Scientific Method and Continual Improvement

David and Sarah Kerridge

Scientific method is the only sure way to learn from experience. It uses theory, but continually tests and improves it. This is the exact opposite of blind faith. The basic rules of science are always the same, but human systems, such as those in management, can not be dealt with as simply as those in subjects like physics. This is why many people think that scientific method does not apply to management. But it does, provided we take extra care. This involves:

1 **Repeated and cautious experiments using the Deming Cycle.**

2 **Careful definition of observations and measurements using the concept of Operational Meaning.**

3 **Stabilisation of systems and measurements, by understanding variation and using control charts.**

4 **A unified approach to improvement, not confined to one method, or just one process at a time, but embracing the whole system.**

Practical Benefits of Scientific Method

The scientific approach:

1. **Makes us reject wrong ideas, even though they seem to work**
2. **Leads to continual improvement of things that do work**
3. **Shows the way to successful innovation**

Wrong ideas poison the whole system. The more obvious they seem, the greater the harm, but the less we can see it. Driving them out is very hard, but essential.

Most innovation is not based on science but on “inspiration”. Someone has a brain-wave, and the drive and enthusiasm to see it through. The result is success, when we say “What a genius!”, or failure, when we say “We told you so.”

This means waste. Ideas are not tried, unless a forceful person is behind them, and bad ideas often ***seem*** to succeed, or ***do*** succeed for a time, from sheer enthusiasm. Even when a good idea is tried, it may fail because it was tried in the wrong way, or at the wrong time. Our future depends on continual improvement and innovation, and waste of ideas will be fatal. The theory of knowledge, which deals in evidence and prediction, shows us a better way.

Science as a way to improve quality has a long history. It was first applied to medicine, and has been a great success. Health is the quality of life itself. Whatever the failings of medicine today, no one would go back to the unscientific medicine of the past, when most treatment did more harm than good.

The cycle of improvement

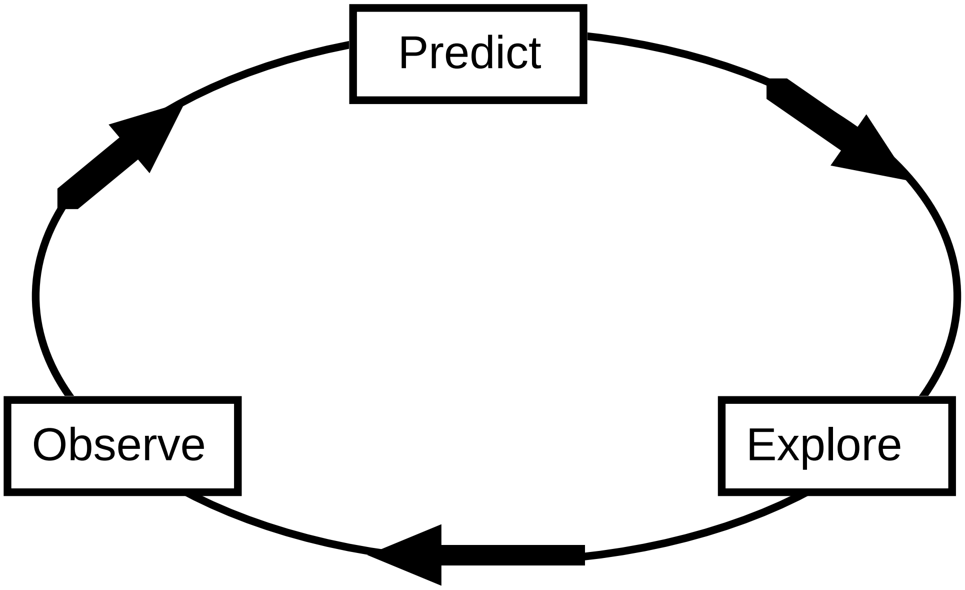
The strategy for learning in pure sciences has three stages:

Observe: Gather facts about the problem as widely as possible.

Predict: Turn the “theoretical” explanation of these facts into a definite rule for predicting the future. It must, of course, agree with what we already know.

