2019

Understanding and Evaluating Teaching Effectiveness in the UK Higher Education Sector using Experimental Design: A Case Study

Antony, J

http://hdl.handle.net/10026.1/11235

10.1108/IJQRM-01-2018-0011
International Journal of Quality and Reliability Management
Emerald

International Journal of Quality & Reliability Management

Understanding and Evaluating Teaching Effectiveness in the UK Higher Education Sector using Experimental Design: A Case Study

Journal:	*International Journal of Quality & Reliability Management*
Manuscript ID	IJQRM-01-2018-0011.R1
Manuscript Type:	Quality Paper
Keywords:	Experimental Design, Design of Experiments, Teaching Effectiveness, Case study, Higher Education
Abstract:	

SCHOLARONE™ Manuscripts

Understanding and Evaluating Teaching Effectiveness in the UK Higher Education Sector using Experimental Design: A Case Study

Abstract

Purpose – The purpose of the paper is to demonstrate the power of experimental design as a technique to understand and evaluate the most important factors which influence teaching effectiveness for a postgraduate course in a Higher Education (HE) context.

Design/methodology/approach – The methodology involves the execution of a case study in the form of an experiment in a business school setting. The experiment was carried out with the assistance of over 100 postgraduate students from 26 countries. The data was collected over a 2 year period (2015 and 2016) from a postgraduate course offered by the same tutor for repeatability reasons.

Findings- The key findings of the experiment have clearly indicated that students' perceptions of teaching effectiveness based on intuition and guesswork are not identical to the outcomes from a simple designed experiment. Moreover, the results of the experiment provided a greater stimulus for the wider applications of the technique to other processes across the case study higher education sector.

Research limitations – One of the limitations of the study is that the experiment was conducted for a popular post graduate course. It would be beneficial to understand the results of the experiment for less popular post graduate courses in the university in order to drive improvements. Moreover, this research was conducted only for post-graduate courses and the results may vary for undergraduate courses. This would be an interesting study to understand the differences in the factors between undergraduate and postgraduate teaching effectiveness.

Practical implications - The outcome of this experiment would help everyone who is involved in teaching to understand the factors and their influences to improve students' satisfaction scores during the delivery of teaching.

Originality/value – This paper shows how experimental design as a pure manufacturing technique can be extended to a higher education setting.

Keywords: *Teaching Effectiveness Higher Education Experimental Design Post-graduate course Case study*

Introduction

Design of experiments or experimental design has been shown to be one of the most powerful techniques for process optimisation (Antony 2014) and has been widely used for improving yield, capability, and performance of various manufacturing processes for several decades (Montgomery 2012). Research has shown that the application of this powerful technique for service processes or non-manufacturing settings is very limited and only a handful number of articles are found in the existing literature (Antony et al. 2011). Experimental design enables researchers and experimenters to understand the effect of several process parameters or variables simultaneously in a minimum number of trials or runs leading to an increased understanding of the process. It is a direct replacement of the traditional "trial and error" approach to experimentation, which depends upon intuition, guesswork, and luck and still does not guarantee success (Antony 1998).

Although experimental design has been around for decades, few business leaders in service organizations have a good grasp of its power in tackling problems associated with service process efficiency and effectiveness (Johnson and Bell 2009). Customers in the service sector and even many public sector organisations are becoming more critical of the quality of service they receive today (Cudney, Furterer, and Dietrich 2013; Antony, Rodgers, and Cudney 2017). It therefore becomes imperative that the most critical factors or variables that affect service quality are evaluated regularly in order to satisfy end users and internal customers effectively. The use of experimental design eliminates the uncertainty involved in determining the critical factors, thereby ensuring reliability and validity.

Relatively few applications of experimental design in a service setting or environment have appeared in the academic literature (Ledolter and Swersey 2006; Blosch and Antony 1999; Kumar, Motwani, and Otero 1996). The next section presents a review of literature on the use of experimental design methods in the context of non-manufacturing scenarios with a greater focus on higher education followed by a review of the literature on teaching effectiveness. The paper then presents some of the challenges on the implementation and illustrates a case study on evaluating the teaching effectiveness for a postgraduate course within the UK higher education sector. In addition, some of the limitations and practical implications of the study are presented followed by future research directions for this study.

