Nok | Nok | Nok |
| 17 | Ok | Nok | Nok | Ok | Ok | Ok | Ok |
| 18 | Nok | Nok | Nok | Nok | Nok | Nok | Nok |
| 19 | Nok | Nok | Nok | Nok | Nok | Ok | Ok |
| 20 | Ok | Ok | Ok | Ok | Ok | Ok | Nok |

Data were statistically analysed using Minitab software. Table 2 shows the level of agreement between the operators themselves. While the highest agreement (95%) is observed in the second operator, the lowest agreement (75%) is observed in the third operator. This measurement shows the agreement between the two measurements made by an individual operator. Table 3 provides the agreement of the measurements of the individual operators with the standard. These inter-operator agreement values were slightly lower than the intra-operator values. Again, while the highest agreement was observed in the measurements of the second operator, the lowest agreement was observed in those of the third operator. Figure 4 provides a graphical representation of the agreements of the operators within themselves and with the standard.

Table 2

Agreement within Operators

| Operator | #Inspected | #Matched | Percent | 95% CI |
|---|---|---|---|---|
| 1 | 20 | 17 | 85,00 | (62,11; 96,79) |
| 2 | 20 | 19 | 95,00 | (75,13; 99,87) |
| 3 | 20 | 15 | 75,00 | (50,90; 91,34) |

Table 3

Agreement Operator vs Standard

| Operator | #Inspected | #Matched | Percent | 95% CI |
|---|---|---|---|---|
| 1 | 20 | 15 | 75,00 | (50,90; 91,34) |
| 2 | 20 | 18 | 90,00 | (68,30; 98,77) |
| 3 | 20 | 13 | 65,00 | (40,78; 84,61) |

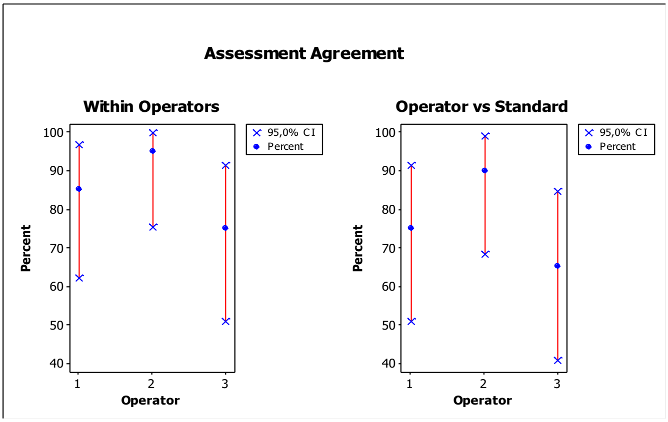


Figure 4 The Levels of Agreement

Table 4 shows the agreement of the operators with each other and with the standard. The same decisions were made in only nine of 20 regions. This is a relatively low rate (45%). The Fleiss Kappa statistic was calculated to be 0.66. Kappa statistics below 0.70 indicate that there is room for improvement in the system (Marques et al., 2018). Additionally, according to the Z-test results, the p-value was calculated as 0.000, indicating that the Kappa value is statistically significantly different from zero.

A more detailed analysis of the results reveals that the second operator is the most senior operator, while the third operator is a new employee. Given the different levels of experience, employees, particularly those with fewer experiences, need further training. Similarly, as calibration is performed to ensure the consistency of the measuring devices, sensitivity should be increased through prospective visual inspection studies.

Table 4

The Overall Agreement

| #Inspected | #Matched | Percent | 95% CI |
|---|---|---|---|
| 20 | 9 | 45,00 | (23,06; 68,47) |

The more consistent the measurement system, the more accurate the decisions about the products. A consistent measurement system not only ensures accurate decisions about the quality of products but also helps identify the sources of variations in products and processes. The earliest version of quality control was based on inspection and removal of defective products. However, today, the focus is on preventing defects and ensuring that defective products are not manufactured. Borkowski and Knop (2016) noted that the value of modern quality control should be measured by the extent to which products are free of defects rather than the number of defects identified.

In quality control studies, it is more objective to evaluate the quantitative data collected by devices, while it is more subjective to evaluate the qualitative inspections performed by humans. Therefore, visual inspection has been a challenging process in all manufacturing industries (Unnikrishnan et al., 2019), as it usually involves some subjectivity. The results may vary according to the nature of subjective evaluations. If there is a bias towards false-positive errors, the defect rate is inadvertently increased, and the waste/reprocessing cost of the business, also known as producer risk, goes up. If there is a bias towards false-negative errors, the defective product is sent to the customer, which is referred to as consumer risk. Therefore, it is necessary to calibrate the inspecting eyes as frequently as possible, just as the calibration of the measurement devices.

