Quality Engineering
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597292

Designed Experiments with Fixed and Varying Factors—A Cautionary Tale

Stefan H. Steiner [a]; R. Jock MacKay [a]
[a] Business and Industrial Statistics Research Group, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada

Online Publication Date: 01 October 2009

To cite this Article Steiner, Stefan H. and Jock MacKay, R.(2009)'Designed Experiments with Fixed and Varying Factors—A Cautionary Tale',Quality Engineering,21:4,384 — 391

To link to this Article: DOI: 10.1080/08982110903167817

URL: http://dx.doi.org/10.1080/08982110903167817

Quality Engineering, 21:384–391, 2009

ISSN: 0898-2112 print/1532-4222 online
DOI: 10.1080/08982110903167817

Designed Experiments with Fixed and Varying Factors—A Cautionary Tale

Stefan H. Steiner,
R. Jock MacKay

Business and Industrial Statistics Research Group, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada

ABSTRACT We present a case study to demonstrate some important points when running experiments to reduce variation in an existing process. One lesson is that we should separate the search for an important cause of the variation from the search for a way to deal with this cause. A key point is that we need to think carefully and differently about factors that normally vary and those that do not.

KEYWORDS case study, experiments, variation reduction

INTRODUCTION

This article is an edited version of a consultation about a designed experiment one of us had with a client. Names and company information are fictional to protect privacy and confidentiality.

Our purpose is to highlight some pitfalls in planning and conducting experiments on a high-volume manufacturing process. For those of you who are not completely familiar with the terminology used in describing experiments, we have included a set of numbered footnotes that briefly explain the terms and provide a reference where appropriate. As you read through the conversations and E-mails, we urge you to try and identify what went wrong and why. To maintain the dramatic tension (just kidding!), we postpone our editorial comments to the final section.

THE FIRST CALL

Consultant: Hello

Rick: Hello this is Rick from ABC Automotive. Remember me? Some time ago I took your two-day short course on DOEs (design of experiments). We ran an experiment and are having trouble making sense of the analysis. I'm hoping you may be able to help.

Consultant: Okay, I remember you. Tell me more. What experiment did you run and why?

Rick: In our assembly operation, we have trouble with variation in parking brake tightness. When you pull on the parking brake lever, it clicks until set. Our specification for tightness is five to eight clicks with a lower number of clicks coming from a tight parking brake. In final inspection we often see out-of-specification values with parking brakes that are too loose or too tight. Then we have expensive rework. We ran a full

Address correspondence to R. Jock MacKay, Business and Industrial Statistics Research Group, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, N2L 3G1 Ontario, Canada. E-mail: rjmackay@uwaterloo.ca

factorial experiment[1] with six factors and two levels each to search for a better way to run the parking brake installation process. For each of the 64 assemblies, we measured the number of clicks. We did an analysis of the data but are unsure how to interpret the results. Can you help?

Consultant: I'll try. You've given me a nice summary of the problem. The goal then is to reduce variation in the number of clicks. You said that the current process results in parking brakes that are both too tight and too loose. It is important to quantify the variation. What's the range of values seen in the current process?

Rick: We see parking brakes with between four and ten clicks. Most are in specification, though.

Consultant: How does the tightness vary over time? Do you see periods where all the inspected cars have high click values and then some time later most of the cars have low click values? Or would you see the full extent of the variation within a few cars?

Rick: Based on my observations, it takes a day or so to see cars with click values from four to ten. Consecutive cars typically have similar click values.

Consultant: What about the measurement system for clicks? If the measurement system is adding a lot of variation, it will be very difficult to get any useful information out of the experiment. Have you checked the measurement system?

Rick: Yes, we do regular gage R&R[2] studies on the measurement system. The measurements are repeatable and we used the same operator to measure all the assemblies in the experiment.

Consultant: You said you ran a six-factor experiment. Tell me about the factors and the levels you used.

Rick: Better still, I'll E-mail you a description so you have a written version.

Consultant: Okay, send me details about the experimental plan, the data, and a brief description of what analysis you did, as well. I'll have a look and get back to you this afternoon.

