Management and Production Engineering Review
Volume 2 • Number 3 • September 2011 • pp. 19–27

THE ESTIMATION OF ALTERNATIVE CONTROL EFFICIENCY WITH THE USE OF THE COHEN'S KAPPA COEFFICIENT

Krzysztof Knop, Stanisław Borkowski

Czestochowa University of Technology, Institute of Production Engineering, Częstochowa, Poland

Corresponding author:
Krzysztof Knop
Czestochowa University of Technology
Institute of Production Engineering
Armii Krajowej 19B, 42-200 Czestochowa, Poland
phone: +48 34 325 03 67
e-mail: kknop@poczta.fm

Received: 15 May 2011
Accepted: 16 August 2011

Abstract
In the article there was prepared the characteristics of the Cohen's Kappa method – one of methods concerning the analysis of measuring systems used for alternative control, recommended by the motorization branch (MSA 4^{th} Edition. Ford, Chrysler, GM, 2010). The analysis of the examined method has been conducted while taking into consideration, as follows: history, use, calculation, results' interpretation, example, pros and cons of the method, research methodology. The Cohen Kappa coefficient, when skillfully used, may cause the increase of control activities efficiency performed by individual inspectors as well as adequate choice of people designed to fulfill the key control operations. The Kappa statistics usability was verified upon estimating the conformity of inspectors. What was underlined is the fact, that statistics of the Kappa type should not be perceived as a determined, unique standard and assumptive way of conformity estimation. Moreover, there was proposed the use of additional helpful indicators during the analysis of alternative control systems efficiency.

Keywords
alternative control, MSA for attributes, Cohen's Kappa coefficient, agreement level, efficiency.

Introduction

Methods of quality control in producing enterprises may be separated into the following specific criteria:

1. The person conducting the control.
2. Number of control degrees.
3. Place of control.
4. Location of control.
5. Time of control processes realizing.
6. Number of controlled objects.
7. Kind of controlled features and object's attributes.
8. Number of controlled features and object's attributes.
9. Estimation kind of the object's attribute.
10. Use of control resources [1–3].

From the *"estimation kind of the object's attribute"* point of view there may be distinguished the numerical and alternative controls. In case, when the estimation concerning product's attributes is based upon measuring features, that means such features that have assigned a specific numerical value, there follows the numerical control. The alternative control is grounded upon non-measuring (or measuring) features while comparing with the formula and issuing a two-stage note, i.e. "appropriate product" or "inappropriate product" or issuing a multivalent note.

The basic difference between numerical and alternative controls has been presented schematically in the Fig. 1.

Let's assume that we are controlling some characteristic, ex. the piston diameter, with the use of numerical and alternative controls. The diameter value of changeable scope depending on process changeability as well as depending value from measuring instrument resolution will be the result of numerical control. On the other hand, while regarding alternative control – the result will be estimated as the number of categories according to specification, i.e. consistent, inconsistent.

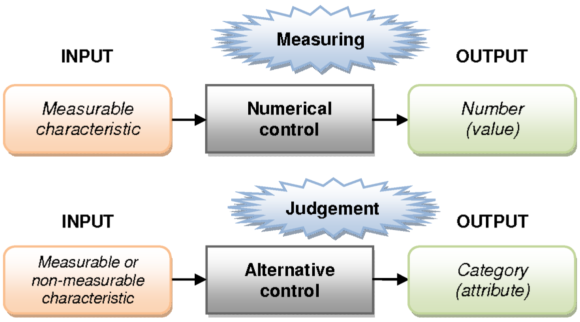

Fig. 1. Numerical and alternative controls.

