

CHALMERS
UNIVERSITY OF TECHNOLOGY

TAMPERING, A MAJOR CHALLENGE FOR CDIO BOOSTING

Downloaded from: https://research.chalmers.se, 2026-11-01 14:22 UTC

Citation for the original published paper (version of record):
Hammersberg, P. (2026). TAMPERING, A MAJOR CHALLENGE FOR CDIO BOOSTING. Proceedings of the International CDIO Conference, 2026

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

TAMPERING, A MAJOR CHALLENGE FOR CDIO BOOSTING

Peter Hammersberg

Department of Mechanical Engineering, Chalmers University of Technology,

ABSTRACT

There is a strong overlap between CDIO and Dr. Deming's System of Profound Knowledge, including his 14 points for management, except for one crucial aspect. In CDIO Standard 12 – Program Evaluation - it is a strong drive for monitoring "program's progress toward attaining its goals". Dr. Deming warns that without a clear common understanding of the difference between common cause of variation (random noise) and special cause of variation (signals) using control charts and SPC-thinking (Statistical Process Control) monitoring tend to drive over-reaction to fluctuations, that is, tampering. That in its turn according to Deming, prevents a shift from focus on individual course follow-up to an aggregated dialog on system development, that is CDIO implementation on programme and multi-program level. This study second Dr. Deming's warning. Course quality monitoring data used to trigger standard course development from 451 course evaluation surveys, from 128 courses given in four different programmes on both basic and advance level over 5 years, have been analyzed relative the definitions of common vs. special cause of variation. More than >98% of the variation in the trigger parameter stems from common cause of variation. Therefore, more than 60% of the 41 Action Plan started (9% of all courses) were done so from false alarms. The general strategi when a process is stable and predictable, that is, only varies due to common cause of variation, is to zoom-out to seek explanations to underlying patterns in the noise. In the noisy course survey data interesting patterns on course roles linked to CDIO Standards are identified. It helps shifting from individual course follow-up, driving blame game, to an aggregated dialogue on program development.

KEYWORDS

tampering, course evaluation, CDIO Standard 3.0, SPC, Common vs. Special cause of variation

INTRODUCTION

It seems to be a gap in literature on the overlap between the CDIO Standards 3.0 (CDIO Standards 3.0. 2026) and Dr. Edwards Deming's 14 Points for Management from his System of Profound Knowledge within Quality Engineering. When these two frameworks are lined up, the overlap is surprisingly strong, revealing both connections and important tensions (Table 1). The system of Profound Knowledge as a base for Quality Improvement consists of four blocks of interacting competences: understanding variation, psychology, knowledge theory and systems thinking. Deming's 14 points of management stems from his book Out of the Crisis (Deming WE. 1982[Ed1] & 2018[Ed3]) further clarified by for example Howard and Shelly Gitlow (Gitlow HS., Gitlow SJ. 1987). Recently, in March 2026, the Deming Institute republished the 14 points of management accompanied by short explanatory videos on their YouTube channel (Deming Institute. 2026) with the intention to highlight the actuality and relevance of Dr Demings almost fifty-year-old message.

Theoretical comparison of Deming's 14 points and the CDIO Standards

In Table 1, a pointwise comparison of the connection between the two frameworks is listed. They largely overlap, but there are some critical conflicts that may cause tension. A summary of the comparison reveals the following connections / tensions:

- **Strongly reinforcement**
    a. System thinking. Both frameworks insist that the unit of improvement is the system, that is, the programme, not individual courses
    b. Faculty development as central, not optional, in both frameworks. A core driver for change
    c. Integration over fragmentation. Deming: break silos. CDIO: Integrate curriculum
    d. Continuous improvement. Both rely on feedback loops
- **Crucial conflict**
    a. Measurement & evaluation
        i. CDIO encourages structured evaluation (CDIO Standard 12 – program evaluation)
        ii. Deming warns that *most variation is common cause* and that *reacting to every fluctuation* = *tampering*, that is, overreaction to variation (Deming point 5 – improving the system)

    Implication: *Overreaction to course evaluations when misinterpreting normal variation as problems*
    b. Accountability vs. system responsibility
        i. CDIO focuses on course-level accountability
        ii. Deming says that 90% of problems are system-level, not individual, and that systems cannot be improved from the inside. It requires an aggregated understanding

    Mismatch can undermine collegial trust and hinder constructive dialogue.

The bottom-line of the synthesis is that when CDIO Standards provide the structure (what to implement in education) Deming provides the philosophy of management (how to implement it effectively) and the Key insights of practical importance suggest to:

- Treat course evaluation data using variation analysis (Statistical Process Control SPC)
- Avoid blaming individual instructors
- Focus on program-level system redesign

And not:

- Rank courses or teachers based on raw scores
- React to small fluctuations in evaluation data

Tampering occurs when there is a lack of understanding the difference between common cause of variation (random noise) and special cause of variation (signal). A synthesis of Tampering understanding has recently been explored resulting definition based on Deming's original thinking: "A response to a perceived problem in the form of an action that is not directed at the fundamental" (Smeds M., 2022). Tampering leads to a deterioration of the process or the process output. It often leads to overreaction and that systems get unstable, like when a driver tries to correct a skid with overly large steering inputs. To build feedback control systems that keeps systems stable is normal engineering practice when done from the outside. However, from the inside the teaching system it is another matter. The teaching system at universities are infrastructures that consists of courses on specific topics taught different teachers in several different programmes, which are run by loosely connected teacher teams led by programme directors that interact in semi-overlapping management teams per disciplines, which in their turn are led by the headmaster's team. To aggregate the understanding of the consequences of tampering is not easy, since the normal engineering practices and tools for system control does not function equally well from the inside.

One problem when interpreting the course evaluation surveys is about understanding if the variation from year to year of the course quality monitoring numbers origin in physical changes of the course or if it stems from routine, common cause of variation. It is particularly important since variation below a certain threshold of a specific indicator in the course surveys (Q8<3) initiates a quality improvement procedure (action plan) for course development pinpointing an individual (bad?) teacher. According to Deming and others (Deming WE. (1982Ed1 & 2018Ed3), Deming WE. (2018), Wheeler DJ. (2002 & 2002). Bergman, B., Bäckström, I., Garvare, R., & Klefsjö, B. (2022), Balestracci, D. (2015). Gitlow HS., Gitlow SJ. (1987) standard operation procedure to filter out routine variation (common cause of variation to see signals in operational studies is to analyze a time series the data using control charts from SPC (Statistical Process Control). Control charts were invented 1924 by Walter Shewhart.

