

The Intraclass Correlation Coefficient

Is your measurement system adequate?

Donald J. Wheeler

In my July column “Where Do Manufacturing Specifications Come From?” we found that the intraclass correlation coefficient is the natural measure of relative utility. This measure is both theoretically sound and also easy to explain. This column will look at how to use the intraclass correlation to characterize the relative utility of a measurement system for use with a particular product.

The issue of relative utility has two components. The first of these is whether or not you can use a measurement system to characterize items or batches as conforming or nonconforming. The answer to this question was provided in my June column “Is the Part in Spec?” and this answer was expanded in my July column mentioned above. In this column I shall address the second aspect of relative utility which is whether or not you can use your measurement system to track your process on a process behavior chart. This aspect of relative utility will consider such questions as: Can you detect process shifts when they occur? Can you track process improvements? When do you need to consider finding an alternative measurement system? However, before we get to these questions, we shall need to have some notation in the interest of clarity.

Let the product measurements be denoted by X . These product measurements may be thought of as consisting of two components. The value of the item being measured may be denoted by Y , while the error of the measurement may be denoted by E . Thus, $X = Y + E$. If we think about these quantities as variables, then the variation in the product measurements, $Var(X)$, can be thought of as:

$$Var(X) = Var(Y) + Var(E)$$

where $Var(Y)$ is the variation in the stream of product values and $Var(E)$ is the variation in the stream of measurement errors.

With this notation, the intraclass correlation coefficient is defined as:

$$\text{Intraclass Correlation Coefficient} = \frac{Var(Y)}{Var(X)} = 1 - \frac{Var(E)}{Var(X)}$$

Thus, the intraclass correlation coefficient is defined to be that proportion of the variation in the product measurements that can be attributed to the product stream, and it is also the complement of that proportion of the variation in the product measurements that can be attributed to the measurement system.

We shall begin with a specific example provided by one of my clients. The average and range chart for the month of June is shown in Figure 1. The daily subgroups consist of eight product values. This plant was running with a 40 hour work week, and the eight items measured each day were selected at the start of each hour during that day’s shift. These items were measured in the lab and the values were reported back to production. The average chart shows a process that

is very unpredictable from day to day.

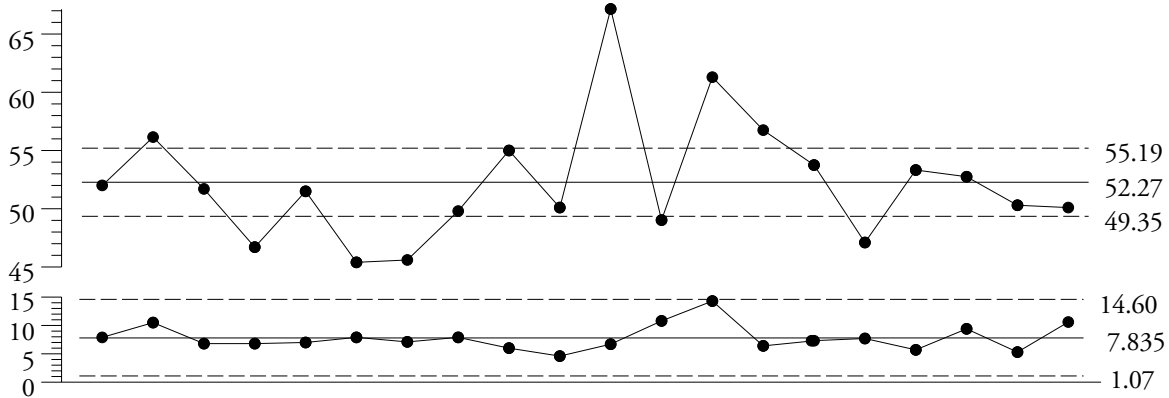


Figure 1: June's Average and Range Chart for Product 2131 Measured with Test Method 65

Since the characteristic measured for the average and range chart of Figure 1 was difficult to measure, the production department blamed the unpredictability seen in Figure 1 on the measurement system. However, the lab was also using this same measurement system to measure a known standard. These readings were done every Monday, and the resulting consistency chart for the first 25 weeks of the year is shown in Figure 2. Here we see that the measurement system is being operated in a predictable manner.