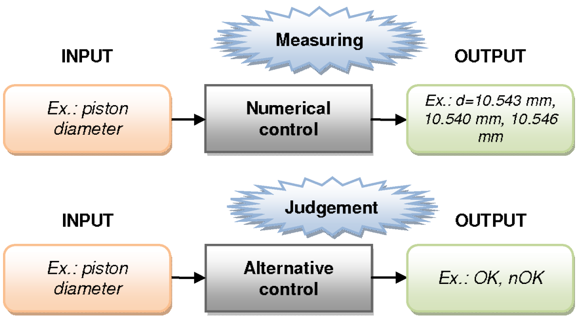

Fig. 2. Results of numerical and alternative control for piston diameter.

Alternative control methods are commonly used, especially where the direct or indirect measurement taken as a numerical value of the interesting feature of product is infeasible, difficult or too expensive. The alternative control may have the form of visual control, the control with the use of examination, the device passing the OK/nOK diagnosis. Such an analysis results in deciding whether a given product meets the specified requirements or not, the product may also be judge in several categories [4].

Despite the development of measuring methods based upon much more objective instruments, in many enterprises the alternative control, that uses only human senses, is the dominant quality control method. Still, there may be noticed many inspectors, that decide about the product further destiny, during inlet, inter or final operation control. The present achievements of metrology allow to replace the man with instrument in every control operation. However, in many cases such project is estimated as unprofitable [5].

It is worthy to learn the efficiency level of quality control, while taking into consideration that the basic aim of such control is exclusion of inappropriate products to client [6], according to the signed agreement requirements as well as while regarding this requirement to attribute control. While estimating the efficiency of the control method, in case of the negative note, there may be learned its causes and because of this fact, the possibility of situation improvement. In such case, there should be used a specific measure that allows to estimate “how well” our control system is able to “indicate losses”. Within the guidelines scope to the analysis of measuring systems designed to the motorization branch [7], there have been showed three methods for alternative control estimation: the Cohen’s Kappa test of agreement between inspectors (Cross Tab Method), Signal Detection Method, Analytic Method. The object of the following thesis is the characteristic of the first within the above mentioned methods, that is: the Cohen’s Kappa method.

The importance of the information concerning control efficiency while regarding the fulfillment of the basic quality control aim, is confirmed by the example presented by J. Obalski [3]: the production consignment, where were found 100 losses, was subject to three complete examination conducted by qualified inspectors. The result was, as follows:
I examination – there were detected → 68 losses.
II examination – there were detected additional → 18 losses.
III examination – there were detected additional → 8 losses. Together: 94 losses.

After three complete examination there were left six undetected losses in the product consignment. The inspector team, specially selected, conduct the fourth examination and detect four more losses. Although, there left two more. What emerges from this situation is the fact, that by the complete examination there may be left undetected from 2 to 32 losses while considering 100 items in the product consignment.

It is worthy to underline, that any control guarantees that the inappropriate product will be delivered to the client. Even 100% control cannot be determined as fully effective. The research indicates that, if the control is conducted by the man, its efficiency does not exceed 90%. The control, in such case, is influenced by inattention, routine, tiredness [2, 6]. However, it is worthy to undertake any efforts to improve quality control systems and increase their efficiency level. Within the scope of the alternative control there may be used the Kappa coefficient as a starting point of undertaking developing activities.

Cohen’s Kappa coefficient

The first evidence of Cohen’s Kappa in print can be attributed to Galton (1982). The seminal paper introducing Kappa as a new technique was published by Jacob Cohen in the Journal Educational and Psychological Measurement in 1960 [8, 9]. At the very beginning, there was used the examination statistics χ^2 or indicator of observed accordance P_0, estimated as a fraction of objects of the same notes, to estimate the accordance level. However, the gained results weren't satisfying. The χ^2 examination determines the relation strength between notes, although it is not always the relation consisting in assigning observations to the same category. Whereas, number of cases, that the supervisors were unanimous about, may be overstated by observations, for which the experts estimated the notes of fate and by accident assigned the same notes. This fact was the cause of suggesting by Cohen the method of improving of accidental conformity P_c determined with the coefficient of the Kappa type.