This study about noise level in survey responses intend to highlight the tampering phenomena and bring forward some understanding of the consequences of variation and how to handle it, when a process is stable and predictable only suffering from common cause of variation. Within Six Sigma and Design for Six Sigma (DfSS, robust engineering) a stable and predictable process containing only common cause of variation needs to be rebuilt. In Quality Engineering the standard procedure to find underlying patterns in a time series from of a stable predicable process that only contain common cause of variation is to stratify the data using describing attributes, such as technical discipline, educational level, compulsory/elective, etc.

The aim with this contribution is to illustrate the tension between CDIO Standard 12 and the Deming point 5 regarding overreaction to noise (tampering) (Table 1) and elevate the discussion about the effect of lack of common understand of how to distinguish signals from common course variation in the course quality follow-up procedure to the program development level and to help the teacher team to define concepts of mutual interest on course exceeding issues and prevent ghostbusting.

Table 1. Lineup of Demings 14 points vs. CDIO Standards 3.0

Dr Deming's 14 points	CDIO Standards 3.0
1. Constancy of Purpose 2. Adopt a new philosophy (culture change) 3. Cease dependence on inspection 4. End the practice of awarding business on price alone; instead minimize total cost by working with a single supplier (N/A) 5. Continuous improvement of the system 6. Institute training on the job 7. Adopt and Institute leadership 8. Drive out fear 9. Break down barriers between staff areas 10. Eliminate slogans, exhortations, and target for the workforce 11. Eliminate numerical quotas for the the workforce and numerical goals for management 12. Remove barriers that rob people of pride of workmanship, and eliminate the annual rating or merit system 13. Institute a vigorous program of education and self-improvement for everyone 14. Put everyone in the company to work accomplishing the transformation	1. The Context 2. Learning Outcomes 3. Integrated Curriculum 4. Introduction to Engineering 5. Design-Implement Experiences 6. Engineering Learning Workspaces 7. Integrated Learning Experience 8. Active Learning 9. Enhancement of Faculty Competence 10. Enhancement of Faculty Teaching Competence 11. Learning Assessment 12. Program Evaluation

Deming's Points	CDIO Standard	Connection/Tension
System purpose & long-term vision		
Deming Point: 1 → improve the system continuously	CDIO Standards: 1 & 2 → define what kind of engineer is educated	Both require a share, long-term purpose → without this faculty drift into *course-level thinking instead of program coherence*
Adopt a new philosophy (culture change)		
Deming point: 2 → transformation requires changing how people think about work	CDIO Standards: 1 & 8 → shift from lecturing to active, experiential learning	Both require faculty mindset change, not just new tools
Build quality into the process (not inspection)		
Deming point: 3 → quality should be built into the system	CDIO Standards: 5 & 6 → learning is embedded in authentic engineering activities, not just exams	Assessment becomes part of learning, not a separate control step
Continuous improvement of the system		
Deming Point: 5 **→Deming warns against misinterpretation of variation and tampering**	CDIO Standard: 12 →CDIO formalizes course evaluation and program reviews that often relies on r**aw metrics without SPC thinking**	Both emphasizes continuous, data-informed improvement **IMPORTANT TENSION**
Training and education of staff		
Deming Points: 6 & 13	CDIO Standards: 9 & 10	Direct match: both require systematic faculty development → No changed outcomes without investment in teach capability
Leadership vs. supervision		
Deming point: 7 → leaders improve the system, not control people	CDIO Standards: 1, 11 &12 → Program leaders must, align learning outcomes & support faculty	Weak leadership → fragmented CDIO implementation
Drive out fear (psychological safety)		
Deming point: 8	CDIO Standard: 9 - 12 + implicit across CDIO	Faculty must feel safe to discuss poor results, experiment with teaching, and critique programme design →Without this: course evaluations become **defensive tools,** not learning tools and collegial dialogue breaks down
Break down silos		
Deming point: 9 → Departments must collaborate	CDIO Standard: 3 Courses must be explicitly connected across the programme	This is one of the **strongest direct overlaps**
Avoid empty slogans and top-down mandates		
Deming Point: 10 → Deming warns: Slogans without system change create frustration	CDIO Standards: Implementation practice (not a single standard) → CDIO can fail when adopted as a 'label' without any real change	
Eliminate numerical quotas and targets		
Deming points: 11 → Deming warns that metrics can distort behavior if misunderstood	CDIO Standards: 11 & 12 → CDIO uses learning outcomes and evaluation metrics	This is a **critical contradiction** CDIO needs measurements Deming warns about misuse of measurement
Remove barriers to pride in workmanship		
Deming Point: 12	CDIO Standards: 9 & 10	Faculty need time, recognition and autonomy. Otherwise, will CDIO become bureaucratic compliance, not meaningful development
Everyone involved in transformation		
Deming Point: 14	CDIO Standards: All – system-wide implementation	CDIO is inherently a program-level transformation that requires alla faculty participation and shared ownership partial adoption = failure mode in both frameworks

PROBLEM – RISK OF TAMPERING (FALSE ACTION PLAN INITIATION)

Understanding the variation in course evaluation numbers is crucial to know when and when not to initiate quality improvement action plans on the course level. The overall course status across programs is monitored with the mean of course survey question 8 (Q8) (Figure 1). The normal procedure to monitor course quality at Chalmers starts with tabulating course survey responses for the courses completed after each study quarter (Figure 2). The numbers in column Q8 are reviewed. Green if above 3,5 and white cells below. Courses with Q8 numbers above 3,0 are left as is for next round. When Q8 numbers are below 3,0 the examiner is asked to investigate why and make appropriate changes and write an action plan report.

1. I had sufficient prior knowledge to be able to follow the course.
2. The learning outcomes clearly described what I was expected to learn in the course.
3. Learning
   - A. The course structure (division into lectures, exercises, laboratories, simulations, etc.) is appropriate for achieving the course learning outcomes.
   - B. The teaching worked well.
   - C. The course literature (including other course materials) supports learning well.
4. The examination (including all compulsory components, written exams, assignments, etc.) assessed whether I had achieved the course learning outcomes.
5. The course administration (ongoing information, course syllabus/PM, course website, etc.) worked well.
6. The course workload in relation to the number of credits was appropriate.
   - NB: scaled with $6^* = 5 - (\text{Question } 6 - 3)^2$, to obtain a max at 3, see below. Question 6 = 3 → Question $6^* = 5$.
7. Gender Equality, Diversity, and Inclusion
   - A. The course structure, teaching, and examination were designed so that I feel I was able to learn and be assessed on equal terms with other students in the course.
   - B. The course materials, such as literature, presentations, assignments, and examples, were selected and designed with the aim of including all students, i.e., not reinforcing stereotypes or being discriminatory.
8. What is your overall impression of the course?