Otherwise, operating costs may increase, which may lead to customer dissatisfaction. However, companies with quality assurance systems, such as ISO 9001, periodically exercise controls and perform calibration procedures for their devices, while they usually fail to do these for visual inspections made by humans.

## 4. Conclusions

The prerequisite for making the right decision about a situation in any walk of life is to assess that situation accurately. More specifically, it is necessary to fully understand the situation in a business to make the right decisions. Having enough data is also essential during this process. The more perfect the system from which these data are obtained, the more accurate and consistent the data will be. Process variation should be reduced to achieve the goal of zero defect, which is frequently expressed in improvement studies in both the production and service sectors today. Bruno et al. (2016) emphasised that the way to reduce process variation is to collect accurate process data.

The quality control process is one of the most critical links in the production chain of an enterprise. If quality control works correctly, the costs of the enterprise do not increase unnecessarily. Wrong decisions made during quality control increase the levels of producer or consumer risks. Although the human eye is very sensitive, there may be problems with consistency due to many factors, ranging from psychology to environmental conditions. In this study, a new AAA approach is proposed to eliminate errors in quality control activities procedures. Thus, it will be possible to reduce costs by minimising risks.

The measurement system from which the data are obtained can rely on quantitative measurements or qualitative decisions. The precision of the instruments used in quantitative measurements is increased by calibration, while the precision of the operators making qualitative decisions can only be increased through training. To understand the current state of the operators, the AAA appears to be a very appropriate tool. Although AAA is not a new application, only piece-based sample (product) checks are performed in known AAA applications. The AAA approach proposed in this study aims to improve the quality control inspections of long-length products. Factors such as employee turnover, training costs, and managers placing less trust in qualitative data compared to quantitative data may have limited research in this area. However, especially in industries like the textile sector, where decision-making relies on observation-based qualitative data, the use of AAA can be highly beneficial for determining and improving agreement.

In the present study, a new AAA approach was developed for the quality control processes of long-length products, which can hardly be checked individually and for which a pass/fail decision can hardly be made. The proposed AAA approach was tested in the quality control processes of a textile company. First, a standard was established for the selected fabric, and three selected operators checked the same fabric twice. During these checks, the operators were not aware of the purpose of their work; none of them knew what the others were doing and did not know that they would perform the check again. The results indicated that although the agreement between the operators was at least 75%, its agreement with the standard and with each other was found to be 45%. This implies that the system needs improvement. In the application phase of the proposed approach, a product with 20 decision regions and 3 operators was used. This demonstrated the feasibility of the proposed method. However, when implemented in an industrial setting, it should be periodically applied to all relevant operators according to a specific schedule. Different products with varying levels of difficulty should be selected for the applications, and the chosen products must contain a sufficient number of decision regions to effectively capture the distinguishability of defects occurring in production.

The AAA approach proposed in the present study can be used in appropriate production systems in prospective studies. It can be used for preliminary preparation before creating training programmes for employees who are decision makers in measurement systems. Training programmes can be designed according to deficiencies that the system in a business suffers from. In addition, more AAA studies can be carried out periodically to keep the teams and their knowledge as fresh as possible.

## Author Contributions

All stages of the study were carried out by İzzettin Hakan Karaçizmeli.

## Conflict of Interest

There is no conflict of interest in this study.

## References

Aust, J., & Pons, D. (2022). Assessment of aircraft engine blade inspection performance using attribute agreement analysis. *Safety,* 8(23), 1-24. Doi: https://doi.org/10.3390/safety8020023

Automotive Industry Action Group (AIAG). (2010). *QS 9000: Measurement systems analysis reference manual*. 4th ed. Michigan: Chrysler Group LLC, Ford Motor Company, General Motors Corporation.

Borkowski, S., & Knop, K. (2016). Challenges faced in modern quality inspection. *Management and Production Engineering Review,* 7(3), 11-22. Doi: https://doi.org/10.1515/mper-2016-0022

Boyles, R. A. (2001). Gauge capability for pass-fail inspection. *Technometrics,* 43(2), 223-229. Doi: https://doi.org/10.1198/004017001750386332

Bruno, R., Pereira, D., Peruchi, R. S., de Paiva, A. P., da Costa, S. C., & Ferreira, J. R. (2016). Combining Scott-Knott and GR&R methods to identify special causes of variation. *Measurement,* 82, 135-144. Doi: https://doi.org/10.1016/j.measurement.2015.12.033

Burdick, R. K., Borror, C. M., & Montgomery, D. C. (2003). A review of methods for measurement systems capability analysis. *Journal of Quality Technology,* 35(4), 342-354. Doi: https://doi.org/10.1080/00224065.2003.11980232

Clonts, R., Thangavelu, R., Hinks, D., Dunn, J., Guzman, P., & Laidlaw, A. (2006). Inter-Instrument agreement in the colorimetric measurement of textile materials. *AATCC Review,* 6(8), 45-48.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ Psychol Meas.,* 20, 37–46. Doi: https://doi.org/10.1177/001316446002000104

Cotter, T. S., & Yesilbas, V. (2018). *An attribute agreement analysis method for HFACS inter-rater reliability assessment.* In Proceedings of the American Society for Engineering Management 2018 International Annual Conference, ed. E-H. Ng, B. Nepal, E. Schott, and H. Keathley, 1-13, Alabama/USA.