Cohen claims that choice tendencies, while assigning the object to the analysed class, should be determined for every expert. Then, there may be estimated the value of P_c that meets the conformity chance by the accident. At the same time, he criticizes the attitude indicating that the people, responsible for estimating, are not able to assign any note and present it of fate. Both of them are of the same probability (Scott Kappa coefficient version) [10]. Kappa is claimed to be a developed use of the P_0 classical conformity indicator (% of conformity).

The most popular conformity note model of the Cohen Kappa type, as found in the subject literature, is the case of binary classification of two experts, in other words the model in the 2×2 set. This set means that 2 experts estimate the object from the point of view of two categories 0 and 1, where: 0 – case, when the object doesn't indicate features typical for the group, doesn't belong to the group, 1 – the object holds the specific feature, it belongs to the group. Obviously, the conformity examination may be conducted by more experts ($R > 2$) and for more categories ($C > 2$). In case, where there is a set, in which two supervisors examine the object from the point of view categories ($R = 2$, $C > 2$), there may be used the Table 1 while regarding the gathered data, where n_{lk} means the number of objects classified by the first expert (A) for the category l, and by the second expert (B) for the category k (l, $k = 1, 2$). By n_{l+} marked boundary numerical force, i.e. the overall numbers of objects assigned to the category l-estimated by the supervisor A, estimated with the use of the formula (1):

$$n_{l+} = \sum_{k=1}^{C} n_{lk}. \tag{1}$$

With the help of n_{+k} marked boundary numerical force, i.e. the overall numbers of objects assigned to the category k-estimated by the supervisor B, estimated with the use of the formula (2),

$$n_{+k} = \sum_{l=1}^{C} n_{lk}, \tag{2}$$

where N means the number of all estimated objects.

Table 1
The data format, in case of two supervisors, with the use of a scale concerning many categories.

		Judgement B						
	Category	1	2	…	k	…	C	Sum
Judgement A	1	n_{11}	n_{12}	…	n_{1k}	…	n_{1C}	n_{1+}
	2	n_{21}	n_{22}	…	n_{2k}	…	n_{2C}	n_{2+}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	l	n_{l1}	n_{l2}	…	n_{lk}	…	n_{lC}	n_{l+}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	C	n_{C1}	n_{C2}	…	n_{Ck}	…	n_{CC}	n_{C+}
	Sum	n_{+1}	n_{+2}	…	n_{+k}	…	n_{+C}	N

The Cohen's Kappa coefficient will be expressed by the following formula (3):

$$\kappa_c = \frac{P_0 - P_c}{1 - P_c} \tag{3}$$

P_0 means relatively observed conformity between supervisors, estimated as follows (4):

$$P_0 = \frac{1}{N} \sum_{l=1}^{C} n_{ll}, \tag{4}$$

where n_{ll} are the elements on the board diagonal 1 and P_c – the degree of accidental conformity, approached as follows (5):

$$P_c = \sum_{l=1}^{C} \left(\frac{n_{l+}}{N} \cdot \frac{n_{+l}}{N} \right), \tag{5}$$

where the values $\frac{n_{+l}}{N}$ and $\frac{n_{l+}}{N}$ occurring in the expression are interpreted as preferences assigning the examined object to the category l by every expert [10]. $1 - P_c$ represents margin of possible agreement between not due to hazard.

Interpretation

The threshold values presented by Landis and Koch, as included in the Table 2, are the most frequently used comparative scale for the Cohen's Kappa coefficient.

Table 2
Interpretation of κ_c index value by Landis and Koch.

Kappa	Agreement level
< 0.00	no agrement, less then chance
0.00–0.20	slight
0.21–0.40	fair
0.41–0.60	moderate
0.61–0.80	substancial
0.81–1.00	almost perfect

These threshold values aren't universally accepted. Landis and Koch supplied no evidence to support it, basing it instead on personal opinion [8].

Fleiss's equally arbitrary guidelines propose simplification, that has been presented in the Table 3.

Table 3
Interpretation of $\kappa_c$index value by Fleiss.