[1]In a factorial design, we have a number of factors, each with two or more predefined levels. In the case, there are six factors (see the table on this page), each with two levels. We then have $2^6 = 64$ possible combinations of the factors and levels. In a full-factorial design, we run all of the possible combinations at least once. In a fractional-factorial design, we use only some of these combinations. See Box et al. (2005) for a complete discussion of factorial experiments.

[2]In a gage repeatability and reproducibility (R&R) study, the goal is to determine what fraction of the overall process variation can be attributed to the measurement system. We can estimate this fraction by repeatedly measuring the same parts or assemblies. If the variation due to the measurement system is large, then we may have real difficulty in isolating important factors in the analysis of the experimental data. For more on R&R studies and measurement system analysis, see Burdick et al. (2005).

E-MAIL FROM RICK

Thanks for your help with this. I have given you the details you asked for below.

****

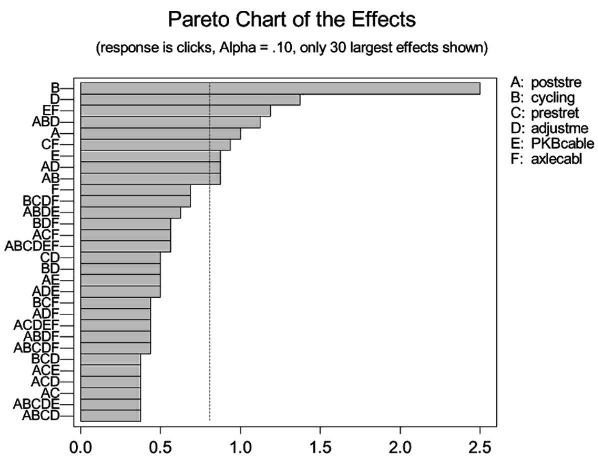
We selected the following factors and levels and ran the 2^6 factorial in standard order. I know you told us to randomize[3] the order but with such a large experiment it was too hard to keep track. The high level for each factor corresponds to the current process setting.

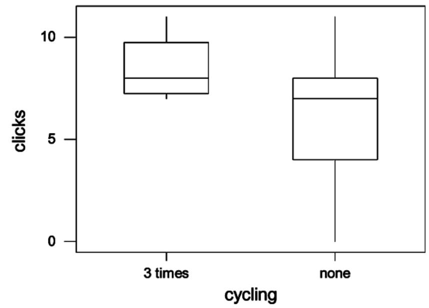
Experimental Factors and Levels

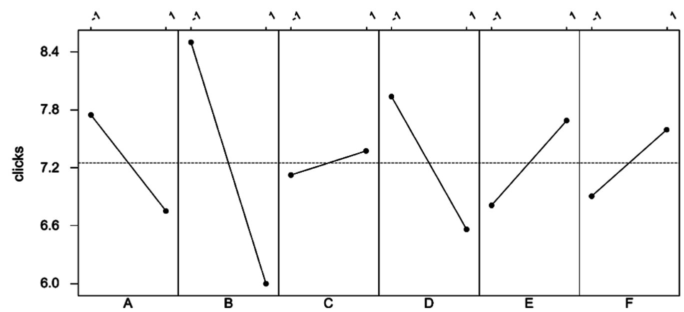
Factor	Label	Low level (−1)	High level (+1) – Current
Post-stretch	A	Yes	No
Cycling	B	Yes	No
Pre stretch	C	Yes	No
Adjust method	D	Method 1	Method 2
PKB cable length	E	Sorted	Current
Axle cable length	F	Sorted	Current