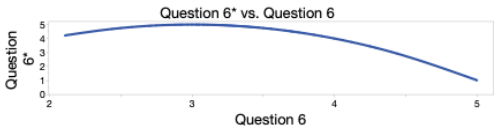
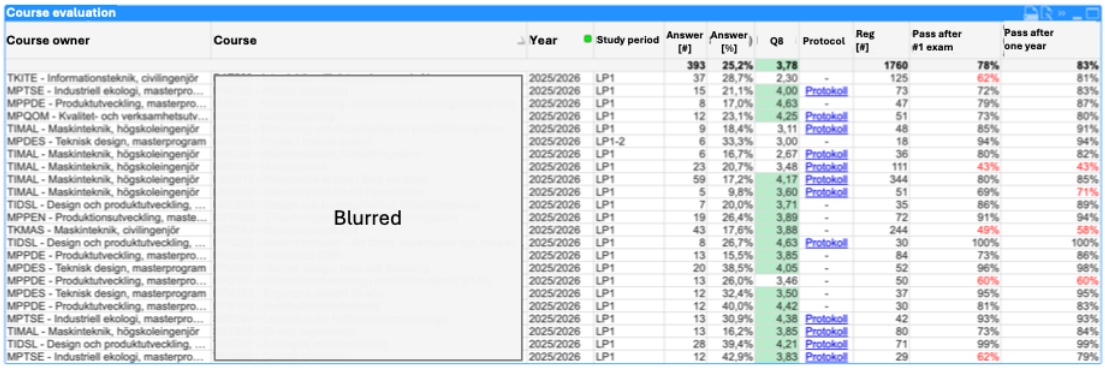


Figure 1. First eight standard questions of the course evaluation survey at Chalmers. Students classify the course they just took on a 1-5 scale for these 8 questions. The higher the better ranking, except on workload Q6 where target is 3. Q6* has been added show Q6=3 to correspond to Q6*=5



Course owner	Course	Year	Study period	Answer [#]	Answer [%]	Q8	Protocol	Reg [#]	Pass after #1 exam	Pass after one year
				393	25,2%	3,78		1760	78%	83%
TKITE - Informationsteknik, civilingenjör		2025/2026	LP1	37	28,7%	2,30	-	125	62%	81%
MPTSE - Industriell ekologi, masterpro...		2025/2026	LP1	15	21,1%	4,00	Protokoll	73	72%	83%
MPPDE - Produktutveckling, masterpro...		2025/2026	LP1	8	17,0%	4,63	-	47	79%	87%
MPQOM - Kvalitet- och verksamhetsutv...		2025/2026	LP1	12	23,1%	4,25	Protokoll	51	73%	80%
TIMAL - Maskinteknik, högskoleingenjör		2025/2026	LP1	9	18,4%	3,11	Protokoll	48	85%	91%
MPDES - Teknisk design, masterprogram		2025/2026	LP1-2	6	33,3%	3,00	-	18	94%	94%
TIMAL - Maskinteknik, högskoleingenjör		2025/2026	LP1	6	16,7%	2,67	Protokoll	36	80%	82%
TIMAL - Maskinteknik, högskoleingenjör		2025/2026	LP1	23	20,7%	3,48	Protokoll	111	43%	43%
TIMAL - Maskinteknik, högskoleingenjör		2025/2026	LP1	59	17,2%	4,17	Protokoll	344	80%	85%
TIMAL - Maskinteknik, högskoleingenjör		2025/2026	LP1	5	9,8%	3,60	Protokoll	51	69%	71%
TIDSL - Design och produktutveckling, ...		2025/2026	LP1	7	20,0%	3,71	-	35	86%	89%
MPPEN - Produktionsutveckling, maste...		2025/2026	LP1	19	26,4%	3,89	-	72	91%	94%
TKMAS - Maskinteknik, civilingenjör		2025/2026	LP1	43	17,6%	3,88	-	244	49%	58%
TIDSL - Design och produktutveckling, ...		2025/2026	LP1	8	26,7%	4,63	Protokoll	30	100%	100%
MPPDE - Produktutveckling, masterpro...		2025/2026	LP1	13	15,5%	3,85	-	84	73%	86%
MPDES - Teknisk design, masterprogram		2025/2026	LP1	20	38,5%	4,05	-	52	96%	98%
MPPDE - Produktutveckling, masterpro...		2025/2026	LP1	13	26,0%	3,46	-	50	60%	60%
MPDES - Teknisk design, masterprogram		2025/2026	LP1	12	32,4%	3,50	-	37	95%	95%
MPPDE - Produktutveckling, masterpro...		2025/2026	LP1	12	40,0%	4,42	-	30	81%	83%
MPTSE - Industriell ekologi, masterpro...		2025/2026	LP1	13	30,9%	4,38	Protokoll	42	93%	93%
TIMAL - Maskinteknik, högskoleingenjör		2025/2026	LP1	13	16,2%	3,85	Protokoll	80	73%	84%
TIDSL - Design och produktutveckling, ...		2025/2026	LP1	28	39,4%	4,21	Protokoll	71	99%	99%
MPTSE - Industriell ekologi, masterpro...		2025/2026	LP1	12	42,9%	3,83	Protokoll	29	62%	79%

Figure 2: Sample of standard course quality monitoring for the courses in first study period academic year 2025/2026.

CONSEQUENCES OF ZOOMING-IN ON SINGLE TABULATED VALUES

Deming (and others) recommend using SPC-thinking to distinguish between common cause of variation and special cause of variation. In Figure 3, the variation of the Q8 number for one of the courses over five academic years is shown using an Individual moving Range control chart (ImR) from the Statistical Process Control (SPC) toolbox. The primary use of control charts is to distinguish between signal and routine variation (noise) and to help avoiding two mistakes in data analysis (Wheeler Donald J. (2002)):

- Mistake One: Interpreting the routine variation of noise as if it amounted to a signal of change in the underlying process, thereby sounding a false alarm
- Mistake Two: Thinking that a signal of change in the underlying process is merely the noise of routine variation, thereby missing a signal

The ImR-chart evolved in the forties from the original control chart invented by Walter Shewhart in 1924. ImR was populated during the fifties and sixties by Dr. W. Edwards Deming and Dr. Donald Wheeler as an aid for managerial follow-up. The theories and techniques are well explored in the general quality literature, for example: 'Quality from customer needs to customer satisfaction' (Bergman, B., Bäckström, I., Garvare, R., & Klefsjö, B. (2022)), 'Understanding variation: The key to managing chaos' (Wheeler, D. J. (2000)), 'Data Sanity' (Balestracci, D. (2015)).