Czichos, H., Saito, T., & Smith, L. E. (2011). *Springer Handbook of Metrology and Testing*. Würzburg: Springer Science & Business Media.

de Almeida, F. A., de Mello, L. G., Romao, E. L., Gomes, G. F., de Freitas Gomes, J. H., de Paiva, A. P., … Balestrassi, P. P. (2021). A PCA-based consistency and sensitivity approach for assessing linkage methods in voltage sag studies. *IEEE Access,* 9, 84871-84885. Doi: https://doi.org/10.1109/ACCESS.2021.3088436

de Mast, J., & van Wieringen, W. N. (2004). Measurement system analysis for bounded ordinal data. *Quality and Reliability Engineering International,* 20, 383-395. Doi: https://doi.org/10.1002/qre.653

de Mast, J., & van Wieringen, W. N. (2007). Measurement system analysis for categorical measurements: agreement and Kappa-Type indices. *Journal of Quality Technology,* 39(3), 191-202. Doi: https://doi.org/10.1080/00224065.2007.11917688

de Oliveira, L. G., Aquila, G., Balestrassi, P. P., de Paiva, A. P., de Queiroz, A. R., de Oliveira Pamplona, E., & Camatta, U. P. (2020). Evaluating economic feasibility and maximization of social welfare of photovoltaic projects developed for the Brazilian northeastern coast: An Attribute Agreement Analysis. *Renewable and Sustainable Energy Reviews,* 123, 1-15. Doi: https://doi.org/10.1016/j.rser.2020.109786

Diering, M., & Kujawińska, A. (2020). Human Aspects of the Measurement System Analysis. *AHFE International Conference, AHFE Open Access*, (12), USA. Doi: https://doi.org/10.54941/ahfe100447

Farías-Rodríguez, R., Monge-Moreno, M. J., Rosales-Vargas, R., & Guzmán-Arroyo, C. (2015). *Assessment of the capability of a foundry shop process measurement system through an attribute agreement analysis.* In IIE Annual Conference Proceedings, 3019-3026, Norcross, Georgia/USA.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychol Bull,* 76, 378–382. Doi: https://doi.org/10.1037/h0031619

Hryb, M., Pawlaczyk, M., Zhou, J., & Majchrzak, J. (2025). Attribute agreement analysis with the multi-variant reference standard. *Management and Production Engineering Review*, 15(1), 1-9. Doi: https://doi.org/10.24425/mper.2025.153936

Karaçizmeli, İ. H. (2023). *Quality control in textile: an attribute agreement analysis*. 13th Research/Expert Conference with International Participation Quality 2023 Proceedings, 71-74, Neum, Bosnia and Herzegovina.

Kulsum, U., Islam, T., Sayed, A., Shahin, M., & Islam, M. N. (2024). Analyzing the differences between spectrophotometer and DigiEye assessments of knit fabric. *International Journal of Textile Science*, 13(2), 11-21. Doi: https://doi.org/10.5923/j.textile.20241302.01

Lashkari, R., & Chenouri, S. (2025). (preprint). A comprehensive framework for statistical inference in measurement system assessment studies. Doi: https://doi.org/10.48550/arXiv.2501.18037

Lyu, J., & Chen, M. N. (2008a). Gauge capability studies for attribute data. *Quality and Reliability Engineering International,* 24(1), 71-82. Doi: https://doi.org/10.1002/qre.868

Lyu, J., & Chen, M. N. (2008b). *A bivariate attribute measurement model for Six Sigma project.* 2008 4th IEEE International Conference on Management of Innovation and Technology, 1129-1134, Bangkok/Thailand. Doi: https://doi.org/10.1109/ICMIT.2008.4654528

Lyu, J., & Chen, M. N. (2010). Evaluating the precision of bivariate attribute measurements. *Journal of Statistical Computation and Simulation,* 80(1), 99-110. Doi: https://doi.org/10.1080/00949650802508891

Marques, C., Lopes, N., Santos, G., Delgado, I., & Delgado, P. (2018). Improving operator evaluation skills for defect classification using training strategy supported by attribute agreement analysis. *Measurement,* 119, 129-141. Doi: https://doi.org/10.1016/j.measurement.2018.01.034

Marques-Mateu, A., Moreno-Ramon, H., Balasch, S., & Ibanez-Asensio, S. (2018). Quantifying the uncertainty of soil colour measurements with Munsell charts using a modified attribute agreement analysis. *CATENA,* 171, 44-53. Doi: https://doi.org/10.1016/j.catena.2018.06.027

McHugh, M. L. (2012). Interrater reliability: the Kappa statistic. *Biochem Med*., 22, 276–282.