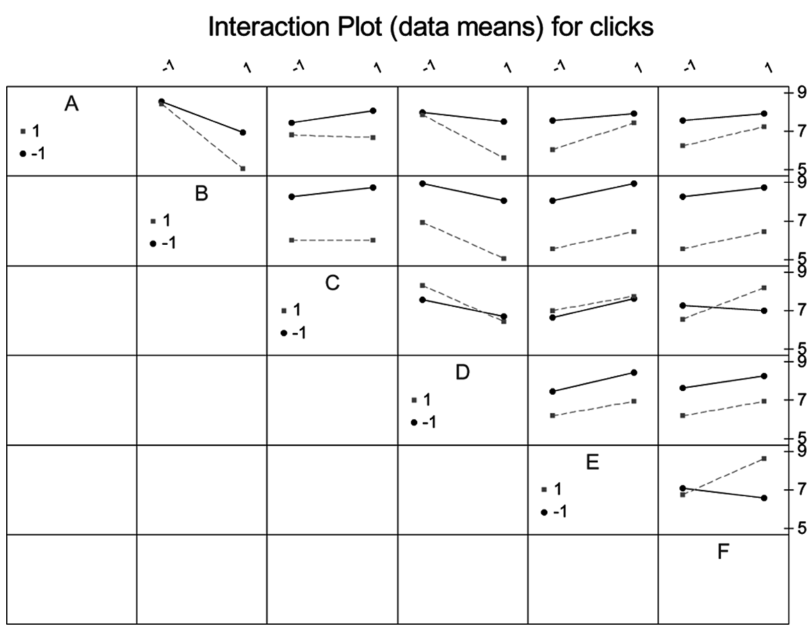
We can change the first four factors (A–D) easily as the process runs. They all involve changes in the parking brake assembly process. The last two factors (E and F) were set by making a special request to our subassembly supplier. For the low levels of E and F, the supplier measured and sorted the two cables used in the parking brake assembly to give us only cables with lengths close to the nominal value. For the high level of E and F, we used a regular batch of subassemblies that had not been sorted.

[3]Randomizing the order of the experimental runs is one of the most valuable and misunderstood statistical concepts. In our case, Rick did not randomize the order and carried out the thirty-two runs with no post-stretch (factor A) followed by the thirty-two runs with post-stretch. Within these two sets of runs, the other five factors were changed in the same systematic way to form a full-factorial design. To assess the effect of post-stretch, we compare the average number of clicks in the two sets of runs. This comparison makes sense because there is exactly one run in each set with exactly the same levels for the other five factors. We call this a *balanced design*. But there are many other unknown changes taking place in the process over the course of the sixty-four runs. Suppose that some unknown factor U that effects tightness is gradually getting larger. Now when we see a difference in the averages and attribute the difference to factor A, we may be mistaken because of the effect of U. This phenomenon is called *confounding*. Randomizing the order of the runs makes it more likely that the levels of U will be close to balanced across the two sets of thirty-two runs and that any difference we see in the averages can be attributed to A. Randomization provides some protection against confounding by unknown factors changing over the course of the experiment.


Pareto Chart of the Effects
(response is clicks, Alpha = .10, only 30 largest effects shown)
A: poststre
B: cycling
C: prestret
D: adjustme
E: PKBcable
F: axlecabl
B
D
EF
ABD
A
CF
E
AD
AB
F
BCDF
ABDE
BDF
ACF
ABCDEF
CD
BD
AE
ADE
BCF
ADF
ACDEF
ABDF
ABCDF
BCD
ACE
ACD
AC
ABCDE
ABCD
0.0 0.5 1.0 1.5 2.0 2.5


clicks
10
5
0
3 times
none
cycling


clicks
8.4
7.8
7.2
6.6
6.0
A
B
C
D
E
F


Interaction Plot (data means) for clicks
A
B
C
D
E
F
1
-1
9
7
5

We used MINITAB to carry out the analysis. See the Pareto chart of the effects and the main and two-way interaction plots.[4] For the largest main effect, cycling, we also show box plots of the data stratified by level. Looking at the box plot, it seems that cycling will really reduce the variation. Can you confirm that this conclusion is correct? (The raw data were also included; we present the data in a rearranged format in the Appendix.)

THE NEXT CALL—A SHORT TIME LATER

Consultant: Hi Rick, I looked over your E-mail and, you know me, I have some additional questions. How did you choose the six factors?

Rick: We used brainstorming and a cause-and-effect diagram. We spent a lot of time narrowing down the possibilities.