Literature Review

Due to the breadth of the research, the literature review focuses on experimental design in service, experimental design in higher education, and teaching effectiveness in higher education. While methodologies such as six sigma (LeMahieu, Nordstrum, and Cudney 2017; Cudney et al. 2014; Cudney and Kanigolla 2014; Kanigolla et al. 2014a), lean (Cudney et al. 2014; Kanigolla et al. 2014b) and quality function deployment (Ezzell, 2015) have been widely used in higher education for continuous improvement, little research employs experimental design.

Experimental Design in Service

Although design of experiments is widely established within the manufacturing sector, the application of the method in non-manufacturing processes is still in its early stages (Antony 2014). Holland and Cravens (1973) presented the essential features of fractional factorial design and illustrated a very interesting example looking into the effect of advertising and other critical factors on the sales of candy bars. Ledolter and Swersey (2006) described the power of a fractional factorial experiment to increase the subscriptions response rate of Mother Jones magazine. Anderson (2009) compiled a number of excellent examples regarding the applications of experimental design in the service environment, which include: identifying the service design parameters that influence the service quality characteristics or CTQs in the eyes of customers; identifying the key service process or system variables that influence the process or system performance; minimizing the time to respond to customer complaints; minimizing errors on service orders; reducing the service delivery time to customers (e.g.: banks, restaurants, etc.) and reducing the turn-around time in producing reports to patients in a healthcare environment, among others.

Kumar, Motwani, and Otero (1996) used a Taguchi robust parameter design methodology in order to improve the response-time performance of an information group operation that was responsible for addressing customer complaints in a small software export company. Blosch and Antony (1999) demonstrated the use of computer simulation and experimental design technique to identify the key risk variables within the manpower planning system at the UK's Royal Navy. This combined approach has provided a greater understanding of the manpower planning system, especially in terms of reducing gapping (a gap occurs when a particular job or task is not being filled by a competent and qualified person) at sea.

Gliatis, Minis, and Myrto (2013) applied experimental design technique in combination with a simulation to investigate the impact of failures on the operational variability of key performance measurements in an operational service process. He et al. (1997) reported on the application of experimental design in the field of software testing. Experimental trials were conducted to detect errors in software. The researchers argued that experimental design can reduce testing time while still detecting as many errors as possible. Besseris (2011) applied experimental design on a maritime vessel in the logistics sector. An experiment was performed to investigate the best combination to maximise vessel speed by simultaneously minimising fuel consumption and exhausted gas temperature.

Experimental Design in Higher Education

Relatively few papers have been published on the use of experimental design methodology applied in the higher education environment; which clearly indicates a research gap and more potential opportunities for its applications in various business processes within the higher education sector. Barone and Lo France (2009) undertook an experimental design approach in combination with the service quality model in an environmental engineering degree program at the University of Palermo, Italy. This study was conducted over a period of two years where data was gathered from 24 students attending two academic statistical courses. The research found that teacher-student interaction is the most influential factor on student satisfaction.

Antony, Sivanathan, and Gijo (2014) performed an experiment with three factors at two levels to evaluate the factors which may influence the quality of an undergraduate and postgraduate course. The three factors included: the presentation of the content, number of speakers, and the time when the class was delivered. The study showed that delivery time of the class and presentation of content are the most significant factors for both undergraduate and postgraduate courses. Further, while undergraduates preferred afternoon classes, postgraduates found morning classes more productive.

Ree, Park, and Yoo (2014) presented a case study on experimental design to improve teaching quality at a university in Seoul, South Korea. The study quantitatively and qualitatively analyzed the factors that affect lecture quality, and selected two control factors that can be controlled by the lecturer and three noise factors that cannot be controlled by the lecturer. The result was analysed to propose the optimum lecturing method. The analysis showed that it is more effective when the effort to form closeness with students is carried out at the beginning of the lecture, and when student presentations are held once every two months.