Mejdell, C. M., Jorgensen, G. H., Rehn, T., Fremstad, K., Keeling, L., & Boe, K. E. (2010). Reliability of an injury scoring system for horses. *Acta Veterinaria Scandinavica,* 52:68. Doi: https://doi.org/10.1186/1751-0147-52-68

Minitab. Attribute Agreement Analysis. Accessed September 01, 2023. https://support.minitab.com/en-us/minitab/21/help-and-how-to/quality-and-process-improvement/.

Montgomery, D. C. (2020). *Introduction to Statistical Quality Control*. 8th ed. New York: John Wiley & Sons.

Murphy, S. A., Moeller, S. E., Page, J. R., Cerqua, J. & Boarman, M. (2009). Leveraging Measurement System Analysis (MSA) to improve library assessment: the attribute gage R&R. *College & Research Libraries,* 70(6), 568-577. Doi: https://doi.org/10.5860/0700568

Pendrill, L., & Petersson, N. (2016). Metrology of human-based and other qualitative measurements. *Measurement Science and Technology,* 27(9), 1-11. Doi: https://doi.org/10.1088/0957-0233/27/9/094003

Petrović, Z., Ćirić, J., Babić-Milijašević, J., Milijašević, M., Lukić, M., Jovanović, J. & Nikolić, A. (2024). Use of attribute agreement analysis (AAA) in the validation of sensory evaluation methods: Case study for the visual determination of parasites in fish. *Meat Technology*, 65(1), 61-69. Doi: https://doi.org/10.18485/meattech.2024.65.1.7

Simion, C. (2018). *Assessment of human capability, an effective tool to obtain confidence in the visual inspection process*. Acta Universitatis Cibiniensis – Technical Series. LXX. Doi: https://doi.org/10.2478/aucts-2018-0001

Simion, C. (2019). Measurement system analysis by attribute, an effective tool to ensure the quality of the visual inspection process within an organization. *MATEC Web of Conferences* 290, 05004. Doi: https://doi.org/10.1051/matecconf/201929005004

Sülar, V., & Okur, A. (2007). Sensory evaluation methods for tactile properties of fabrics. *Journal of Sensory Studies,* 22, 1-16. Doi: https://doi.org/10.1111/j.1745-459X.2007.00090.x

Sybilska, W., Walawska, A., & Matusiak, M. (2017). Comparison of spectrophotometric and DigiEye colour measurements of woven fabrics. *Textile and Apparel,* 27(1), 53-59.

Unnikrishnan, S., Donovan, J., Macpherson, R., & Tormey, D. (2019). Machine learning for automated quality evaluation in pharmaceutical manufacturing of emulsions. *Journal of Pharmaceutical Innovation,* 15, 1-12. Doi: https://doi.org/10.1007/s12247-019-09390-8

Vetter, T. R., & Schober, P. (2018). Agreement analysis: what he said, she said versus you said. *Anesthesia & Analgesia,* 126(6), 2123-2128. Doi: https://doi.org/10.1213/ANE.0000000000002924

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the Kappa statistic. *Fam Med,* 37, 360–363.

Wang, Y., Li, W., & Wang, J. (2021). Color prediction model for hybrid multifilament fabric. *Textile Research Journal*, 92(7-8), 1038-1048. Doi: https://doi.org/10.1177/00405175211047114

Xenarios, S., Kakumanu, K. R., Nagothu, U. S., & Kotapati, G. R. (2017). Gender differentiated impacts from weather extremes: Insight from rural communities in South India. *Environmental Development,* 24, 156-169. Doi: https://doi.org/10.1016/j.envdev.2017.05.002

Xiang, R., Yuan, X., Zhang, Y., & Zhang, X. (2025). Quantitative analysis of the labeling quality of biological images for semantic segmentation based on attribute agreement analysis. *Agriculture*, 15(7), 680. Doi: https://doi.org/10.3390/agriculture15070680

Yadav, N., Mathiyazhagan, K., & Kumar, K. (2019). Application of Six Sigma to minimize the defects in glass manufacturing industry. *Journal of Advances in Management Research,* 16(4), 594-624. Doi: https://doi.org/10.1108/JAMR-11-2018-0102

Zanobini, A., Sereni, B., Catelani, M., & Ciani, L. 2016. Repeatability and reproducibility techniques for the analysis of measurement systems. *Measurement,* 86, 125-32. Doi: https://doi.org/10.1016/j.measurement.2016.02.041