Kappa	Agreement level
< 0.40	slight
0.40–0.74	moderate or good
0.75–1.00	perfect

In the Table 4 there have been passed different simplified threshold values proposed by Cicchetti and others [10].

Table 4
Interpretation of κ_c index value by Cicchetti.

Kappa	Agreement level
< 0.40	slight
0.40–0.59	moderate
0.60–0.74	good
0.75–1.00	excellent

Example

There were asked two inspectors (A and B) who were supposed to make an independent classification of 10 products (containing 4 appropriate and 6 inappropriate products) to two quality categories: appropriate product – OK, inappropriate product – nOK. Every inspector was supposed to examine every product three times. The expert also made his own quality evaluation of products.

The inspectors A and B should be estimated from the following points of view:

a) mutual conformity in their decisions,
b) conformity of their decisions with the expert decision,
c) internal conformity of every inspector,
d) choice of the best inspector supposed to fulfilling key control activities.

The examinations have been conducted with the use of the Kappa Cohen coefficient as well as additional indicators. The segregation results of individual products by the inspectors A and B and the expert have been presented in the Table 5.

Table 5
The chart summarizing the classification results of products by the inspectors A and B and the expert.

Number of product	Inspector A			Inspector B			Expert
	Sample 1	Sample 2	Sample 3	Sample 1	Sample 2	Sample 3	
1	OK	OK	OK	OK	OK	nOK	nOK
2	OK	OK	OK	OK	OK	OK	OK
3	OK	OK	OK	OK	nOK	OK	OK
4	nOK	nOK	nOK	nOK	nOK	OK	nOK
5	OK	nOK	nOK	nOK	OK	OK	nOK
6	nOK	nOK	nOK	nOK	nOK	nOK	nOK
7	OK	OK	nOK	OK	nOK	OK	OK
8	nOK	nOK	nOK	nOK	nOK	OK	nOK
9	nOK	nOK	nOK	nOK	nOK	OK	nOK
10	OK	OK	OK	OK	nOK	OK	OK

a) There occurs a case that two inspectors make products evaluation with the use of a two-stage category: OK and nOK. The data format will be presented in the following set 2×2, as in the Table 6.

Table 6
Data format in 2×2 system.

Inspector A	Inspector B		
	nOK	OK	Total
nOK	n_{11}	n_{12}	$\boldsymbol{n_{1+}}$
OK	n_{21}	n_{22}	$\boldsymbol{n_{2+}}$
Total	$\boldsymbol{n_{+1}}$	$\boldsymbol{n_{+2}}$	$\boldsymbol{N}$

In case to indicate characteristic relations, there was made a summary of conformity evaluation results (Table 7). The basis to evaluate the conformity was the number of decisions classifying products to the given two categories.

Table 7
Summary table for A and B inspector decision.

Inspector A	Inspector B		
	nOK	OK	Total
nOK	**9**	**6**	15
OK	**5**	**10**	15
Total	14	16	30

As a result of the examination, the inspectors' decisions were consistent in 19 cases (among 30 of them) and inconsistent in 11 cases.

There was made a conformity calculation between inspectors with the use of the Cohen's Kappa coefficient. The Table presented as 8a), shows the cross chart (of contingencies) concerning actually observed results of sorting. Whereas, the Table presented as 8b), shows the distribution of results frequency that are expected by an accident and were estimated by boundary values (sums). In both charts, the fields representing conformity in sorting are darkened.