Explore: Check the rule by ***actively trying*** to find situations in which it either predicts wrongly, or makes no definite prediction.

Every time we prove the rule wrong, or find a “new” situation, we have new information, and improve our predictions. This produces a cycle of ever-increasing knowledge, which is not just abstract knowledge, but tells us what will happen if we change things. So we take action with greater confidence, and to a greater effect. This is the cycle, which we use to improve a general theory.



:

We can join the cycle at any point, often starting with vague and ill defined ideas, but we must keep going round it in the right direction. Many improvements are small refinements which do not change the basic idea. Others are more radical. But this approach works with any “fundamental” theory, that holds at all times and places.

Prediction not just explanation

Sometimes our explanation of ***why*** our predictions work changes completely as we progress, so that it seems as if we are starting from nothing. For example, Copernicus decided that the earth goes round the sun, instead of the sun going round the earth. This made astronomy easier to understand, at the time, though we now know that the two theories are equally true. Although explanations often change, we have immediate practical benefits. At each stage the rule works better than before, and can be used until another improvement is found.

An old theory may still be useful, if it predicts well. We gain by knowing its limitations better. Newton’s theory of Gravitation is known to be “wrong” in the sense that Einstein’s Relativity predicts better, and under a wider range of circumstances. But the old theory is easier to use, and still good enough to plan a journey to the moon. The more we know about the cases where it fails (such as near black holes) the safer it is to use it.

When we talk of “theory”, we may mean a whole range of different things, such as a vague hunch, or an attractive but untested explanation. These can all help us to learn, but we must never trust an untested theory. The theory of Gravitation, because it has been fully tested, and its limitations known, is safe and useful. We can learn, even with a wrong theory, if we know how.

Failure to learn

There are many striking examples of failure to learn. Our ancestors believed that bad smells made you ill. This led to the name malaria (bad air), a disease that was common near swamps. It was a useful theory, because it told you to avoid smelly things like sewage, rotten fish, and marshes, all of which are dangerous. Unfortunately they used the theory to predict that a bunch of sweet-smelling flowers, which disguised the smells, would keep you healthy. They did not check the theory. If they had they would have found a better one.

We are too easily satisfied with an explanation that sounds convincing. When we are used to a theory, especially if we thought of it ourselves, we become attached to it, and stop looking for weaknesses in it. Just the opposite: we see what we expect to see. We even fail to see exceptions when they are staring us in the face. This explains the survival of many political and management theories. We must train ourselves to look for gaps in every theory.

Application to management

The first successes of science were in subjects like astronomy. Most variation in results was due to errors of measurement. They were small compared to the movements of the planets, so that it was easy to check a theory. When scientists began to study plants, animals, and people, variation became more important. This led to the development of statistical methods.

The new problems also made scientists rethink the idea of “cause”. Instead of a single cause for every effect, various factors combine to increase or decrease the probability that an effect will occur. For example, the “cause” of

AIDS, or any infection, is not just a virus, but the combination of risk factors. One is the virus, some may not yet be known. This way of thinking about causes was the start of system thinking.

Then scientists saw the need for very precise definitions, and developed the concept of Operational Meaning. Every measurement or prediction rule must be defined in terms of actions, rather than concepts. This is explained in more detail in the section which starts on page 10 of this document.

Shewhart and stability

Walter Shewhart was one of the most profound thinkers of our age. He saw the importance for management of scientific method, since effective action depends on prediction. But he also saw that scientific method needed one more vital development. It must have a time dimension.

The "laws of nature" investigated by early scientists were timeless. A chemical formula, once discovered, does not change. But in management, and other complex systems, change is the rule, not the exception. The best way to do something today may not be the best tomorrow, because of innovation, or change elsewhere in the system. We need to know how long our predictions will continue to work. To study stability through time, Walter Shewhart developed the control chart. He called a stable system "under statistical control." In this state there are no special causes, but only common cause variation.

In the form developed by Shewhart, scientific method becomes a universal way of learning, whereas before it could only be used in a limited range of "fundamental" problems, which are basically already stable. The study of these fundamental problems provides "subject knowledge": essential, but not enough to manage any particular system.