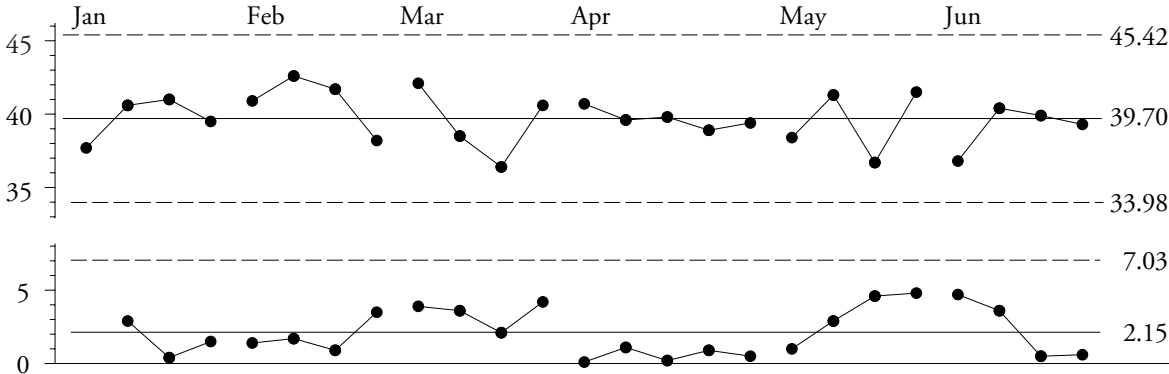


Figure 2: *XmR* Chart for Repeated Measurements of a Known Standard Using Test Method 65

Since the chart in Figure 2 does not show evidence of any inconsistency in the measurement system, we have to conclude that the unpredictability seen in Figure 1 is coming from the operation of the production process, rather than being attributable the measurement system.

Given the information provided by these two charts we can estimate the intraclass correlation coefficient for the use of test method 65 with product 2131. The consistency chart in Figure 2 provides us with an estimate of measurement error. Here we divide the average moving range of 2.15 by $d_2 = 1.128$ to estimate the standard deviation of test-retest error:

$$\text{Estimate of } SD(E) = \frac{2.15}{1.128} = 1.906$$

At the same time, the average range of 7.835 in Figure 1 will, when divided by $d_2 = 2.847$, provide an estimate of the routine variation of the product measurements, X .

$$\text{Estimate of } SD(X) = \frac{7.835}{2.847} = 2.752$$

We convert these two estimates of dispersion into an estimate of the intraclass correlation coefficient by squaring each, forming the ratio, and subtracting from 1.0:

$$\text{Intraclass Correlation} = 1 - \frac{\text{Var}(E)}{\text{Var}(X)} = 1 - \frac{(1.906)^2}{(2.752)^2} = 0.520$$

Thus, when we use test method 65 to measure product 2131 we can say that 52 percent of the routine variation in the product measurements comes from the product stream, while 48 percent of the routine variation in the product measurements comes from the measurement system. While this does not sound good, Figure 1 shows that it is good enough to reveal the problems in production. Moreover, when they try to remove these problems from the production process, test method 65 will still be sufficient to tell them if they have improved things.

Therefore, the first lesson of the example above is that process behavior charts do not require measurement systems that are perfect, or even nearly perfect. They work even when the data contain substantial amounts of measurement error. So, the question becomes "How bad do the measurements have to be before they are useless on a process behavior chart?" To answer this question we will need to resort to some theoretical results.

HOW SIGNALS GET ATTENUATED

Since the question above concerns our ability to detect signals that occur within the production process, it will be helpful to see how signals are attenuated and how that relates to the intraclass correlation coefficient. We begin with our model:

$$X = Y + E$$

and consider what would happen if the average value for the production process was shifted by an amount equal to $3 SD(Y)$. If the intraclass correlation coefficient happened to be 0.81, then the square root of the intraclass correlation would be 0.90 and we would have:

$$3 SD(Y) = 3 [0.90 SD(X)] = 2.7 SD(X)$$

So that a 3 standard deviation shift in the process values, Y , ends up looking like a 2.7 standard deviation shift in the product measurements, X . Hence, measurement error will attenuate production process signals so that they show up in the product measurements with a reduced signal strength, and that reduction is defined by the square root of the intraclass correlation:

$$\text{Production Process Signal Strength} = \sqrt{\text{Intraclass Correlation}}$$

Figure 3 shows the relationship between production process signal strength and the intraclass correlation. From the curve and the preceding example we can see that an intraclass correlation in the vicinity of 0.80 will result in approximately a 10 percent attenuation of signals from the production process. Likewise, an intraclass correlation in the vicinity of 0.50 will result in approximately a 30 percent attenuation of any signals originating in the production process. Thus, with an intraclass correlation of 0.52, the signals we see in Figure 1 have been attenuated by

approximately 28 percent due to the effects of measurement error! While this may have obscured some of the smaller signals, we still have more than enough to use in looking for assignable causes in this production process.