Teaching Effectiveness in Higher Education

The use of various instruments for measuring teachers and their performance led to a harsh debate for the appropriateness of those instruments (Clayson 1999). That debate is fuelled by the lack of a clear definition of teaching effectiveness (Ding and Sherman 2006). According to Ding and Sherman (2006), the concepts of teaching effects and teaching effectiveness are distinctively different. For example, a teacher with a deep knowledge of their field might be an inefficient teacher. The authors are of the opinion that teaching effectiveness is about 'doing the right thing' both for the students and for the teacher. Therefore, an effective teacher is better than an efficient one, who is only doing the process of teaching in the right way and not doing the things that are right for the students and teacher alike.

In our research, we follow the definition provided by Seidel and Shavelson (2007, 456) "We speak of teaching effects or teaching effectiveness when referring to the effects of teaching on student learning and how satisfied students are from their learning experience". The rationale behind the use of this definition lies with the fact that this definition relates teaching effectiveness to students' learning experience. In our research, the authors would like to understand the factors which influence teaching effectiveness from a learning perspective. According to Seidel and Shavelson (2007), studies conducted in the past decade related to teaching effectiveness were dominated by correlational survey studies, but they were proven different from the teaching–learning process.

"Differences in teacher effectiveness is the single largest factor affecting academic growth of populations of students" (Sanders 2000, 8). Thus, according to Sanders (2000), teacher effects have a larger impact than class size effects, spending differences, and several other factors believed to impact student learning. Teaching methods have also been found to have a direct impact on student motivation (Ezzell and Cudney 2017). In another study, Marsh and Hattie (2002) used 'student perception' of teaching as the measure of teaching effectiveness, rather than an assessment that attempts to directly measure student-learning outcomes. We are using a similar approach in the case study and a scientific experiment is executed purely based on students' perception on teaching effectiveness.

McKeachie (1997, 385) defines teaching effectiveness as "the degree to which one has facilitated student achievement of education goals". Marsh and Hattie (2002) measure teaching effectiveness by the overall ratings of the teacher and the value of the course, while Galbraith and Merrill (2012) measure teaching effectiveness by the learning outcomes. The lack of a universally acceptable model of 'good teaching' causes poor measures of teaching effectiveness (Hinton, 1993). Teaching effectiveness obviously is a highly complex and very personal process of evaluation that includes a multitude of variables (Galbraith and Merrill, 2012). The quality of teaching effectiveness has been reported to have a direct relationship with the student learning outcomes (Darling-Hammond and Young, 2002).

Several factors affect teaching effectiveness according to various academics. However, according to Ding and Sherman (2006) and Kupermintz, Lorrie, and Robert (2002) there is a strong relationship between the teaching effectiveness and the teacher effect. Thus, the factors mentioned below according to McBer (2000) affect the teacher effectiveness and could also affect to a certain degree teaching effectiveness. Teacher effectiveness could be affected by teacher characterstics, which include professionalism (e.g., commitment, confidence, trustworthiness, and respect), thinking/reasoning (e.g., analytical thinking and conceptual thinking), expectations (e.g., drive for improvement, information seeking, and initiative), and leadership (e.g., flexibility, accountability, and passion for learning). Some additional elements that affect teacher effectiveness, according to Rowan, Chiang, and Miller (1997) include content knowledge, teaching practices, classroom management skills, motivation, and classroom context.

Clayson (1999) examined how teacher effectiveness is affected over time. Three categories were identified, factors with strong change over time, factors with positive change over time, and factors with negative change over time. The factors and their classifications are described in Table 1.

[Table 1 near here]

Case Study

A case study was conducted for two purposes (1) to remove the myth that experimental design is confined to improvement of process performance in manufacturing settings, and (2) to demonstrate the power of experimental design in a higher education context. The case study encompasses the delivery of a postgraduate course to postgraduate students from 26 countries at a higher education institution in the UK. The case study was carried out in four different phases: (1) planning the experiment, (2) designing the experimental layout, (3) conducting the experiment, and (4) analyzing the experiment.