Learning is difficult when everything keeps changing. Even if we do nothing, things do not stay the same, so it is hard to see whether change is due to our action, or would have happened anyway. ***Nothing*** is completely stable, in the long run, but the longer it remains stable, the faster we can learn. Another problem is that whenever we change anything in a complicated system, the short term effects are often very different from the long-term effects, and changes in one part of a system have unexpected consequences elsewhere.

All management systems involve people, so that there are important psychological dimensions to every problem. One is the tendency to self-deception, which we must overcome. Apart from this, we cannot completely separate the mechanical effect of a change from the psychological effect, and should not wish to. What matters is that we get predictable and lasting improvement.

The Deming Cycle (PDSA cycle)

The extra care needed in applying scientific method to management means that we add two rules to our earlier form of the improvement cycle. The new form has several names. Dr. Deming himself called it the Shewhart cycle, since Walter Shewhart first saw how to apply scientific method in this new way. Most people call it the PDSA cycle, from the steps described below, others the PDCA cycle. The Japanese call it the Deming wheel. A problem with any name based on initial letters is that it is only meaningful in English. To avoid confusion, we will call it the Deming cycle. These are the two new rules we must apply:

1. **Observations and theories must satisfy the rules of operational meaning.**

2. **Final judgment on the accuracy of prediction must wait until the thing we are predicting (not necessarily the whole system) has reached a state of statistical control.**

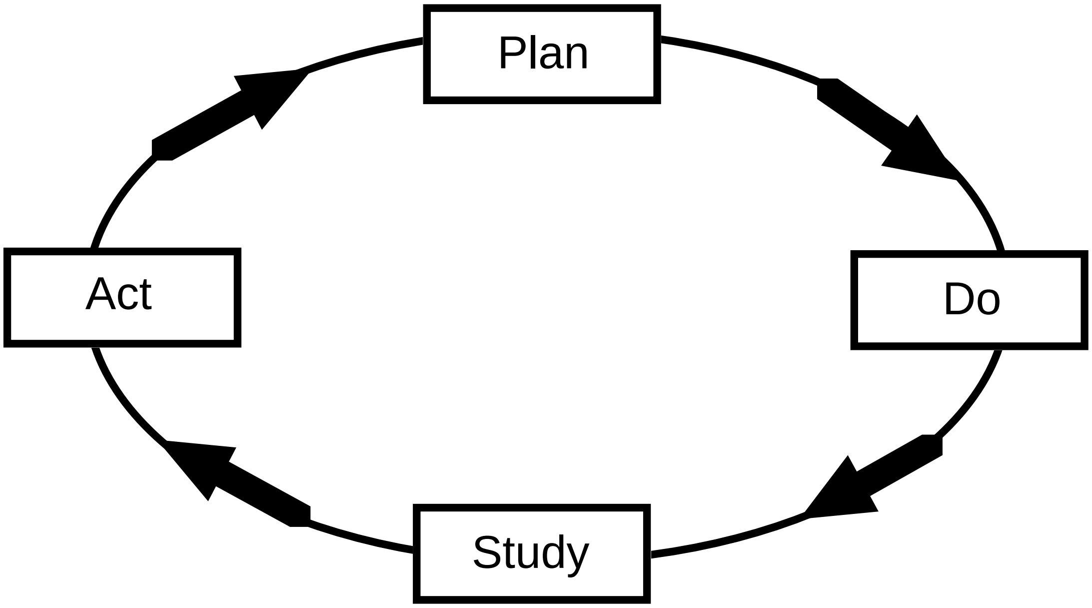
These rules both contribute to stability. Operational meaning makes sure that different people will observe the same thing, and make the same predictions under the same circumstances. This ensures stability from one observer or operator to another. Statistical control ensures that the predictions will be stable over time, and so continue to work in the future.

The value of small changes

The most important changes are innovations, such as new technology, a new product, or a major reorganisation. These are large changes. But when fine-tuning an existing process, we make progress, if possible, in small steps. This is because the system soon returns to a state of statistical control after a small change, but may take a long time to recover from a large one.