Figure 3 exhibit two panels, the upper shows the time sequence of Q8 over five years. The lower shows the moving range (mR) between two consecutive years in absolute numbers (positive). For this course, Q8 is stable and predictable and will most likely continue to vary with 50% chance to be over or under the average between the Upper and Lower Control Limit (UCL and LCL), respectively, for the years to come. The average moving Range mR is 0.3, which is the base for the control limits calculation using the well-established standard multiplicators 3,27 and 2,66 for mR and Individual observation panels, respectively (red lines). The control chart tells that all variation within the control limits is common cause of variation without plausible explanation. In 50% of the cases Q8 changes more than 0.3 from year to year and can be as large as 0,9 without reason.

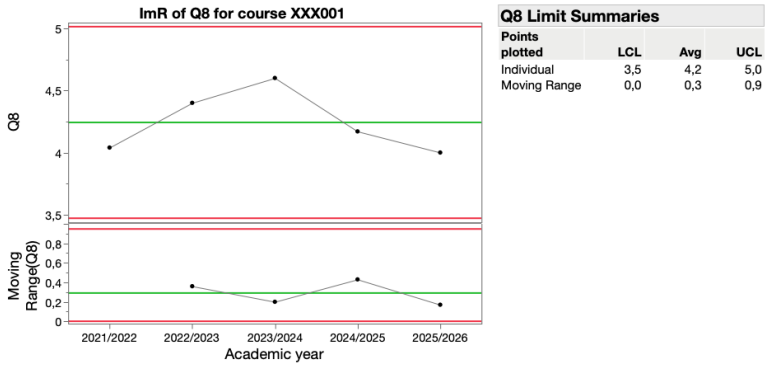


Figure 3: Individual & Moving range chart (ImR) visualizes how Q8 vary for a typical course 2021-2025. It shows no signs of signals, only common cause of variation average mR=0,3 The course quality is stable and predictable and will most likely continue to perform with Q8 on average at 4.2 but can vary from 3.5 to 5 because of common cause of variation.

In Figure 4, four different patterns of the Q8 variation have been connected to the decisions taken on action plan initiation depending on Q8<3 or not. The quarters represent the following:

- Upper left (Yes-Yes): Action plan correctly initiated when five-year average < 3.0 (left) OR single observations outside the control limits AND below 3.0 (right) (*)
- Upper right (No-Yes): Action plan is not initiated (No) even though five-year average of Q8 < 3.0. Slipped aways (▼).
- Lower left (Yes-No): Action plans initiates (Yes) because Q8 < 3, but it is inside control limits and only common cause of variation (No) (▨)
- Lower right (No-No): No action plan initiated AND Q8 & Q8 average are both >3 (•)

The intention with the classification in Figure 4 is to be able calculate the misclassification rate and the Flase-Discovery-Rate FDR as metrics of quality of this judgement procedure.

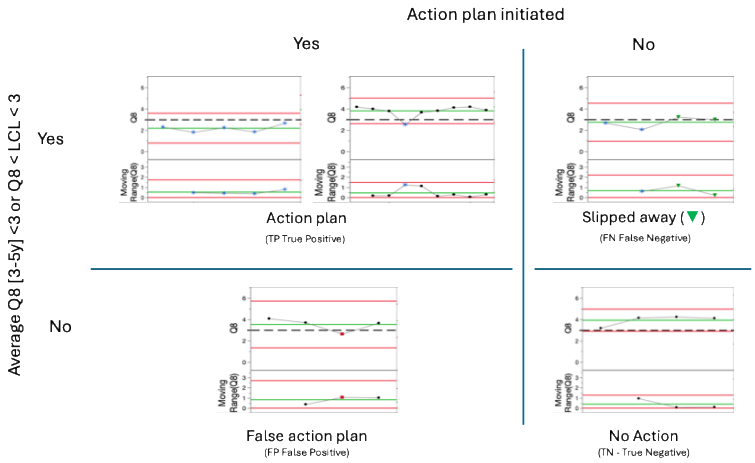


Figure 4: ImR of Q8 scores for five of the courses in relation to the trigger level of Q8=3. Yes/Yes: Action plan initiated correctly ∗, Yes/No: Action plans missed ▼, No/Yes: Action Plans on False Alarms ▧, No/No: Action plan correctly not initiated

Sample Data

Figure 5 display ImR charts of Q8 for 128 courses given in two basic (B.Sc) engineering programs, TIMAL and TIDSL, and two progammes on advance level (M.Sc), MPPDE and MPDES from Mechanical Engineering and Technical Design Engineering, respectively. Compulsory program courses are regularly run every year (five data points) and the elective courses vary between 1 and 10 depending on if students elected them and if they run two times a year (large courses). The threshold for action plan initiation (Q8 = 3,0) is marked with a horizontal black line.

- Mechanical Engineering
  - TIMAL – Mechanical Engineering 3y B.Sc.
  - MPPDE – Product Development +2y M.Sc.
- Technical Design Engineering
  - TIDSL – Design and Product Development 3y B.Sc.
  - MPDES – Technical Design +2y M.Sc.

Even if the time series are short, 6-8 data points is recommended for high precision of the control limits (Wheeler DJ. 2002), the lions' share of the 128 control charts shows a stable and predictable behaviour where most variation is common course of variation. For all courses, except three, the five-year average (green) is above 3,0. Two courses in TIMAL and one in TIDSL with correctly initiated Action Plans (∗). All together have 41 Action Plans (∗ correct + ▧ false alarm) been initiated, which is 9% of all surveys.

Figure 6, visualizes the classification of all 451 surveys according to the definitions in Figure 4. The misclassification rate is defined as the total number of misclassifications in relation to the total number of surveys, that is, 7%. In other words, 1 out of 15 surveys is misclassified, which is 1–2 surveys per study period. Out of the 41 initiated Action Plans initiated 61% (FDR – False Detection Rate) were initiated from common cause of variation down within the noise band, when the average Q8 was > 3, that is False alarms.