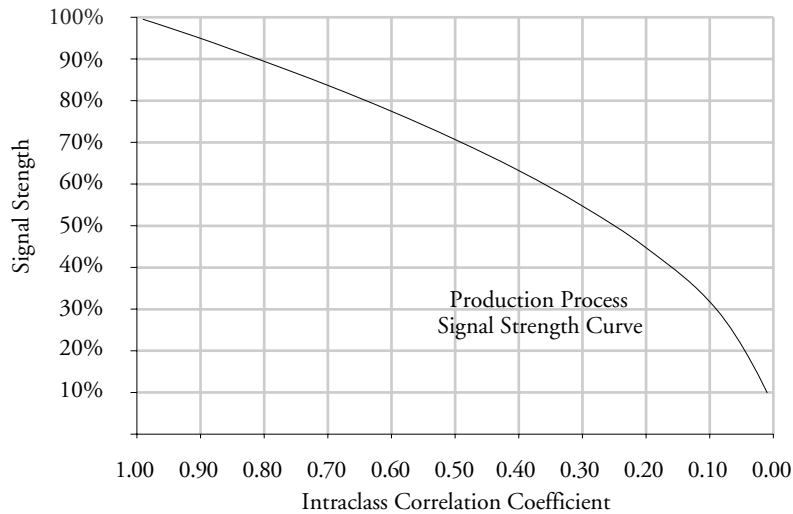


Figure 3: Production Process Signal Strength versus the Intraclass Correlation Coefficient

An alternate way of expressing this same idea is to consider the extent to which measurement error will inflate the limits on a process behavior chart. From the preceding argument, this inflation will be the inverse of the amount by which the signal strength is deflated:

$$\text{Inflation Factor for Process Behavior Chart Limits} = \frac{1}{\sqrt{\text{Intraclass Correlation}}}$$

For the chart in Figure 1, with an intraclass correlation of 0.52, this inflation factor turns out to be 1.387. So we can say that the limits in Figure 1 have been inflated by 39 percent, or we may say that the signals from the production process have been attenuated by 28 percent (the signal strength is 0.721). Either way we are describing the same thing.

HOW THIS ATTENUATION AFFECTS OUR ABILITY TO DETECT SIGNALS

Traditionally we use power function curves to characterize how a process behavior chart will detect signals. However, these power function curves are always computed with the implicit assumption that there is no measurement error. The curve in Figure 3 allows us to adapt these theoretical power function curves to allow for the signal attenuation effects of measurement error. When we do this we obtain the curves shown in Figures 4 and 5.

Figure 4 is for the case where we are only using Detection Rule One (where a point outside the three-sigma limits is interpreted as a signal). The vertical axis shows the theoretical probability of detecting a signal within $k = 10$ subgroups of when that signal occurs, while the horizontal axis shows the size of the signal. The various curves represent the power function for intraclass correlation coefficients ranging from 1.00 to 0.00. The dots show how the probability of detecting a three standard error shift will change with the changing intraclass correlation.