Another reason for making changes in small steps, when we can, is that total failure costs money. A pure scientist is delighted when he can push things to extremes, so that a theory fails completely. A manager must usually keep the system working while trying to improve it. We add extra care to each step, and an extra step to the learning cycle. After studying, preferably on a small scale, the effects of a change, and improving our theory, we must apply the results more widely. What works under experimental conditions may not work so well when used routinely. So we study the effect on the whole system before repeating the cycle. The system is now:

Plan:	Decide what to try and predict the changes which will result.
Do:	Carry out the plan, preferably on a small scale.
Study:	Observe the results, and study them carefully.
Act::	Apply the lessons learned to the whole system.



The pure scientist may start with “observe” but the manager is more likely to start with plan and predict. The system is usually working, so predictions are already in use, based on experience, or just a vague hunch. As we make progress, the planning improves, the predictions are stated more precisely, and the observations are collected in a more disciplined way. Having a theory guides us in all these things.

Changing the system

This sounds as if the improvement of management is more difficult than other applications of scientific method, but in many ways it is easier. Pure scientists must not change the nature of the system they are studying. They want to be able to predict how it will work when they are not looking at it.

In management we actually want to change the system. Every change that makes an organisation less chaotic is not only an improvement in itself, but makes the PDSA cycle work better. The same applies to all the other changes brought about by applying the Deming philosophy of management. One of the bonuses we get when we apply the 14 Points in an organisation is it makes learning easier.

The problems of learning are also different because we are dealing with a specific process or system, not a general scientific law. In one way this makes things more difficult, because we can never assume that what works elsewhere will necessarily work in our system. Each system is in some way unique. What is more, we can not assume that what used to work well, will work equally well for ever. Systems change, quite apart from the changes we make deliberately, in the course of continual improvement.

If we are studying a familiar system, with which we have been working for some time, there will be a store of accumulated knowledge. This usually means that there will be plenty of suggestions for improvement.

Many of these will be wrong, but used correctly, even wrong ideas help us to learn.

Distrust of theory

Many people distrust theory. There is good reason for this, because it is seldom used properly. Anyone who learns a theory by rote, and trusts it blindly is asking for trouble. But it is not theory that is wrong, just the way it is used. A car is safe and useful if you know how to drive: if not, it is a public menace. The Deming cycle uses theory safely. It is self-correcting, and gives long-term certainty to the improvement process.

Putting theory to work

Checking and rechecking is an essential part of the learning cycle. This is even more important in management than in pure science. When a pure scientist makes a mistake, scientists all over the world can repeat the experiment and put things right. Inside an organisation we can rely on no one but ourselves to check what we do.

Mistakes will certainly happen. Every change involves some risk, but the discipline of the Deming cycle makes sure that mistakes are small, quickly seen, and rapidly corrected. Without it, the risks are far greater.

Few people combine the enthusiasm for trying new ideas with the self-discipline necessary in the study phase. This is a good reason for the PDSA cycle, when applied to a working process, to be operated by a well balanced team, not a single individual, however talented. Finally, we must always remember that the scientific method guarantees long-term improvement, but not necessarily the improvement we expect.

Learning with the aid of theory makes us see things we would rather not see, and ask questions we would rather not ask. Only if we genuinely want improvement will we dare to use it consistently.

Operational Meaning

To make a new idea achieve a useful result, we must give it operational meaning. This is a definite form that can be communicated, tested, and improved. For example, a chef has a new idea for a cake. Others could understand the general idea, but not produce the same result. The operational form of the idea is a recipe which specifies the ingredients and the way to cook it. It is easy to think of other examples: an engineer plans a new type of car. The blueprints give the idea operational meaning, from which others can build a prototype, test it, and revise the blueprints.

Making ideas operational is a technique that has many uses. Used correctly, it is an important tool of continual improvement. It makes learning and communication much easier, and has a direct effect in reducing variation. Used without understanding, it can put ideas in a straight-jacket and hold up progress.

Operational definitions

The first, and simplest use is to construct operational definitions of concepts such as “time”, or “size”, or “a breakdown”. An operational definition is not so much “true” as useful, for a particular purpose. Its value is that it states or implies the same actions to everyone who uses it. That it “works”, in this sense is a theory that should be tested.