Table 8
Contingency table for conformities results: a) observed, b) chance expected [11].

a) Observed

Inspector A	Inspector B		
	nOK	OK	Total
nOK	**9**	6	15
OK	5	**10**	15
Total	14	16	30

Observed Concordant Decisions:
count = 19
proportion = 0.63

b) Chance Expected

Inspector A	Inspector B		
	nOK	OK	Total
nOK	**7**	8	15
OK	7	**8**	15
Total	14	16	30

Observed Concordant Decisions:
count = 15
proportion = 0.5

There has been noted that the inspectors A and B were consistent in 11 decisions concerning product acceptance, and in 10 "NO" decisions – products should be rejected. Thus, the observed percentage agreement is:

$$P_0 = \frac{1}{N}\sum_{l=1}^{C} n_{ll} = \frac{1}{30}(9+10) = 0.63.$$

To calculate P_c (the probability of random agreement) we note that:

– The inspector A indicated OK 15 times and said nOK 15 times as well. Thus, the inspector A said „OK" 50% of the time.
– The inspector B made a decision concerning OK 16 times and nOK 14 times. Thus inspector B said OK 533% of the time.

Therefore the probability that both of them would say OK randomly is $0.500 * 0.533 = 0.267$ and the probability that both of them would say nOK is $0.500 * 0.467 = 0.233$. Thus the overall probability of random agreement is $P_c = 0.267 + 0.233 = 0.50$. In case to confirm the result, there has been used the following formula to calculate the conformity probability by the P_c case:

$$P_c = \sum_{l=1}^{C}\left(\frac{n_{l+}}{N}\cdot\frac{n_{+l}}{N}\right) =$$

$$= \frac{15\cdot 14 + 15\cdot 16}{30\cdot 30} = \frac{450}{900} = 0.50.$$

When the formula of calculating the Cohen Kappa coefficient is used, there is obtained as follows:

$$\kappa_c = \frac{P_0 - P_c}{1 - P_c} = \frac{0.63 - 0.50}{1 - 0.50} = 0.26,$$

what is an equivalent of poor conformity between the people evaluating products.

b) The expert estimated his own decision concerning the conformity of every product with requirements. As a result of his decision, among 10 of examined products, 6 of them were inappropriate and 4 – appropriate. The Table 9 presents mutual relations between decisions of the inspectors A and B and reality, that is the supervised inspector's decision – the expert's.

Table 9
Contingency table for conformities results: a) observed, b) chance expected.

a) Observed

Inspector A	Expert		
	nOK	OK	Total
nOK	**14**	1	15
OK	4	**11**	15
Total	18	12	30

Observed Concordant Decisions:
count = 25
proportion = 0.83

b) Chance Expected

Inspector A	Expert		
	nOK	OK	Total
nOK	**9**	6	15
OK	9	**6**	15
Total	18	12	30

Observed Concordant Decisions:
count = 15
proportion = 0.5

There was made a calculation of the conformity degree between the inspector A and the expert with the use of the Kappa Cohen coefficient.

$$P_0 = \frac{1}{N}\sum_{l=1}^{C} n_{ll} = \frac{1}{30}(14+11) = 0.83,$$

$$P_c = \sum_{l=1}^{C}\left(\frac{n_{l+}}{N}\cdot\frac{n_{+l}}{N}\right) = \frac{15\cdot 18 + 12\cdot 15}{30\cdot 30} = \frac{450}{900} = 0.50,$$

$$\kappa_c = \frac{P_0 - P_c}{1 - P_c} = \frac{0.83 - 0.50}{1 - 0.50} = 0.66.$$

The value $\kappa_c = 0.66$ responds to proper conformity between the inspector A and the expert.

There was conducted an analogical conformity examination between the inspector B and the expert. In this case there was used the Table 10.

Table 10
Contingency table for conformities results: a) observed, b) chance expected.

a) Observed

Inspector A	Expert		
	nOK	OK	Total
nOK	**11**	3	14
OK	7	**9**	16
Total	18	12	30

Observed Concordant Decisions:
count = 20
proportion = 0.66

b) Chance Expected

Inspector A	Expert		
	nOK	OK	Total
nOK	**8.4**	5.6	14
OK	9.6	**6.4**	16
Total	18	12	30