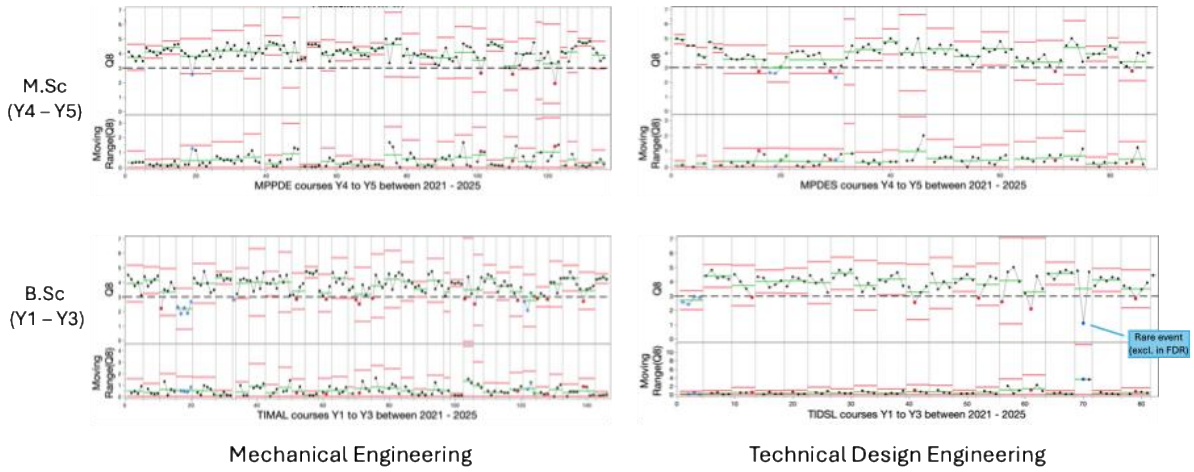


Figure 5: Variation in Q8 over five years for 128 courses in four programmes on basic (B.Sc - below) and advanced level (M.Sc - above) visualized using ImR control chart from SPC-thinking: Mechanical Engineering left and Technical Design Engineering right. Each course has a vertical segment where the upper part shows Q8 and the lower part the corresponding moving range mR. Surveys that initiated Action Plans relative Q8 <3 are coded according to Figure 4.

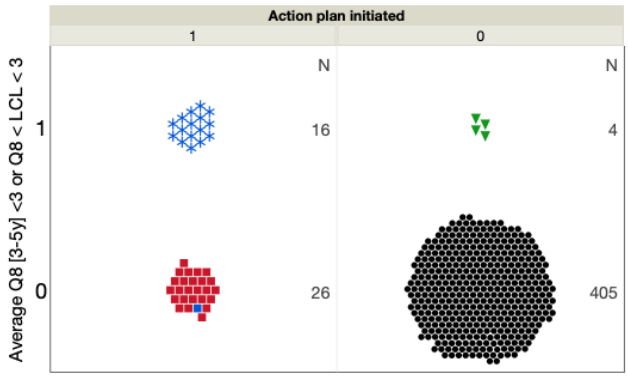


Figure 6: $Misclassification\ Rate = \frac{Number\ of\ Misclassifications}{Total\ Number\ of\ Observations} = \frac{30}{450} \rightarrow 7$ %, that is,1:15, meaning 1-2 surveys per study period are incorrectly classified as Q8<3. The precision, the ratio of false alarms, is defined as: $Precision = \frac{TP}{TP+FP} = \frac{16}{41} = 0{,}39 \rightarrow 39\%$, implying that the False Discovery Rate FDR (1-Precision) is 61%, that is, the risk to falsely initiate an action plan with Q8 < 3 is over 60%. One rare event is excluded from the calculations.

In Figure 7, the effect of the common cause of variation is visualized in the same ImR without the individual course averages. This assumes that the common course of variation in the survey process is the same based on the same level of answer frequency (Figure 8). The average mR across all courses is 0,5. The average moving range (0,5) indicates that annual changes of Q8 up to 1,5 units can be considered common cause of variation (random noise). The control limits in the upper panels +/- 1,2 units indicate the limits for random variation of Q8. More than >98% of the Q8 variation in these 452 surveys is noise.

The Answer frequency in Figure 8 is also a stable and practicable process that vary with average 28% [1,6%, 54,2%] without any reason. In the standard survey process students get a survey-link in the mail after the course is complete and exam done. There are some observations above the upper control limit with survey responses in the range of 60-80%.

These results were obtained when the survey was distributed and filled in during the last lecture, that is, by another process.

The consequence of not considering the routine (random) variation in the Q8 response increase the risk starting action plans based on false alarms. This results in a tampering behaviour that overcompensate for variation, leading to unnecessary changes and frustration among teachers and programme administration.

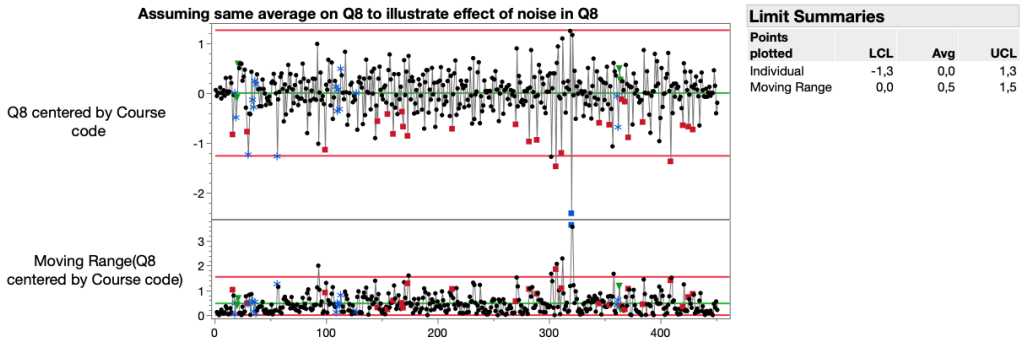


Figure 7. Almost all variation (>98%) of Q8 is within the control limits assuming a grand average moving range (mR) for all courses is 0,5 and the average Q8 level for all individual course is removed. This reflects the decision take from the table in Figure 2. This implies that Q8-scores varies +/- 1,3 without any reason.