A military unit is planning a night attack. The officer in command gives the order “Synchronise watches.” Everyone then sets their own watch to the time shown by the watch of the leader. This provides the operational definition of time for the duration of the particular sortie. Whether it agrees with the “right” time, determined by the Greenwich Observatory, does not matter. It makes sure that they keep exactly in step with each other.

In the same way there is no “right” definition of a clean table, (to use Dr. Deming’s example), a hot dinner, or a patient’s temperature. An Operational Definition must be fit for a particular purpose, and result in sufficiently close agreement when used by different people, or at different times and places.

Making sure of agreement in this way not only increases clarity, it improves stability. Without it, a process can appear to be out of control when it is really the observation system that is unstable. So “corrective” action is taken which actually makes things worse. Finally, it directly reduces variation.

Does it suit the purpose?

It is easy enough to find out whether a suggested operational definition produces good enough agreement. We can set up an experiment in which, for example, different people observe the same thing. If they record the same result, to the accuracy that matters, the agreement is satisfactory. If not, improve it using the Deming cycle. In the same way, we can check on consistency over time, or from place to place.

By contrast, there is no easy way to decide whether an operational definition represents the original concept sufficiently well. We must usually rely on understanding of the subject matter. Should a patient’s temperature be measured on the skin, under the arm, in the mouth, or in the rectum? It depends on the purpose. If the purpose is not obvious, consider the action to be taken when the result is known.

The ideal way is to give operational meaning to “better for the purpose” would be to see if one suggested operational definition enables us to make more accurate prediction than another. This is far more difficult and time consuming than simply testing whether different users agree. For example, we could compare different rules for classifying children as dyslexic, and see which system more accurately predicts each child’s reading ability in several years time.

Sometimes this kind of investigation is profitable, but less often than we might think. Borderline cases, however defined, will usually be those where the effect is least. Naturally, in the case of engineering blueprints, a slight change in the drawings might make the difference between success and failure. But different rules for making drawings, if everyone understands them, could represent the exact design equally well. So it pays to follow whatever rules are current practice.

Since the gain when any reasonably good operational definition is introduced is so marked, and the difficulty of further refinement by a prediction study so great, other improvements usually take higher priority. Still, it is always wise to look for exceptions.

Formulation of a theory

A more elaborate use of operational meaning arises when we try to turn a suggestion for improvement into a form which can be tested. What often happens is that a group get together, and try to explain why things are going wrong. However convincing the various explanations sound, we must not rely on persuasive arguments, but facts. From a possible explanation, we derive a suggestion for action. Then to test this idea, we must state it as a precise prediction, that if we do A, under conditions B, we will observe C.

To take a simple example, consider the idea that vitamin supplements in pregnancy lead to healthy babies. To test it we must specify which vitamins, in what dose, given for how long, to which group of pregnant women, and how we are to define a healthy baby.

An example from crime prevention

Police forces want to find ways of reducing crime. One possibility is that video cameras scanning the streets would be effective. This may be a good idea, but it is not in a definite enough form to be tested. We must ask a great many questions, which define the ***system*** to be used.

How precisely are the cameras to be used?

How many cameras?

How high above the street?

Are they to be visible, or concealed?

Are the pictures to be viewed live at police head quarters, or recorded on tape or both?

If cameras show a crime, what action is to be taken? Is the idea to send a police patrol at once, or to identify the criminals and prosecute later?

Under what conditions is the system to be tested?

Is it to be tried in city centres, or residential areas?

Does it work only in well-lit streets? (Some types of camera can “see” in the dark)

Which types of crime do we expect to deter? For example drug-dealing, car theft, or violence?

How do we decide if it works?

Do we mean that doing this ***in addition*** to what is done now would reduce crime. Or do we mean that with the limited resources available, video cameras are more effective than spending the same amount in other ways? (More frequent patrols, or better street lighting for example)

How would we know if there is less crime? Do we measure incidents reported, or numbers of convictions, for example?

Is it a success or a failure if the crime simply moves elsewhere?