Observed Concordant Decisions:
count = 14.8
proportion = 0.49

The calculations of the Kappa coefficient have been presented below:

$$P_0 = \frac{1}{N}\sum_{l=1}^{C} n_{ll} = \frac{1}{30}(11+9) = 0.66,$$

$$P_c = \sum_{l=1}^{C}\left(\frac{n_{l+}}{N}\cdot\frac{n_{+l}}{N}\right) = \frac{14\cdot 18+12\cdot 16}{30\cdot 30} = \frac{444}{900} = 0.49,$$

$$\kappa_c = \frac{P_0 - P_c}{1-P_c} = \frac{0.66-0.49}{1-0.49} = 0.33.$$

The value $\kappa_c = 0.33$ responds to weak conformity of the inspector B and the expert. The analysis concerning results of Cohen's Kappa conformity coefficients indicates that the conformity between the inspector A and the expert is better than the conformity between the inspector B and the expert.

c) There was executed an internal evaluation of decisions' cohesion concerning individual inspectors while products classifying to one of categories. Moreover, there was conducted the evaluation of decisions' cohesion concerning inspectors and the expert. The results have been presented in the Table 11. The inspector A is characterized by high evaluation repeatability – 80%, whereas the participation of the same decisions in relation to the same products is low and amounts 20% for the inspector B. While taking into consideration the conformity of the same decisions to the same products with the expert opinion, there is reached a conclusion that the inspector A is much better than the inspector B. Their parts are presented correspondingly 70 and 20%. When analyzing the fact, that two inspectors create one control system – this system is rather poor – only 20% of conformities.

Table 11
The indicators of internal conformity concerning every inspector (with the expert).

	The number of all examined products	The participation of products identically evaluated in relation to all products
Inspector A	10	$8/10 = 0.8 * 100\% = 80\%$
Inspector B	10	$2/10 = 0.2 * 100\% = 20\%$
Inspector B and B together (control system)	10	$2/10 = 0.2 * 100\% = 20\%$
	The number of all examined products	**The participation of products identically evaluated and consistent with the expert's decision in relation to all products**
Inspector A	10	$7/10 = 0.7 * 100\% = 70\%$
Inspector B	10	$2/10 = 0.2 * 100\% = 20\%$
Inspector B and B together (control system)	10	$2/10 = 0.2 * 100\% = 20\%$

d) To sum up, the first place in the inspector statement has been awarded to the inspector A, who is characterized with high internal conformity and conformity with the expert in his evaluations. Although, within his evaluations there arisen his tendency of admitting the inappropriate products! The inspector B, because of his weak results while concerning all indicators, should be transferred to the other position or be sent to the improving trainings. After the training, his skills should be verified once again.

Pros and cons

One of the problems with Cohen's Kappa is that it does not always produce the expected answer. In the Table 12 there has been presented the retention of the Cohen Kappa coefficient by the similar level of the observed conformity in all cases $P = 0.60$, based upon accidental data.

The higher evaluation concentration is in one field cornerwise, that resembles the consistent observations, the lower is the value of the Kappa coefficient. The shifting of cases, that the inspectors assigned different notes, does not lead to the essential change within the value of the conformity indicator. As it can be seen in the example, the similar level of the observed conformity between the inspectors $P = 0.60$ does not mean the identical values of the Kappa coefficients. The coefficient values differ from 0.05 to 0.31, that resembles weak and average level of conformity.

Table 12
The retention of the Cohen Kappa coefficient κ_c.

Class	YES	NO	κ_c	Agreement level according to Landis and Koch
YES	25	10	0.05	slight
NO	10	5		
Class	YES	NO	κ_c	
YES	20	10	0.17	slight
NO	10	10		
Class	YES	NO	κ_c	
YES	15	10	0.20	slight
NO	10	15		
Class	YES	NO	κ_c	
YES	25	0	0.20	slight
NO	20	5		
Class	YES	NO	κ_c	
YES	20	0	0.29	fair
NO	20	10		
Class	YES	NO	κ_c	
YES	15	0	0.31	fair
NO	20	15		

Pros of Kappa index:
- Kappa statistics are easily calculated and software is readily available.
- Kappa statistics are appropriate for testing whether agreement exceeds chance levels for binary and nominal ratings.