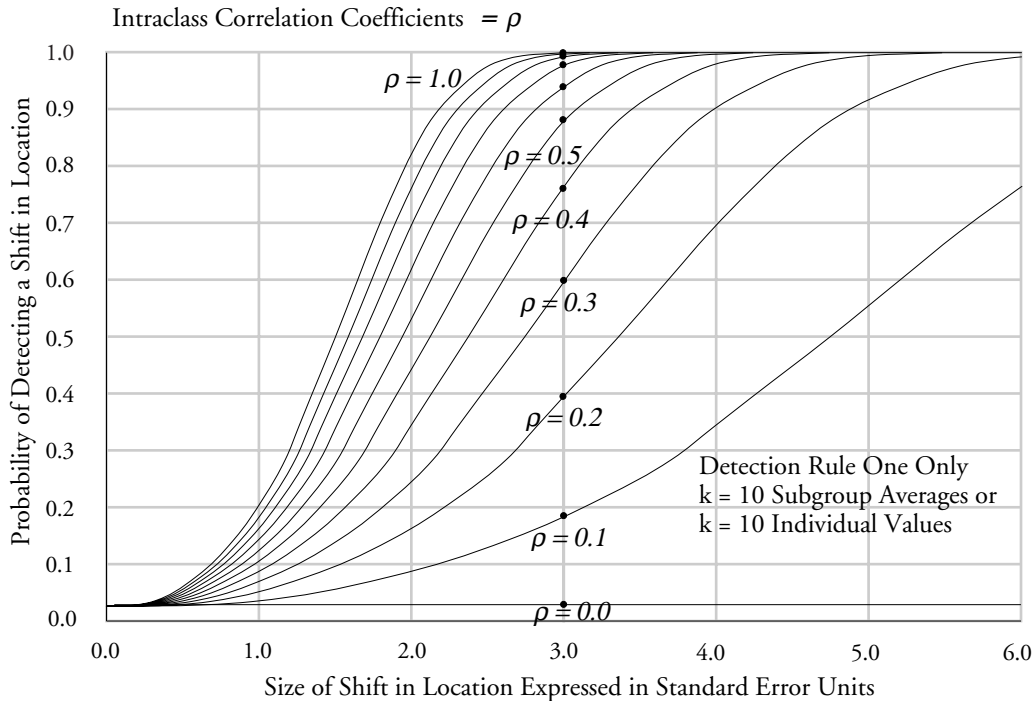


Figure 4: The Effect of the Intraclass Correlation Coefficient Upon Power Functions for Charts for Location

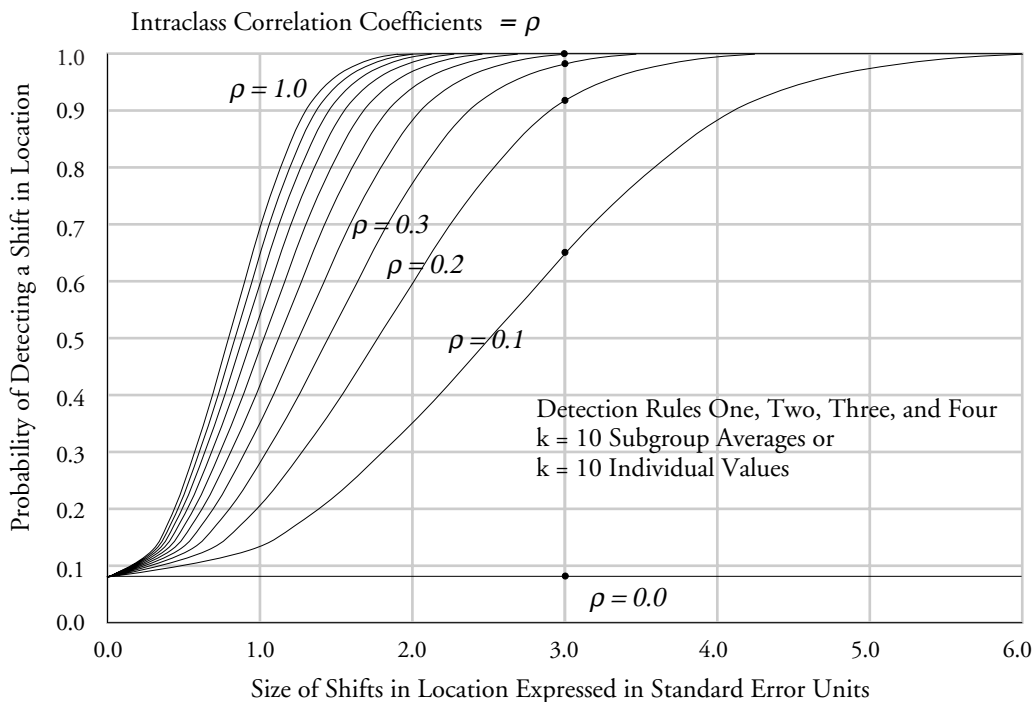


Figure 5: The Effect of the Intraclass Correlation Coefficient Upon Power Functions for Charts for Location

Figure 5 is for the case where we are using Detection Rules One, Two, Three, and Four. With Detection Rule Two a signal is indicated whenever two out of three successive values are on the

same side and are more than two sigma units away from the central line. With Detection Rule Three a signal is indicated whenever four out of five successive values are on the same side and are more than one sigma unit away from the central line. With Detection Rule Four a signal is indicated whenever eight successive values are on the same side of the central line.