Phase 1: Planning the experiment

In the planning phase, the students were asked to define teaching effectiveness in their own perspective. The study was carried out in two successive years (2015 and 2016) and included over 100 postgraduate students. The students were put in groups, with each group consisting of no more than eight students. From the students' perspective, it was apparent that there were two components that constitute teaching effectiveness. The first component was the content of the course taught by the tutor or instructor has to be practical and can be readily applied in a business context. The second component was the course material can be easily understood and can be learned efficiently and effectively. The tutor of the class has 20 years of experience with the course content and 18 years of industrial experience on the topic. The tutor asked the students to identify the potential factors or process variables that could influence teaching effectiveness. In this study, the response or quality characteristic of interest is the teaching effectiveness. For simplicity reasons, the students were asked to keep each factor at two levels. This assumes linearity for each factor and the definitions of each level for each factor were determined by the students in groups. Further, the tutor was involved in guiding the students in developing a definition that was agreeable with everyone in the classroom. A thorough brainstorming session was conducted for an hour, during which the students initially identified 20 potential factors and their levels as defined in Table 2.

[Table 2 near here]

The student groups were then asked to utilize multi-voting to reduce the number of factors to a manageable number. All students participated in this exercise and identified the top 11 factors in order for a screening design to be utilised to further identify the most important factors from the study. Table 3 presents the list of factors that were included in the screening experiment. In order to study 11 factors at 2-levels, a non-geometric Plackett-Burman 12-trial design was utilized.

[Table 3 near here]

There were 14 groups all together for years 2015 and 2016. Each group consisted of 7 to 8 students. Each group was asked to rate teaching effectiveness on a Likert scale of 1 to 10; 1 being the lowest score and 10 being the maximum possible score. The average teaching effectiveness for year 2015 for all possible combinations of factors were recorded and similarly the same exercise was repeated in 2016 for repeatability. The data collection and the experimental layout is discussed in the next phase.

Phase 2: Designing the experimental layout

In this phase, the design of the experimental layout was developed for experimental trials based on the various combinations of factors selected from the brainstorming session in Phase 1. In this phase, a Plackett-Burman screening design with 12 trials was employed (Antony 2014). In Plackett-Burman designs, the main effects have a complicated confounding relationship with 2-factor interactions. Therefore, these designs should be used to study the main effects only and should not be used when strong interactions are to be studied or analysed in an experiment. Plackett-Burman designs are very powerful in identifying the most important factors in a minimum number of experimental runs or trials. For instance, if a full factorial design is utilized for studying 11 factors at 2-levels, the total number of experimental trials or runs would be 2^{11} or 2048 runs. This large number of trials would be time consuming in our investigation and was, therefore, determined not feasible to execute. Table 4 presents the design matrix or experimental layout in coded format (showing all the factor levels in coded form). The low level of each factor in the experimental layout was replaced by “-1” and high level by “+1”.

[Table 4 near here]

Phase 3: Conducting the experiment

Data was collected during the second semester of 2015 and second semester of 2016 from over 100 post-graduate students representing 26 countries. The same factors were studied using the Plackett-Burman design for repeatability purposes and the average teaching effectiveness for both years were recorded by the course tutor, which resulted in 24 data points from 12 experimental trials. This allows an experimenter to create enough degrees of freedom to work out the experimental error or error variance. The average scores for teaching effectiveness for years 2015 and 2016 were entered into the last column of Table 5, which indicated differences in the average scores between the years at each trial condition.

[Table 5 near here]

The next phase focuses on analysing the results. Therefore, the key objectives from the analyses were discussed with the students. The following objectives were set by the tutor in order to conduct statistical analysis on the collected data.

1. What are the most important factors (from a statistical perspective) that influence the average teaching effectiveness?
2. What are the least important factors that influence the average teaching effectiveness?
3. What are the best settings of the factors to maximize teaching effectiveness?

Phase 4: Analysing the experiment

The purpose of this phase is to analyse the data in the experimental layout and interpret the results in order to derive valid and sound conclusions. Minitab software system version 17 was used for the analysis of data and a number of graphical tools were utilized to validate the results of the experiment. The first task in the analysis was to identify the most important factors that have an impact on the average teaching effectiveness scores. Quite often, people pay too much attention to the most important factors from an experiment. In fact, it is equally important for experimenters to understand the least important factors that influence the quality characteristics of a product or service. The goal is to set the least important factors at their most economical levels for cost savings. Figures 1 and 2 show the Pareto plot and half-normal probability plot of the main effects of the factors that influence the average teaching effectiveness.