Consultant: If I understand your E-mail correctly, the last two factors, PKB cable length and axle cable length, vary in regular production. For the experiment, you got your supplier to measure the cables and create two batches of subassemblies. The sorted batch has cable lengths close to nominal and the current batch has the usual variation. Is that right?

Rick: Yes, you've got it. We originally planned to measure the cable lengths in the subassemblies ourselves but we later realized that we have no way to make the measurements. So we got the supplier to do it.

Consultant: Why did you think that the length of the two cables would matter? Had you done any preliminary investigations?

Rick: That was based on our engineering knowledge. It makes physical sense that the parking brake tightness would be impacted by the cable lengths.

Consultant: True, but you can't be sure their impact is large without some data. The other four factors don't normally change in production. Is that correct?

Rick: Yes—we changed these factors in the experiment because we thought they would affect the tightness.

Consultant: How did you determine the new levels for these four?

Rick: We used our judgment. We were pretty sure we could continue the operation at all the combinations we would see in the experiment.

Consultant: Rather than give you a lecture over the phone, I'll write down my thoughts in an E-mail and then we can talk again. I'll try to get it to you by tomorrow.

CONSULTANT'S E-MAIL

Hi Rick.

Find below a brief set of comments about your experiment and analysis. Sorry to be so negative. Give me a call when you're ready to talk again.

****

1. You told me that the problem was excess variation in tightness. The experiment seems to have two purposes. First, you wanted to verify that the two cable lengths were important causes of variation in the parking brake tightness. Second, and this was probably your main goal, you hoped to find a better installation process that would result in less variation in tightness. Running an experiment with these two purposes is likely not the best strategy. It is better to run two simpler experiments, each planned with a single purpose in mind. Remember in the DOE course, we talked about the diagnostic and remedial journeys.[5] First identify the cause of the problem and then look for a remedy. It looks to me if you tried to find the cause of tightness variation and the remedy within the same study.

   You need to think about the normally fixed inputs (e.g., factors A, B, C and D) and the potential causes of the variation (e.g., factors E and F that varying in normal production) in different ways. To make the process better, you need to change one or more normally fixed inputs. These inputs are under your control. Other than yelling at your supplier, you have no means to influence the variation in cable lengths.

   Here is a summary from the DOE course notes modified to your experiment.

[4]The main effect plots for each factor show the averages over the thirty-two runs with the factor at each of its two levels. Large differences in these averages indicate that changing that factor is likely to shift the process average. The interaction plots for two factors show the averages over sixteen runs for each of the four combinations of levels. Nonparallel lines in the plot indicate that the effect of changing the first factor depends on the level of the second and vice versa. This is called *interaction*.

[5]The ''diagnostic and remedial journey'' recommends for any well-formulated problem that you first isolate the important causes and then look for ways to mitigate the effects of these causes. See Juran and Gryna (1980). This paradigm is the basis for most structured problem-solving systems such as Six Sigma (Breyfogle, 1999) and Statistical Engineering (Steiner and MacKay, 2005).

Excerpt from Design of Experiments Course Notes

Fixed inputs	Varying inputs
Fixed in regular production	Varying in regular production
e.g. Adjust method, product design, target cable length	e.g. Length of cable, operator, plant temperature
Can't be the cause of output variation	Are small or large causes of output variation
Easy to control in an investigation	May be difficult to control in an investigation
Need to change one or more fixed input to make the process better	Can't be set to a single value in regular production

2. Back to identifying the important causes of the tightness variation. Remember in the course, we talked a lot about observational studies[6] to help identify important causes of variation. There are often a lot of clues. You told me that the short-term variation in tightness was small but that it took a day or so to see the full extent of the variation. Do the cable lengths vary in the same manner? You could check with your supplier. Or you could get the supplier to measure the cable lengths and tag the subassemblies so you could keep track. You could plot number of clicks against cable length and see if the cable lengths are important causes of the tightness variation.
3. If we assume for the moment that the cable lengths are the important causes of the tightness variation, you decided to pursue a solution to the excess variation in tightness based on process desensitization.[7] With this approach, you try to find new settings for the factors A, B, C, D that make the process less sensitive to variation in the cable lengths. I guess that you felt it was not possible to reduce variation in the cable lengths directly. Just a reminder that there are other approaches to reducing the variation that we discussed in the course—see the table below from the course notes. You picked approach #2.