To see the effect of measurement error upon our ability to detect a signal of a given size we need to look at Figures 4 and 5 differently. Using the dots in Figures 4 and 5 we construct a pair of curves that show how the probability of detecting a three sigma shift using an XmR Chart will vary with the value of the intraclass correlation. These curves are in Figure 6.

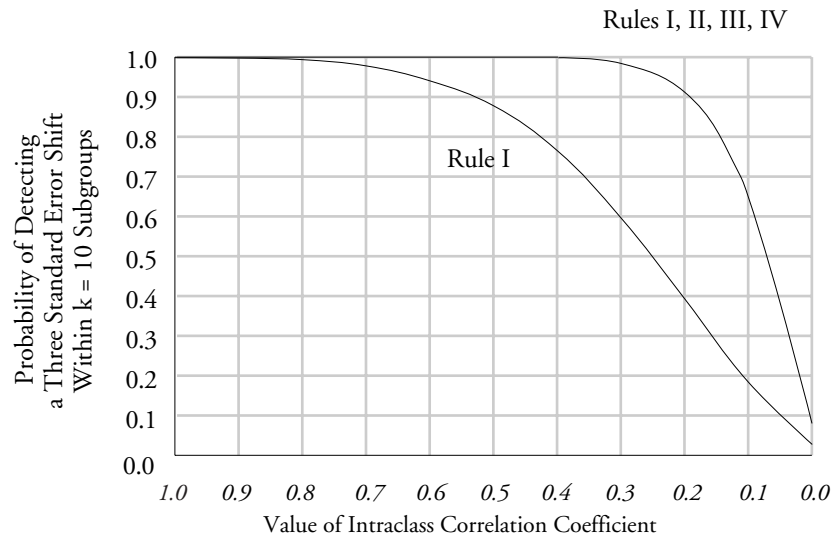


Figure 6: How Measurement Error Affects Our Ability to Detect a Three Sigma Shift

The curves in Figure 6 define three natural breaks to use in characterizing the relative utility of a measurement system for a particular application.

- When the intraclass correlation exceeds 0.80 we are virtually certain to detect our three sigma shift using Rule One alone. In this case the measurement system will be said to provide a First Class Monitor for the product being measured.
- When the intraclass correlation exceeds 0.50 we have better than a 90 percent chance of detecting our three sigma shift using Rule One alone. Therefore, when the intraclass correlation is between 0.80 and 0.50 the measurement system will be said to provide a Second Class Monitor.
- When the intraclass correlation exceeds 0.20 we have better than a 90 percent chance of detecting our three sigma shift using Rules One, Two, Three, and Four together. Therefore, when the intraclass correlation is between 0.50 and 0.20 the measurement system will be said to provide a Third Class Monitor.
- As the intraclass correlation falls below 0.20 the probabilities of detecting our three sigma shift will rapidly vanish. When the intraclass correlation is below 0.20 the measurement system will be said to provide a Fourth Class Monitor.

THE RELATIONSHIP WITH PROCESS CAPABILITY

The capability ratio describes the relationship between the specification limits and the variation in the process measurements, $SD(X)$. Specifically:

$$Capability\ Ratio = \frac{Upper\ Spec - Lower\ Spec}{6\ SD(X)}$$

From our definition of the intraclass correlation coefficient:

$$Intraclass\ Correlation = 1 - \frac{Var(E)}{Var(X)}$$

we can rewrite the expression for the capability ratio as:

$$Capability\ Ratio = \frac{Upper\ Spec - Lower\ Spec}{6\ SD(E)} \sqrt{1 - Intraclass\ Correlation}$$

Now consider what happens whenever we improve the production process. As $SD(Y)$ gets smaller $SD(X)$ will also get smaller, but $SD(E)$ will remain the same. As $SD(X)$ gets smaller the Capability Ratio will increase, but the intraclass correlation coefficient will get smaller.