Cons of Kappa index:
- Kappa is not really a chance-corrected measure of agreement.
- Kappa is an omnibus index of agreement. It does not make distinctions among various types and sources of disagreement.
- Kappa is influenced by trait prevalence (distribution) and base-rates. As a result, kappas are seldom comparable across studies, procedures, or populations.
- Kappa may be low even though there are high levels of agreement and even though individual ratings are accurate. Whether a given kappa value implies a good or a bad rating system or diagnostic method depends on what model one assumes about the decision-making of raters.
- With ordered category data, one must select weights arbitrarily to calculate weighted kappa.
- Kappa requires that two rater/procedures use the same rating categories. There are situations where one is interested in measuring the consistency of ratings for raters that use different categories (e.g., one uses a scale of 1 to 3, another uses a scale of 1 to 5).
- Tables that purport to categorize ranges of kappa as "good", "fair", "poor", etc. are inappropriate; do not use them [12, 13].

There is wide disagreement about the usefulness of kappa statistics to assess rater agreement. Kappa doesn't take into account the level of agreement between both appraisers, it only analyzes if there is concordance or not. This analysis doesn't tell how the measurement system sorts good parts bad ones (for this analysis we use the Effectiveness; Effectiveness = Number of correct decisions/Total opportunities for a decision) [14].

There are two Kappa paradoxes. The first paradox is connected with term "…the proportion of units for which agreement is expected by chance" alone. This term is relevant only in case the raters are independent, but the clear absence of independence calls its relevance into questions. The second Kappa paradox is characterized by a low Kappa value associated with good agreement among raters on marginal counts. The root cause of the second Kappa paradox is high trait prevalence [15].

Methodology

The scope of activities, within the evaluation of measuring system used to alternative control in the enterprise, includes:

1. Preparation of examinations, that is the choice of inspector team.
2. Preparation of a product sample. The sample should not be converted from the process at random. It should be selected by the qualified staff and must be determined as conforming or nonconforming. The sample should include both inappropriate and appropriate parts. The parts chosen to the selection should be appropriate in 1/3, inappropriate in 1/3 and 1/3 of them should be at the boundary of conformity and non-conformity. The results in the total sample will be only appropriate and inappropriate [16]. The sample size depends on the number of appraisers and the number of gages used during testing. According to the opinion most often expressed in the literature, around 20 to 30 samples and two to three measurement repetitions are required for the testing results to be reliable. The sample size and number of repetitions depend on the nature of the process itself, arrangements between the supplier and customer of parts, as well as on measurement procedure. Recommended sample sizes and repetition numbers put together in Table 13. These data should be treat as recommendation in the choice of minimal quantity of parts but not as standard [4].
3. The examination execution – conducting and supervising of the inspectors team, registering of the results. After selecting the correct sample, the parts are evaluated by inspectors in the accidental order, and the results noted in the sheet. From the

technical point of view, it is recommended to indicate products previously, i.e. assigning the number on one part of the product. Such indication should not be visible for any inspector during the examination. Only the supervisor should be able to notice such indication. In case of many repetitions, it is recommended that the examination be performed by individual inspectors by turns. Then, the supervisor makes calculations by indicating the control system capacity indicators with the use of the Kappa method.

4. Performing the statement that contains the interpretation of the gained results, the suggestions concerning possibilities of improvement [4].

Table 13
Recommended sample sizes and repetition numbers.

Quantity of Appraisers	Quantity of Gages	Minimum number of parts	Minimum number of measurement per part
1	0 (without any gage, e.g. visual assessment)	24	5
1	1		
2	0	18	4
2 or More	1		
2	2 or More		
3 or More	0	12	3
2 or More	2 or More		

Conclusions

The Cohen's Kappa method is the method of evaluating the conformity between two inspectors. This method is recommended by the motorization branch and described in the guidelines *Measurement System Analysis (MSA)*. Kappa index is still evolving. Kappa index also occurs in the form of weighted Kappa (counts disagreements differently and is especially useful when codes are ordered) and Kappa maximum (helps interpret the value of kappa actually obtained) [8]. The examination execution, with the use of the Kappa coefficient, may fundamentally affect the improvement of activities performed by the quality control service in the enterprise. The calculation of the Kappa coefficients make up the system of supporting indicators may affect decisions concerning the best inspectors to individual tasks.