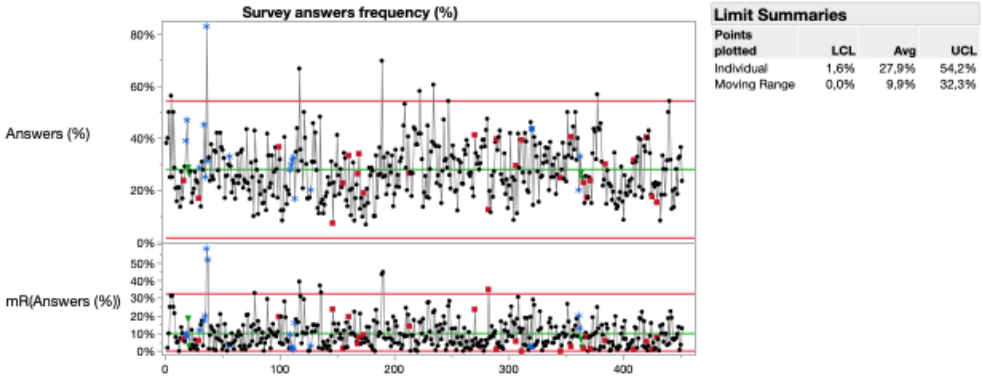


Figure 8. Answer frequency [%]. The standard process of mailing a link to the course survey results in a stable and predictable process. The points outside upper control limit, which are occasions when another survey distribution process been used, in class during final lecture.

PROBLEM ANALYSIS SUMMARY

The risk of falsely initiating action plans based on a single table value that go under a threshold is high. The False-Detection-Rate is 61%, meaning that when Q8 drops below 3.0 there is a larger chance to take the right decision when tossing a coin. The result second Dr Demings warning that managerial data of this kind mainly consist of common cause of variation driving fragmentation. Also, Davis Balestracci second this argument in his book Data Sanity (Balestracci, D. 2015) with cases from the medical field, where time series data of this kind consists of 85% - 90% common cause of variation. Consequence is that zooming in on specific data points tend to drag programme management attention to details driving the blame game making teachers feeling guilty and frustrated since they cannot find a rout cause for the ‘bad’ quality, instead of shifting to an supportive overall understanding of how the programme

function and on system-level. When fitting a normal distribution to Q8 data from Figure 2 it is difficult to say that the two numbers below 3 not is part of the same distribution of common cause of variation as the rest of the numbers from that Study Period $N[\bar{x}, \sigma] = N[3.8, 0.59]$.

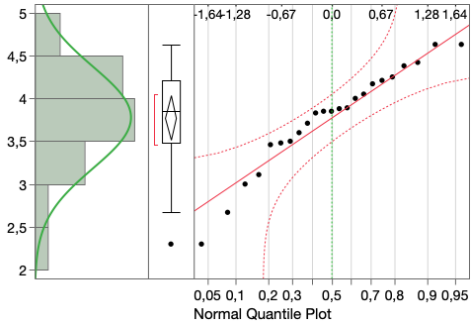


Figure 9: Fit of Normal distribution to the Q8 numbers from Study Period 1 2025 in Figure 2

BEYOND TAMPERING - DECOMPOSING COMMON CAUSE OF VARIATION

When a process only contains common cause of variation (random noise) it is said to be in statistical control and is stable and predictable. The role of control charts within statistical process control is to filter out the random noise to be able to see the signals, that is, changes in the system that has an explanation. As shown with Figure 7, there is not much information in a Q8 number, since it can vary +/- 1,3 from the year before without any reason.

However, just we cannot see any structure in the noise it does not mean there isn't any. Absence of evidence is not evidence of absence, often referred to Carl Sagan or William Wright. In radiography, the challenge is to take good decision about cancer, without increasing the general radiation dose to the society that will increase the risk of cancer. In Quality Engineering a stable and predictable process cannot be improved by fixing it, it needs to be rebuilt to understand and mitigate the influence of common cause of variation. To understand if there exist underlying patterns in the noise the strategy is to zoom-out since individual data-point won't give a clue.

However, zooming out from a single educational program is challenging for the individual PA (programme director). Challenging in several different aspects. The PA-team needs:

- a common language on common cause vs. special cause of variation, that is, signal vs. noise.
- a clear operational definition of what to monitor, how to test on aggregated level and what decisions to take, as in W. Edwards Demings point 5 (Gitlow HS., Gitlow SJ. (1987))
- a common view of the overall system and how to communicate and aggregate systematic variation and structures hidden in the noisy data

Both Deming and Einstein claim that systems cannot be understood and improved from the inside and formulated it, for example, as: "Problems that are created by our current level of thinking can't be solved by that same level of thinking".

The question is to advance the understanding of how to use the course survey results not only to initiate single course improvement (or maybe rather not to initiate action plans), to zoom-out and reinforce a structural programme development dialogue. The question is: Can the

standardised course surveys be used to create a common language and structural program development on a higher hiearachical level from inside the teaching system?

REVEALING STRUCTURES IN THE UNDERLYING SYSTEM OF Q8 SCORES

Stratification using ANOVA on the natural attributes connected to the programmes Figure 10 reveal some minor effects. There is a systematic higher level of Q8 between the level of studies in Mechanical Engineering (+0,28) but not in Technical Design Engineering. Courses seem to be more appreciated in later years compared to the first (+0,45).

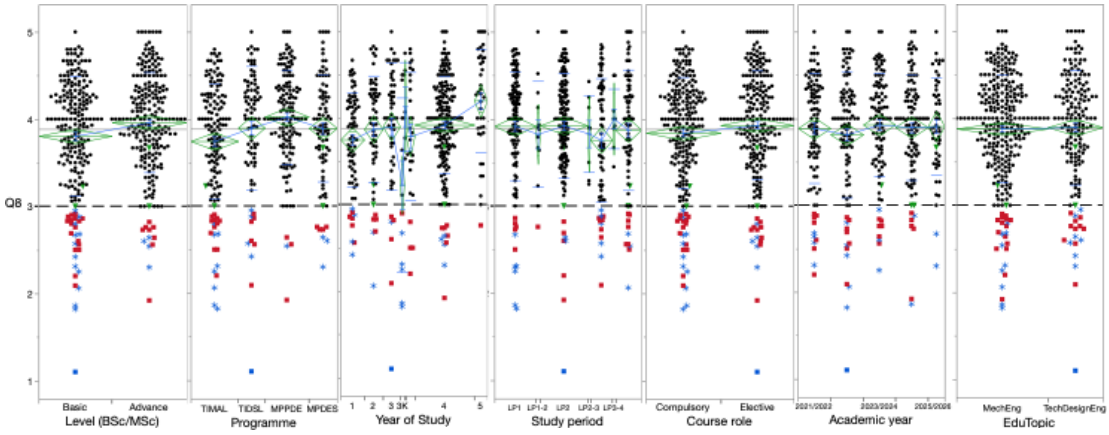


Figure 10: Some minor systematic effects explaining systematic variation in Q8 is revealed using ANOVA. The three to the left have significant shift of the means, whereas the four to right does not. From left M.Sc. programmes have systematically 0,15 higher on Q8 than B.Sc. courses, which interacts with the field of studies. No difference between levels in Technical Design Engineering, whereas in Mechanical Engineering it is. The MP is 0,28 higher on Q8. Third panel: Over time the Q8 response increase 0,45 from year 1 to year 5.