Inspection of the expression above will reveal that the ratio on the right is the inverse of the Precision to Tolerance Ratio, or P/T Ratio. Denoting this inverse of the P/T ratio by λ we could rewrite the expression above as:

$$Capability\ Ratio = \lambda \sqrt{1 - Intraclass\ Correlation}$$

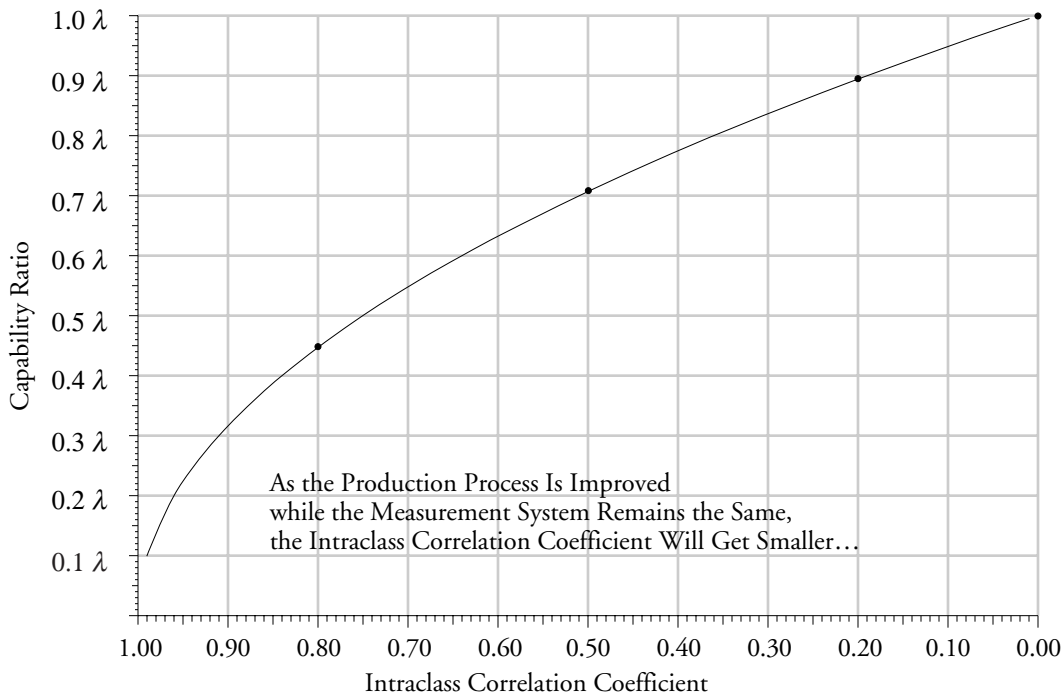


Figure 7: The Relationship between Intraclass Correlation and the Capability Ratio

The curve in Figure 7 can be helpful in determining when you need to look for a new

measurement system. It defines those capability ratios that can be attained with the current measurement system. Specifically, the dots in Figure 7 define the maximum values for the capability ratio while the measurement system continues to operate within each classification.

For example, if you have a First Class Monitor and you improve the production process to the point that your intraclass correlation drops to 0.80 you will have a crossover capability ratio equal to:

$$C_{p80} = \frac{USL - LSL}{6 SD(E)} \sqrt{1 - 0.80}$$

This crossover capability defines the maximum capability that your process can attain with the current measurement system while that measurement system remains a First Class Monitor. Improvements that result in larger capabilities will also degrade the measurement system to a Second Class Monitor.

If you have a Second Class Monitor and you improve the production process to the point that your intraclass correlation drops to 0.50 you will have a crossover capability ratio equal to:

$$C_{p50} = \frac{USL - LSL}{6 SD(E)} \sqrt{1 - 0.50}$$

This crossover capability defines the maximum capability that your process can attain with the current measurement system while that measurement system remains a Second Class Monitor. Improvements that result in larger capabilities will also degrade the measurement system to a Third Class Monitor.

If you have a Third Class Monitor and you improve the production process to the point that your intraclass correlation drops to 0.20 you will have a crossover capability ratio equal to:

$$C_{p20} = \frac{USL - LSL}{6 SD(E)} \sqrt{1 - 0.20}$$

This crossover capability defines the maximum capability that your process can attain with the current measurement system while that measurement system remains a Third Class Monitor. Improvements that result in larger capabilities will also degrade the measurement system to a Fourth Class Monitor. This crossover capability will always be twice the size of the C_{p80} value. From a practical perspective C_{p20} is the maximum capability that you can expect to achieve with a given measurement system. This is due to the way that measurement error seriously erodes our ability to detect process improvements with a Fourth Class Monitor.