[Figure 1 near here]

The Pareto plot displays the absolute values of the effects and draws a reference line on the chart. Any effect that extends past the reference line is statistically significant at a 5% significance level (Antony 2014). The significance level is the risk of stating that a factor is significant when, in fact, it is not. In other words, it is the probability of the observed significant effect due to pure chance. The findings from the Pareto plot were further confirmed with a half-normal probability plot (HNPP) of the estimates of the effects. A HNPP provides the absolute value of the effects of factors. Unimportant (that is, near-zero) effects manifest themselves as being near zero and on a line; while important (that is, large) effects manifest themselves by being off the line and well-displaced from zero. A red line through the insignificant factors helps to graphically delineate the difference between significant and insignificant factors.

[Figure 2 near here]

Figures 1 and 2 suggest that the most important factors (arranged in the order of importance) from the screening experiment are:

1. Instructor background
2. Professionalism of the instructor
3. Presentation style of the instructor
4. Interaction between the students and instructor in the classroom
5. Facilities
6. Method of course assessment
7. Types of exercise set by the instructor

The unimportant factors include feedback provided to the students, course content, frequency of lectures, and supporting materials provided in the virtual learning environment (e.g. case studies).

The last part of the analysis phase was to understand the optimal settings of the factors to maximize the average teaching effectiveness score. In order to accomplish this objective, a main effects plot was utilized (Launsby and Weese 1995). A main effects plot is a plot of the mean response values of each level of a factor. One can use this tool to compare the relative strength of the effects of various factors in an experiment. The sign of a main effect indicates the direction of the effect; i.e., whether the average teaching effectiveness scores increases or decreases. The magnitude tells us the strength of the effect. If the effect of a factor is positive, it implies that the average teaching effectiveness score is higher at a high level than at a low level for that specific factor. Figure 3 illustrates the main effects plot of all factors influencing the average teaching effectiveness.

[Figure 3 near here]

The optimal settings for maximizing the teaching effectiveness score are provided in Table 6.

[Table 6 near here]

Discussion, practical implications and limitations of the study

This case study demystifies the myth that experimental design is confined to primarily improving manufacturing processes. This research demonstrates the power of experimental design methods in understanding and scientifically evaluating the most influential factors that affect teaching effectiveness in the delivery of a postgraduate course within the higher education sector. The students were asked to identify the top five factors before the experiment was executed. The students used brainstorming and multi-voting methods to determine the top five factors they believe are important to teaching effectiveness. The results from students' perceptions were compared to the results from the screening experiment conducted with the input of over 100 students (Table 7).

[Table 7 near here]

It was very surprising to the authors that the most important factor identified by the students was not statistically significant in the screening experiment. Moreover, the method of assessment was important to the students' perception but it was not in the top five factors based on the analysis of data obtained from the experiment. In addition, the students did not rate the background of the instructor and facilities in the top five factors, but these factors were statistically significant and listed among the top five factors from the screening experiment.

One of the major challenges of the study was that the experiment took considerable time to plan, design, conduct, and analyze the data collated from the experiment. The authors also noticed that there were differences in the teaching effectiveness scores between the groups. It is believed that this is primarily due to students from varied cultural backgrounds; however, this information was not captured as part of the study. Another limitation of the study is that the experiment was conducted for a popular post graduate course. It would be beneficial to understand the results of the experiment for less popular post graduate courses in the university in order to drive improvements. Moreover, this research was conducted only for post-graduate courses and the results may vary for undergraduate courses. Finally, the authors would like to capture the perceptions of teaching effectiveness with a number of key academics through semi-structured interviews with the purpose of understanding their perceptions on teaching effectiveness.

The authors would like to highlight some of the practical implications of our findings in this section. First of all, it is a common practice that we assume we understand our processes and what factors are critical to the processes we work with on a daily basis. However this should not be the case in real life scenarios. A scientific approach such as experimental design is a powerful tool to obtain a greater understanding of the process and the most influential process parameters or factors which affect the critical to quality characteristics or response (also called process output). Secondly, quite often organisations put tolerances on all process parameters which results in exorbitantly high costs to the organisation. Experimental design can play a significant role in such scenarios in reducing operational costs by identifying the most critical process parameters which influence the process output and put tolerances on such critical parameters and relax tolerances on the insignificant parameters. The same principle is applicable to many processes within Higher Education (HE) sector too. However there will be challenges in the implementation due to lack of understanding about the benefits of experimental design, lack of potential buy-in from senior managers in the HE sector, intangibility associated with the process outcomes and the difficulty of measuring performance and the influence of human beings who work with the processes compared to machines in the case of manufacturing sector.