To further explore if there are any underlying patterns. The courses are classified into one of five elevated generic categories/role linked to CDIO Standards aiming to capture the course role in the programmes (Table 2). The sampled programmes all have roughly the same number of course types in each category Figure 11. A stratification on course categories and level of education Figure 12 reveals that the only notable difference is on the course category: Project and Project and applied practice (open projects - applying toolbox in profession). On advance level these courses are systematically given 0.2 more than other course categories and on basic they score 0,7 lower.

RESULT

The course evaluation response of Question Q8 (What is your general impression of the course on a scale 1-5?) vary almost only depending on common cause of variation, that is noise. It can vary +/- 1.3 units from year to year without any explanation. When Q8 drop below 3.0, the current course evaluation process requires that an Action Plan for course development is initiated. Out of 451 surveys from 128 courses in four different programs on basic and advance level, this criterion has led to the initiation of 41 Actions Plans (9%).

When considering the uncertainty in the Q8 number +/- 1.3 units, the misclassification rate is 7%. One out of fifteen surveys are incorrectly classified on the wrong side of the threshold. The FDR, the False-Detection-Rate, is large than 60%.

Decomposing the common cause of variation relative programme characteristic attribute reveal some underlying structure in the noise. Courses later in the programs get systematically ~0,5 units higher score on Q8 than courses in the first year. One underlying explanation of this is that courses with open-ended projects scores 0.9 unit lower on bachelor level compared to master programme level.

Table 2: Course characterization / role vs. CDIO Standards

Course classification/roles	CDIO Standards connection
Applied engineering courses (incl. tailored projects to train specific methods)	Standard 5 (design-implement experiences) Standard 7 (integrated learning experience)
Fundamental theory (programme spine and basic building blocks)	Standard 2 (disciplinary knowledge Standard 3 (integrated curriculum) *Conceive phase*
Methodology and tools (master the toolbox)	Standard 3 (integrated curriculum) Standard 5 (design-implement experiences)
Professional competencies and context (understand professional role, demo-projects)	Standard 2 (personal and interpersonal learning outcomes, Section 2 & 3) Standard 8 (active learning) Standard 9 (faculty competence) *Operate phase*
Project and applied practice (open projects - applying toolbox in profession)	Standard 5 (design-implement experiences) Standard 4 (introduction to engineering) Standard 12 (program evaluation)

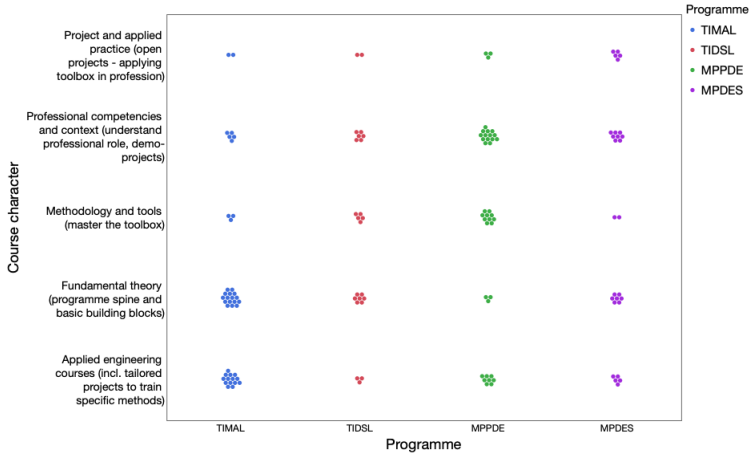


Figure 11. Classification of courses in the sampled programmes in Mechanical Engineering and Technical Design Engineering at Chalmers

DISCUSSION

Details

One plausible explanation for the different score on open-ended problem courses would be that open-ended problem solving in the beginning of the programmes are more challenging for the students where the course structure is much more unclear, in contrast to open-ended projects in the final years that address problems closer connected to their selected exit profile. Since these courses on bachelor level often fall below the Q8-threshold for Action Plan course development, it is relevant to ask if these courses should be judged by the same threshold since they are core courses in the CDIO framework? Handling of open-ended projects with unclear solutions are at the core of the engineer profession and what we intend to educate them to handle.

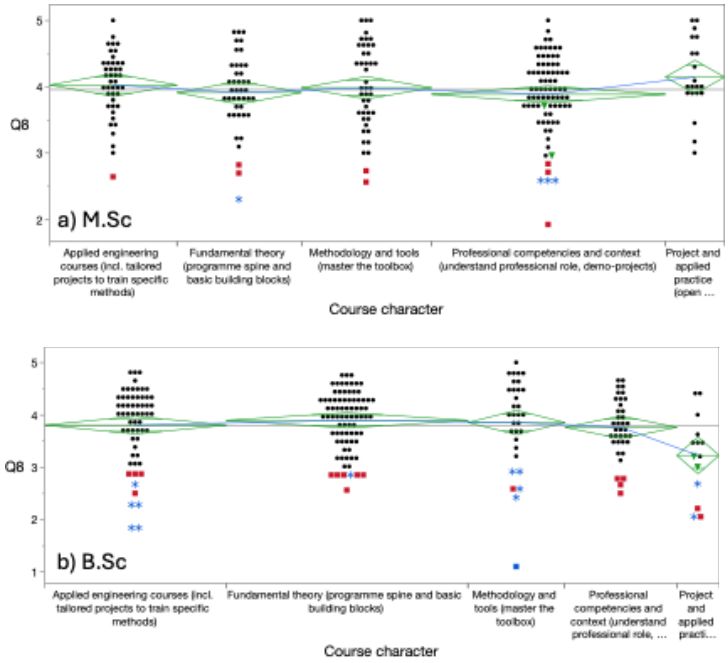


Figure 12: There are no statistically significant difference between the different course characters on advanced level, but a strong tendency that open-ended project courses overall are well accepted among students. On basic level is there a significant difference that open-ended project courses are not well received. a) 3.9 vs. 4.1, that is 0.2 systematically higher on Q8. b) 3.9 vs 3.2, that is 0,7 systematically lower. An overall difference of 0,9.