Variation Reduction Approaches (Excerpt from DOE Course Notes)

The possible approaches to reduce variation (Steiner and MacKay, 2005) include:
1. fixing the obvious based on knowledge of a dominant cause of variation
2. desensitizing the process to variation in a dominant cause
3. feed-forward control based on a dominant cause
4. feedback control
5. making the process robust to cause variation
6. 100% inspection

4. Choice of levels for the varying inputs. Here is the place that I think you made a serious error. For the low level, the cable lengths are all near to nominal. However, for the high level where there was no sorting, you have no idea of the cable lengths—they could be large, small, or close to nominal. This makes it virtually impossible (with the sample size in the experiment) to see whether cable lengths are important causes of the tightness variation. Instead, you should have gotten your supplier to sort cables so that for the experiment the two levels correspond to short and long cables. Remember, in the course we suggested that you make sure the levels of the varying factors are close to the highest and smallest values seen in normal production. If you look at the data again (given in the Appendix) and supposing you had selected the levels for the cable lengths as I suggested, then the four tightness measurements in each row would correspond to the same four combinations of long and short cables. What you are looking for is a row where there is little variation in the tightness. The levels of the normally fixed factors for this row are a robust solution.
5. What can you learn from this experiment? First, it looks like cycling has an effect on the average tightness. If you cycle, then you can increase the tightness average. But that doesn't help with your problem since you are already getting values that are both too large and too small. You asked about the side-by-side box plots comparing the tightness values when you cycled compared to the normal procedure. First, you can see clearly that cycling increases the average tightness and

[6]An observational study is one in which we observe and measure process inputs and outputs without making any deliberate changes to the process. In an experiment, we intervene and change or control certain inputs.

[7]In the context of reducing variation in high-volume processes, the desensitization and robustness approaches were popularized by Taguchi (1985). See Steiner and MacKay (2005) for explanations and a list of references.

the observed variation in the experiment when you cycled was substantially smaller. You need to ask yourself whether the observed variation fairly reflects what you will see over the long term. In the experiment, you changed other fixed inputs (A, B, D) that do not normally change. For that reason, you may see more variation in the experimental results than you would see in the regular process. In the experimental runs with no cycling, the number of clicks varied from 0 to 11. This is much more variation than you are experiencing in normal production. So I would be very cautious about concluding that you can reduce the variation by cycling.

THE NEXT PHONE CALL

Rick: Hi this is Rick again. Thanks for the E-mail—it seems we really messed up. I've got a few more questions.

Consultant: Let's hear them.

Rick: I can see that we didn't define the levels of the cable lengths properly. But suppose that these lengths are not important. Does it do any harm to include them in the experiment?

Consultant: It added a lot of cost and complexity and cluttered up the experiment. Since you were looking for a robust solution, you need to vary the important causes of variation within the experiment. So if the cable lengths are not these causes, then you didn't have a chance.

Rick: That was another one of my questions. You reminded me that there are other approaches to reducing the variation? Do you think any of these would work here?

Consultant: It's hard to tell without more information, but one idea is feedback control.[8] You told me that consecutive cars have tightness values that are close to one another. Suppose the values started to get large. Do you have any way to adjust the tightness level?

Rick: Actually, come to think of it, we have a way to adjust the number of clicks. To rework any parking brake that is too tight or too loose we adjust the depth of a nut. So if we saw the values getting too high or low at inspection, we could change the nut depth at the installation station.

Consultant: That's the idea—it may not work too well if there is a large number of cars between installation and inspection.

[8]Feedback control is a variation reduction strategy that uses the observed values of the response to signal the need for an adjustment to the process level. If the process output gradually increases or decreases, you can predict future values from the current ones and adjust accordingly. See Box and Luceno (1997).

Rick: So what else might we try?

Consultant: It would be a good idea to first verify that cable lengths really are important causes of the variation.