Conclusion and directions for further research

The application of the experimental design technique for process understanding is unknown in many service environments due to the lack of understanding its benefits. Further, there is an incorrect, pre-conceived mindset that it is confined to manufacturing processes and fear among senior managers that it is a time consuming exercise that involves advances statistics. The purpose of the paper is to illustrate the use of experimental design methodology in a higher education context. The key findings of the experiment have clearly indicated that students' perceptions of teaching effectiveness based on intuition and guesswork are not identical to the outcomes from a simple screening experiment. Moreover, the results of the experiment also provided a greater stimulus for the wider applications of the technique to other processes across the case study higher education sector. The authors would like to pursue more experiments in the future to understand the factors influencing teaching effectiveness for other postgraduate and undergraduate courses. In addition, the authors would like to explore how the average teaching scores vary for different cultural backgrounds and if any differences exist in the top five factors.

References

Antony, J. et al. 2011. "Design of Experiments for Non-manufacturing Processes: Benefits, Challenges and Some Examples." *Journal of Engineering Manufacture: Part B (IMechE)* 225 (11): 2078-2087.

Antony, J. 1998. "Some Key Things Industrial Engineers Should Know About Experimental Design." *Logistics Information Management* 11(6): 386-392.

Antony, J. 2014. *Design of Experiments for Engineers and Scientists*. London: Elsevier Ltd.

Antony, J., B. Rodgers and E. Cudney. 2017. "Lean six sigma for public sector organizations: Is it a myth or reality?" *International Journal of Quality and Reliability Management* 34(9): 1402-1411.

Antony, J., L., Sivanathan & E.V., Gijo. 2014. "Design of experiments in a higher education setting." *International Journal of Productivity and Performance Management* 63(4): 513-521.

Besseris, G.J. 2011. "Applying the DOE toolkit on a Lean-and-Green Six Sigma Maritime-Operation Improvement Project." *International Journal of Lean Six Sigma 2*(3): 270-284.

Barone, S., and E., Lo Franco. 2009. "Design of a university course quality by Teaching Experiments and Student Feedback (TESF)". *Total Quality Management & Business Excellence*. 20(7): 687-703.

Clayson, D. E. 1999. "Students' Evaluation of Teaching Effectiveness: Some Implications of Stability" *Journal of Marketing Education 21*(1): 68-75.

Cudney, E., S. Corns, and A. Gadre. 2014. "Virtual modelling for simulation based lean education." *International Journal of Lean Enterprise* 1(1): 3-21.

Cudney, E., C. Elrod, and S. Stanley. 2014. "A systematic literature review of six sigma practices in education." *International Journal of Six Sigma and Competitive Advantage* 8(3/4): 163-175.

Cudney, E., S. Furterer, and D. Dietrich. 2013. *Lean systems: Applications and case studies in manufacturing, service, and healthcare*. New York, NY: CRC Press.

Cudney, E., and D. Kanigolla. 2014. "Measuring the impact of project based learning in six sigma education." *Journal of Enterprise Transformation* 4(3): 272-288.

Darling-Hammond, L. and P. Young. 2002. "Defining "highly qualified teachers": What does "scientifically based research" actually tell us?." *Educational Researcher* 31(9): 13-25.

Ding, C. and H. Sherman. 2006. "Teaching effectiveness and student achievement: Examining the relationship." *Educational Research Quarterly* 29(4): 40-51.

Ezzell, J. M., and E. Cudney. 2017. "Evaluating the impact of teaching methods on student motivation." *Journal of STEM Education 18*(1): 5-18.

Ezzell, J.M. (2015), Utilizing learning style preferences and quality function deployment for curriculum development, MSc Thesis, Missouri University of Science and Technology, USA.