If the measure chosen shows an improvement, is this due to the publicity surrounding the project, or to the video cameras themselves?

This may sound like a statistician making difficulties. But these questions must be asked. Fortunately few problems in management are as difficult as the crime problem. The point is that ***an idea, in the abstract, can not be tested.*** We can only test the prediction that a well defined system which uses video cameras will under particular circumstances reduce particular measures of crime.

If our test demonstrates an improvement, we have made the idea work. If not, the idea may still be a good one, but must be operationalised in a different way: for example, we could try different numbers of cameras, or use them in different types of area. This is yet another example of the Deming cycle.

From explanation to prediction

When a hunch or an “explanation” has been given operational meaning, it becomes a “theory”, in the technical sense used in discussing scientific method. It can be used to make predictions, and if the predictions are correct, the theory is “accepted”. This does not mean that we believe in it without question, but that we use it until something better is found. The next step is to look for exceptions, so that we know the range of conditions under which it applies. You do not understand a theory unless you know when it will not work.

If an explanation is not expressed in operational terms as a prediction, it does not help us to learn. This is because it can never be proved wrong.

Although the idea of operational meaning is useful and simple, it can cause problems if introduced without sufficient education. Anyone who does not understand why it this care is necessary will think we are just being difficult.

When someone has thought hard about a problem, and come up with a creative suggestion, it seems hard-hearted and obstructive to ask so many questions, and to insist on strict definition and testing. This is particularly true in the prevailing business culture, which mistakes instant, decisive action for leadership. A useful approach is to say “This is such an important idea that it deserves very careful study.” This is both true and tactful, and should avoid hurt feelings.

Improving a Process

There is not one way to improve a process, but many. These are not alternatives. Used with understanding, all contribute to the continual improvement of the particular process, related processes, and the whole system. Each makes other methods more effective, and so they should be used together. To illustrate this, we concentrate on the practical problems of using the Deming Cycle, and show how other actions help it work.

A unified approach

This combined and unified approach to improvement is typical of the Deming Philosophy. Instead of learning separate techniques, and applying them as much as we can, we should take a system view.

This is how we tackle anything complicated. An automobile is simple compared to most processes. If we want it to run well, we do not spend all our time on the electrical systems, and ignore the fuel supply, or concentrate on the tyres and forget the brakes. Of course, if there is a break-down, it probably affects just one part, and naturally we find out which, and work on that first. But for trouble-free motoring, we make sure that all the essential parts are regularly serviced. We do not wait for something to go wrong.

Even this analogy does not go far enough. We can change systems, to make them less likely to go wrong: as if we could continually redesign the automobile to make it perform better.

Ways to improve a process

We express these as actions and questions. Any one action may produce dramatic improvement on its own. For example, improvement in the measurement process, even though it does not directly affect the process, may reduce tampering. More often it is the interaction between these approaches that produces results. What is more, we must see the investigation of this one process as an integral part of the transformation of the whole organisation.

Without this transformation, it is hard to improve an individual process, and the improvement, even if we achieve it, seldom lasts. But working on a process can make some of the ideas of overall transformation more concrete, and fix them in everyone’s minds. The first question is ***always*** “What is your aim?”

1 **Study the customers’ needs. Is the output of our process the most helpful that could be given to them? Is it causing problems in a later process? There is no point in improving a process until you know what a good result really means.**

2 **Flow chart the process. Are there unnecessary stages, or examples of rules 2-4 of the funnel? Have you identified all the internal and external customers and suppliers? Do you listen to them?**

3 **Improve the training of the process operators. Introduce Operational Definitions, and make sure they work.**

4 **Study ways to measure outputs and inputs. What measures are most relevant to success of the process? Check that the measurement processes are under statistical control before attempting to use the measurements to study the process.**

5 **Reduce variability of the inputs. The inputs include every way in which the rest of the system affects the process. Can you reduce the numbers of internal or external suppliers to the process? Do the suppliers understand your process?**

6 **Study the outputs and inputs of the process using control charts. Remove special causes. Eliminate tampering.**

7 **Collect suggestions for improving the process, and test them using the Deming Cycle.**

There are more ways to improve a process, including various special experimental designs. However these seven we have listed are enough to make the point. The Deming Cycle relies on checking the results of such a change, using measurement. When the process itself varies less, and measurements on it are more reliable, it is easier to see the effect of a change. Besides which, the understanding of the process which comes from all these ways of studying it will suggest changes that should be tried. For example, patterns we see in a control chart often suggest useful theories. Both minor changes and major innovations may result from this increased understanding.