Rick: Yes, I understand how we can do that. But then what?

Consultant: Well, if cable lengths are important, you could get the supplier to measure the cable lengths and send the measured values with each subassembly. Then you could make an adjustment during installation using the depth of the nut that you mentioned.

Rick: Well that sounds good in theory but not very feasible in practice. Our suppliers would not be pleased. And it would drive the installers crazy.

Consultant: Okay. But speaking of your suppliers, why not help them reduce variation in cable lengths if they turn out to be the important causes.

Rick: That's a better idea. We could teach the suppliers how to run experiments since we are so good at it.

Consultant: I'm glad you haven't lost your sense of humor at least.

Rick: One other question. You keep telling me that we were looking for a robust solution. But then you talked about a desensitization approach in the chart you sent me. I'm a bit confused about these two ideas.

Consultant: Well that's probably my fault. You assumed that the cable lengths were important causes of tightness variation and controlled their levels during the experiment. What you hoped to find was a setting for the other factors A to D that would reduce the effects of the variation in cable lengths. I call this a desensitization approach. Another possibility that you could try is to run a 2^4 experiment using only factors A to D. But now produce enough cars under each combination so that you would expect to see the full range of tightness variation four to ten clicks. You can then calculate the variation (e.g., standard deviation) of the tightness for each of these runs and use that as your response in the experiment. This is a robustness approach.

Rick: But I told you it would take at least a day to see the full extent of tightness variation. Are you suggesting that we run an experiment that would take sixteen days? There's no way that is going to happen!

Consultant: Right, I'm a professor. You wouldn't expect all of my suggestions to be feasible. But if you can show that the cable lengths are important causes of the variation, you could do the desensitization experiment over. That's one reason why identifying the causes is so important. But you'd need to set the levels of the varying factors more carefully next time.

Rick: So what do you think we should do now?

Consultant: Well, you could try adding cycling to the process and see if the variation is reduced. You'd need to use your adjustment nut to get lower tightness values. If it's not expensive to cycle, then this is worth a try. But if you want to do more investigations, I'd start with trying to see whether the cable lengths are important causes of the variation. Once you sort that out, you can decide how to proceed.

Rick: Thanks for all of your help with this—next time I'll call you before we run another experiment.

Consultant: Well, I hope you can solve your problem. And thanks for calling me. You have given me a really good example that I can use in my classes and with my colleagues. Don't worry—I'll make sure to disguise where I got it.

SUMMARY AND LESSONS LEARNED

Despite Rick's excellent intentions, many mistakes were made in both the planning and analysis of this experiment where the primary goal was to reduce the tightness variation.

Mistakes during Planning

- Failing to consider alternatives to a planned experiment to look for a way to solve the problem.
- Not checking that the chosen varying factors were important causes of the tightness variation.
- Choosing the levels for the varying factors incorrectly.
- Failing to randomize the run order in the experiment.

Mistakes during Analysis

- Failing to treat the fixed and varying factors differently.
- Using the average as a performance measure when the goal was to reduce variation.
- Assuming the variation seen in the experiment reflects long term process variation.

To avoid these sorts of mistakes, we make some general suggestions:

1. Use a structured problem-solving method to put the experiment into a context. Using a structured approach, such as DMAIC in Six Sigma (Breyfogle, 1999) or Statistical Engineering (Steiner and MacKay, 2005) helps ensure that each investigation (observational or experimental) has a clear purpose. In most cases we will proceed from the definition of the problem to search for a cause to looking for a solution. If we find an important cause of the variation, we are usually a long way toward finding a solution.
2. Have a single purpose to keep the experiment simple. We recommend separating the search for important causes of the variation from the search for a solution. To avoid cost and disruptions, we can search for important varying inputs using observational investigations before we conduct any experiments. Cause-and-effect diagrams and brainstorming can be useful to identify fixed inputs to use as factors in an experiment designed to look for process improvements. However, they are not the best tools for identifying important varying inputs for an existing process.
3. Distinguish between fixed and varying inputs when planning and analyzing experiments. A fixed input does not change, whereas a varying input changes from unit to unit or time to time in regular production. The purpose of the experiment dictates whether fixed or varying inputs or both should be chosen as factors—see the table below and Steiner and MacKay (2005) for more details. We should choose the low and high level of varying inputs based on the small and large values seen for that input during regular production. For fixed inputs, the levels can be chosen based on engineering judgment. We want to avoid choosing factor levels based on measures of variation.
4. The appropriate analysis of an experiment depends on the purpose. In the example, the goal