Supporting indicators describe detailed aspects of functioning the quality control system; they can be divided into two categories: the first category includes those indicators that describe overall efficiency of quality control system, e.g.: *% Appraiser, % Score vs. Attribute, System % Effective Score and System % Effective Score vs. Reference*; the second category includes those indicators that describe efficiency of quality control system, taking into account the correctness of products classification (error risk of I^{st} and II^{nd} type), e.g. *false negative, false positive, false alarm rate, miss rate, bias (regarding products and decisions)*.

These indicators should be used in addition to the index Kappa because Kappa does not give us complete information about the quality control system we have in the company – evaluates the control system, only from the standpoint of the degree of agreement between evaluators. If you would like to know something more, it is recommended to calculate supporting indicators.

Calculation of supporting indicators helps us, in a comprehensive way to assess the control system, pay attention to those aspects of the system that require special attention because of the company costs and customer satisfaction (controller tendency to make mistakes I^{st} and II^{nd} type). Based on this knowledge you can take decisions related to the improvement of the current control system, for example, you can choose a team of people that are the best in fulfilling the key control activities. It really encourages to perform such examinations in the company that using alternative control system.

References

[1] Czyżewski B., *Internal quality control in an industrial enterprise*, Greater Quality Club NOT, Poznan, 2009 (in Polish).

[2] Hamrol A., *Quality management with examples*, PWN, Warsaw, 2008 (in Polish).

[3] Olejnik T., Wieczorek R., *Inspection and quality control*, PWN, 1982 (in Polish).

[4] Feliks J., Lichota A., *Computer-aided measurement analysis*, Archives of Foundry Engineering, 3, 2010.

[5] http://www.qnowhow.pl/obiekty/mod/szczegoly/Obi_id/334/czy-ufasz-kontroli-wzrokowej.html

[6] Borkowski S., Siekański K., *Analysis of the Quality Control System in the Foundry with Implementing the Quality Management System According to ISO 9001*, W: 7 Medzinarodne vedecke sympozium. Kvalita a spol'ahlivost strojov. Sprievodna akcia Medzinarodneho strojarskeho vel'trhu 2002 v Nitre, Nitra, 2002.

[7] AIAG Measurement Systems Analysis (MSA). Reference manual by Chrysler, Ford and General Motors. 4^{th} Edition, 2010.

[8] http://en.wikipedia.org/wiki/Cohen%27s_kappa

[9] Wood J.M., *Understanding and Computing Cohen's Kappa: A Tutorial WebPsychEmpiricist*, Web Journal at http://wpe.info/, 2007.

[10] Jarosz-Nowak J., *Models assessing agreement between two experts, using the kappa coefficient*, Applied Mathematics 8, Warsaw, 2007 (in Polish).

[11] Cunningham M., *More than Just the Kappa Coefficient: A Program to Fully Characterize Inter-Rater Reliability between Two Raters*, SAS Global Forum 2009, Statistics and Data Analysis, Paper 242–2009.

[12] http://faculty.vassar.edu/lowry/kappaexp.html

[13] www.john-uebersax.com/stat/kappa.htm#procon

[14] Casey P., Altobelli G., Pignatelli P., *Application of the hypothesis analysis method using Cohen's Kappa index to measure the agreement between leather sorter.*

[15] Gwet K., *Handbook of Inter-Rater Reliability (Second Edition)*, Advanced Analytics, LLC, Gaithersburg, USA, 2010.

[16] Hradesky J., *Total Quality Management Handbook*, McGraw Hill, 1995.