Finally, an unreachable upper bound for the capability ratio is seen from Figure 7 to be λ , the inverse of the P/T ratio. This corresponds to the impossible situation where the process variation goes to zero.

For an example, assume that the watershed specifications for product 2131 in Figure 1 are 42.5 to 57.5. Then with $SD(E) = 1.9$, we find the inverse of the P/T ratio to be $\lambda = 1.316$. The current measurement system is unlikely to take us beyond the point where the intraclass drops to 0.20, and at this point we would have a capability ratio of 1.18. If the customer is asking for capabilities in excess of 1.33, there is simply no way that the current measurement system can take us there, and both we and the customer need to understand this. Either the capability target needs to be relaxed, or else we need a new measurement system.

HOW THE INTRACLASS CORRELATION CHARACTERIZES RELATIVE UTILITY

Now we have the ability to answer the questions regarding the relative utility of a measurement system for tracking process performance. The intraclass correlation allows us to establish four classes of process monitors.

When the intraclass correlation exceeds 0.80 you will have a First Class Monitor. Signals coming from the production process will be attenuated by 10 percent or less. You will be virtually certain to detect a three standard error shift within ten subgroups using a process behavior chart and Rule One alone. You will be able to track process improvements up to the point where you reach a process capability equal to the crossover capability C_{p80} without any significant degradation in the relative utility of the measurement system.

| Intraclass Correlation | Attenuation of Process Signals | Chance of Detecting a 3 Std. Error Shift | Ability to Track Process Improvements |
|------------------------|--------------------------------|--|---------------------------------------|
| 1.00 | | | |
| First Class Monitors | Less than 10 Percent | More than 99% with Rule One | Up to C_{p80} |
| 0.80 | | | |
| Second Class Monitors | From 10 % to 30 % | More than 88% with Rule One | Up to C_{p50} |
| 0.50 | | | |
| Third Class Monitors | From 30% to 55% | More than 91% w/ Rules 1, 2, 3, 4 | Up to C_{p20} |
| 0.20 | | | |
| Fourth Class Monitors | More than 55 Percent | Rapidly Vanishes | Unable to Track |
| 0.00 | | | |

Figure 8: The Four Classes of Product Monitors

When the intraclass correlation falls between 0.80 and 0.50 you will have a Second Class Monitor. Signals coming from the production process will be attenuated by anywhere from 10 percent to 30 percent. You will be virtually certain to detect a three standard error shift within ten subgroups using a process behavior chart and Rules One, Two, Three, and Four. With Rule One alone you will still have better than a 90 percent chance of detecting this shift. You will be able to track process improvements up to the point where you reach a process capability equal to the crossover capability C_{p50} while your measurement system remains a Second Class Monitor. With low-end Second Class Monitors it can be very helpful to maintain a consistency chart like that in Figure 2 for the measurement system.

When the intraclass correlation falls between 0.50 and 0.20 you will have a Third Class Monitor. Signals coming from the production process will be attenuated by anywhere from 30 percent to 55 percent. You will be still have better than a 90 percent chance of detecting a three standard error shift within ten subgroups using a process behavior chart and Rules One, Two, Three, and Four. You will be able to track process improvements up to the point where you reach a process capability equal to the crossover capability C_{p20} while your measurement system remains a Third Class Monitor. With Third Class Monitors it is important to maintain a

consistency chart for the measurement system so that you can isolate the sources of any signals seen on the chart for the production process.

When your intraclass correlation falls below 0.20 you will have a Fourth Class Monitor. While such monitors may detect very large process changes, they are virtually worthless for tracking process improvements. (This is because any process improvement will substantially reduce the effectiveness of the Fourth Class Monitor.)

When do you need to consider finding a new measurement system? When you have a low-end Third Class Monitor you are approaching the limit of what your current measurement system can do in terms of process monitoring. While Fourth Class Monitors may sometimes be useful in sorting product relative to specifications, they cannot detect small process changes, nor can they be used to track process improvements. So while the answer to the question of replacing one measurement system with another will depend upon the economics of the situation, understanding the utility, or the limitations, of your current measurement system is a key element in making this decision.

For more information on this topic see my book *EMP III: Evaluating the Measurement Process and Using Imperfect Data* available from SPC Press.