Intermediate

The high level of random noise in the survey data, visualized in Figure 7, tend to drive over-reaction on course survey response variation, which results in unnecessary initiation of action plans that raise stress levels and frustrations for nothing. In the book 'Data Sanity' (Balestracci, D. (2015)), the author claims that 85-90% of the variation seen in managerial data is common cause of variation and only a small part of the variation has assignable cause. The dilemma according to the author is that most organisations have routines for problem solving for special cause (zooming-in) but lack strategies to understand and reduce the level of common cause of variation that require zooming-out and elevated expansion of the problematic system. Problems tend to be pushed too low in organisations, too late in the chain of events in the lap of the symptom owners that only can handle them reactively instead of making it possible to identify the true root cause of the problem. There is also a built-in lurking driver for programme managers to act rather than not to. This is due to underlying cognitive that trigger tampering including mechanisms, such as fear of the consequences of not acting if not individual feel safe or there is lack of time and room for elevated discussions on system-level (Smeds, M., J. Martin, M. Elg and I. Gremyr. 2023). Avoiding tampering go hand in hand with Deming's point 8 about driving out fear.

An interesting observation of the shift in mind-set was noted during a teachers meeting for the Mechanical Engineering bachelor programme when changing from the normal follow-up visualization in Figure 2 to the aggregated multi-course control chart in Figure 5. Spontaneously, the examiner of the fourth course from the left, who struggles frustratingly with course improvement Action Plans every year without knowing what to do to convince the students said: "Maybe this course is not the right one for this programme". A second reflective comment from another teacher was: "if we are going to use this data for a constructive dialogue on program development, we need a better process to collect student opinions". This

spontaneous shift of mind-set from reactive to proactive when tables and bar charts are shifted to control charts visualizing signals vs. noise have been noticed before by the author and colleagues (Ericson Öberg, A., et. al. 2016).

Overall

The overlap of CDIO and Dr Deming's System of Profound Knowledge is strong, except for one crucial point about basic data quality. This study second Dr. Deming's warnings regarding his point 5: If there is a lack of understanding of the difference between common cause of variation and special cause of variation the shift from individual follow-up, driving blame game, to an aggregated dialog on system development hindered, that is, programme and multi-programme development and management.

CONCLUSION

There is a strong overlap between CDIO Standards 3.0 and Dr. Deming's 14 point for management, except for the one point with tension: CDIO Standard 12 (Program Evaluation) vs Dr Deming's point 5 (Improving the System). The former formalizes course evaluation and program reviews whereas Deming warns against misinterpretation of variation and tampering based on a lack of a common language and joint organisational strategies to differentiate between common cause and special cause of variation. This study second Dr Deming's warning. Almost all variation in the trigger parameter for quality improvement action stems from common cause of variation that results in many actions based on False Alarms, that is, overreaction to random fluctuations – tampering. This tends to hinder a shift of joint mindset from course details to program function (system level) that needs a systematic approach to detect underlying patterns in the noisy data.

ACKNOWLEDGEMENTS

Many thanks to Programme Directors Lecturer Jonas Tuveson (Technical Design Engineering) and Senior Lecturer Dr. Andreas Dagman (Product Development) for course classification, Professor Emeritus Bo Bergman (Quality Engineering) and Martin Gibson (Director Process Excellence) for reviewing manuscript. The author received no financial support for this work.

REFERENCES

CDIO Standards 3.0. (2026). https://www.cdio.org/content/cdio-standards-30

Deming WE. (1982[Ed1] & 2018[Ed3]). Out of the Crisis. Cambridge. ISBN 0-521-30553-5

Deming WE. (2018). The New Economics 3ed. The MIT Press. ISBN 9780262535939

Deming Institute (2026). Dr. Demings 14 Points for Management. https://deming.org/explore/fourteen-points?utm_source=chatgpt.com and https://www.youtube.com/user/deminginstitute

Wheeler DJ. (2002). Twenty things you need to know. SPC Press. ISBN 978-0945320685

Bergman, B., Bäckström, I., Garvare, R., & Klefsjö, B. (2022). Quality from customer needs to customer satisfaction (4th ed.). Studentlitteratur AB. ISBN 978-9144140261

Wheeler, DJ. (2000). Understanding variation: The key to managing chaos (2nd ed.). SPC Press. ISBN-13: 978-0945320531

Balestracci, D. (2015). Data Santity: a quantum leap to unprecedented results – second edition. Medical Group Mamagement Association MGMA.ISBN: 978-11568294384

Gitlow HS., Gitlow SJ. (1987). The Deming Guide to Quality and Competitive Position. Prentice Hall, Englewood Cliffs, NJ 07632, ISBN 0-13-198441-1

Smeds M. (2022), Deming's tampering revisited: definition and future research agenda, Int. J. of Quality and Service Sciences, Vol. 14 No. 5. Pp. 37-51, Emeral Publishing Limited, http://dx.doi.org/10.1108/IJQSS-03-2021-0041

Smeds, M., J. Martin, M. Elg and I. Gremyr (2023). Why won't you leave the process alone? Exploring emotional, motivational and cognitive mechanisms triggering tampering, Total Quality Management, Vol. 34, No. 6, 791-805, https://doi.org/10.1080/14783363.2022.2112514

Ericson Öberg, A., S. Braunias, P. Hammersberg, C. Andersson (2016), Changing from watermelon measures to real decision support: including information about variation in performance measurements, 5th World Production and Operations Management Conference, P&OM Havana

BIOGRAPHICAL INFORMATION

Peter Hammersberg is senior teaching fellow in Material Engineering and Quality Engineering at The Dept. of Mechanical Engineering at Chalmers University of Technology Sweden. He holds an MSc degree in Applied Physics and Electrical Engineering and a PhD degree in Engineering Materials. He also holds an industrial certified degree as Six Sigma Master Black Belt and trained about 600 students and practitioners as Lean Six Sigma Black Belts. Peter is particularly interested in training and coaching students to handle open-ended questions in real engineering challenges bridging academia and industry/healthcare.

Corresponding author

Peter Hammersberg
Dept. of Mechanical Engineering, Chalmers University of Technology
Rännvägen 2, SE-412 96 Göteborg, SWEDEN
peter.hammersberg@chalmers.se



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.