What should we do first?

For a process that has not been studied before, the order given above is reasonably good, but the priorities depend on the circumstances. This does not mean that we finish one before going on to the next: we usually do several at the same time.

Even if the process suffers from a major problem which must be “fixed”, never neglect methods which improve the overall system. There is a good reason for this. If the cause of the problem had been obvious, such as something broken, it would have been put right immediately. So we expect the investigation to take time. Occasionally a problem disappears, still unexplained, as part of overall improvement. Almost always the cause is easier to trace when we improve the whole system.

Here are some examples of this kind. One factory of a group had consistently bad results over many years. Every so often a trouble-shooting team went out from head office, found a problem and fixed it: But things were soon just as bad as before. Then control charts were plotted for inputs and outputs, and the process improved without further action. In another case, a long standing problem disappeared after a change to a single supplier.

When we have been studying a process for some time, we will often have a good idea what to try next. But when an inexperienced team faces its first improvement project, we may have either too many ideas, or too few. There will usually be plenty of ideas to choose from, provided everyone understands that learning about the process is more important in the long run than guessing the “right” answer. If a change makes no difference, or even makes things worse, establishing that fact adds to process knowledge, and will eventually help to make things better, or to get an equally good result more easily and cheaply.

We usually test one change at a time (unless we use advanced experimental designs) so if there are too many suggestions, we need ways to choose between them. Here are some useful questions:

1 **Can it be tested on a small scale?**

2 **Will the effect, if there is one, be seen reasonably quickly?**

3 **Will the effect be easy to measure?**

4 **Will the types of measurements on the process which are routinely plotted on control charts be sufficient to show the effect?**

5 **Has the measurement already been studied and shown to be stable?**

6 **Is the test simple and inexpensive?**

7 **Can the test be done without disturbing the ordinary running of the process?**

These questions concentrate on whether a change is easy to try, rather than whether it seems important, or likely to make a big improvement. This is because we do not want an inexperienced team to meet many technical difficulties at its first attempt. The number of “yes” answers gives a crude measure of the ease of applying the Deming Cycle.

It is important to maintain the enthusiasm of the team, and it is encouraging to find a big improvement. So if a change seems likely to be a “winner” in this sense, it is tempting to go for it. Our cautious advice would be to try at least one “easy” project first. In any case, as overall system improvement proceeds, other changes will become easier to investigate. This is one of the important effects of applying the 14 Points.

As the team gains experience, confidence, and understanding of the process, it may be more reasonable to set priorities on the basis of probable benefit, if there is any reasonable way to estimate this. In the long run, a change that reduces variation, without making the average worse, is more desirable than one that improves the average, leaving the variation as great as before. This is because reduced variation almost inevitably means better quality and productivity. It also means lower costs, though these may be “unmeasured and unmeasurable.” Besides all this, it makes other improvements easier to find.

Be systematic

Once a choice has been made, make sure that you do not waste any of the information from the experiment. Keep systematic records of each stage, especially all the things that did not work at the time. Never rely on memory alone: it plays too many tricks.

If an idea does not appear to work, it does not mean that the suggestion should be forgotten. When variation has been reduced by other means, it may be possible to show that the suggestion does work, after all.

Conclusion

Although we must stress the need to define exactly, and to test each new idea thoroughly, we must maintain balance. One of the follies and failures of our education system in the past has been to present Arts and Humanities as the creative subjects, and the Sciences as cold and disciplined. This is nonsense: ask a dancer or musician whether discipline is needed, or a scientist about the excitement of discovery.

All learning comes from a living partnership between creativity and imagination, on the one hand, and patience, caution, and discipline on the other.