Purposes of experimental plan	Type(s) of inputs used as factors
Verify a dominant cause of the variation	Varying
Search for a way to move the process center	Fixed
Investigate solutions with a known cause (desensitization)	Fixed and varying
Investigate solutions with an unknown cause (robustness)	Fixed

was to reduce the *variation* in number of clicks, not change the *average* number of clicks. However, if the output is itself a measure of variation, e.g., out-of-roundness, or a defect rate, reducing variation and decreasing the average are similar. With a defect rate, lower is better and reducing the average rate can be thought of as equivalent to reducing variation.

ACKNOWLEDGEMENTS

We thank the referees and editor, who made some excellent suggestions to improve the article. Both Jock MacKay and Stefan Steiner received support from the National Sciences and Engineering Research Council of Canada. And thanks again to Rick.

ABOUT THE AUTHORS

Stefan Steiner and Jock MacKay are both Associate Proessors in the Department of Statistics and Actuarial Science at the University of Waterloo. They are also active consultants who have worked with organizations from a wide range of sectors including automotive, telecommunications, aerospace, government and more.

REFERENCES

Box, G. E. P., Hunter, J. S., Hunter, W. G. (2005). *Statistics for Experimenters: Design, Innovation, and Discovery,* 2nd ed. New York: John Wiley & Sons.

Box, G. E. P., Luceno, A. (1997). *Statistical Control by Monitoring and Feedback Adjustment*. New York: John Wiley & Sons.

Breyfogle, F. W., III. (1999). *Implementing Six Sigma: Smarter Solutions Using Statistical Methods*. New York: John Wiley & Sons.

Burdick, R. K., Borror, C. M., Montgomery, D. C. (2005). *Design and Analysis of Gauge R&R Studies: Making Decisions with Confidence Intervals in Random and Mixed ANOVA*. ASA-SIAM Series on Statistics and Applied Probability. Philadelphia: SIAM.

Juran, J. M., Gryna, F. M. (1980). *Quality Planning and Analysis,* 2nd ed. New York: McGraw-Hill.

Steiner, S. H., MacKay, R. J. (2005). *Statistical Engineering: An Algorithm for Reducing Variation in Manufacturing Processes*. Milwaukee, WI: Quality Press, American Society for Quality.

Taguchi, G., Wu, Y. (1985). *Introduction to Off-Line Quality Control*. Nagaya, Japan: Central Japan Quality Control Association, Nagaya.

APPENDIX

Reformatted Data from the Parking Brake Experiment

				Number of clicks			
Post-stretch (A)	Cycling (B)	Pre-stretch (C)	Adjustment method (D)	PKB cable current Axle cable current	PKB cable sorted Axle cable current	PKB cable current Axle cable sorted	PKB cable sorted Axle cable sorted
No	No	No	Method 2	3	2	0	3
Yes	No	No	Method 2	8	7	8	6
No	Yes	No	Method 2	10	8	8	7
Yes	Yes	No	Method 2	11	8	7	7
No	No	Yes	Method 2	6	4	4	3
Yes	No	Yes	Method 2	7	5	7	8
No	Yes	Yes	Method 2	10	8	7	7
Yes	Yes	Yes	Method 2	7	7	8	9
No	No	No	Method 1	10	7	8	5
Yes	No	No	Method 1	11	9	2	7
No	Yes	No	Method 1	10	9	8	9
Yes	Yes	No	Method 1	10	8	10	10
No	No	Yes	Method 1	9	3	7	7
Yes	No	Yes	Method 1	7	5	6	8
No	Yes	Yes	Method 1	10	7	9	8
Yes	Yes	Yes	Method 1	9	8	9	9