Galbraith, C. S., and G. B. Merrill. 2012 "Faculty Research Productivity and Standardized Student Learning Outcomes in a University Teaching Environment: A Bayesian Analysis of Relationships." *Studies in Higher Education* 37(4): 469-480.

Gliatis, V., I. Minis, and L. K. Myrto. 2013. "Assessing the impact of failures in service operations using experimental design with simulation." *International Journal of Quality & Reliability Management* 30(1): 23-46.

He, Z., G. Staples, M. Ross, I. Court, and K. Hazzard. 1997. "Orthogonal Software Testing: Taguchi Methods in Software Unit and Subsystem Testing." *Logistics Information Management* 5(5): 189-194.

Hinton, H., 1993. "Reliability and validity of student evaluations: Testing models versus survey research models." *Political Science & Politics* 26(3): 562-569.

Holland, C. W. and D. W. Cravens. 1973. "Fractional Factorial Designs in Marketing Research." *Journal of Marketing Research* 11: 286-294.

Johnson, L., and G. Bell. 2009. Designed Experiments in Service Quality Applications. ASQ World Congress on Quality and Improvement, Minneapolis, MN, USA, May 18-20.

Kanigolla, D., E. Cudney, S. Corns, and V. A. Samaranayake. 2014a. "Project Based Learning for Quality and Six Sigma Education." *International Journal of Six Sigma and Competitive Advantage* 8(1): 51-68.

Kanigolla, D., E. Cudney, S. Corns, V. A. Samaranayake. 2014b. "Enhancing engineering education using project-based learning for lean and six sigma." *International Journal of Lean Six Sigma* 5(1): 45-61.

Kumar, A., J. Motwani, and L. Otero. 1996. "An Application of Taguchi's Robust Experimental Design Technique to Improve Service Performance." *International Journal of Quality and Reliability Management* 13(4): 85-98.

Kupermintz, H., S. Lorrie, and L. Robert. 2002. "Teacher Effects as a Measure of Teacher Effectiveness: Construct Validity Considerations in TVAAS." Los Angeles, CA: CRESST/University of Colorado, Boulder.

Launsby, R., and D. Weese. 1995. "Straight Talk on Designing Experiments." Launsby Consulting, Colorado Springs, CO, USA.

Ledolter, J., and A. Swersey. 2006. "Using a Fractional Factorial Design to Increase Direct Mail Response at Mother Jones Magazine." *Quality Engineering* 18 (June): 469-475.

LeMahieu, P., L. Nordstrum, and E. Cudney. 2017. "Six sigma in education." *Quality Assurance in Education* 25(1): 91-108.

Marsh, H. W., and J. Hattie. 2002. "The Relation Between Research Productivity and Teaching Effectiveness: Complementary, Antagonistic, or Independent Constructs?" *The Journal of Higher Education* 73(5): 603-641.

McBer, H. 2000. *Research into teacher effectiveness: A model of teacher effectiveness (Research report, N. 216).* Norwich: The Crown copyright Unit.

McKeachie, W. 1997. Student ratings: The validity of use, *American Psychologist*. 52(11), 1218-1225.

Montgomery, D. C. 2012. Design and Analysis of Experiments, 8th Edition, John Wiley and Sons, USA.

Plackett, R. L., and J. P. Burman. 1946. "The Design of Optimum Multifactorial Experiments." *Biometrika* 33(4): 305–25.

Ree, S., Y. H. Park, and H. Yoo.2014. "A study on education quality using the Taguchi method." *Total Quality Management and Business Excellence* 25(8): 935–943.

Rowan, B., F.-S. Chiang, and R. J. Miller1997. "Using research on employees' performance to study the effects of teachers on students' achievement." *Sociology of Education* 70: 256-284.

Sanders, W. L. 2000. *Value-added assessment from student achievement data*, Cary, NC: Create National Evaluation Institute.

Seidel, T., .R. J. Shavelson 2007. "Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results." *Review of educational research* 77(4): 454-499.

Table 1 Teacher classification categories and classifications

Strong Change Over Time	Positive Change Over Time	Negative Change Over Time
Organization of class Knowledge Breadth of knowledge	Students are learning Individual rapport Appropriate workloads Encouraging student learning Fairness of exams Group interactions	Intellectually stimulating Open to points of view Enthusiasm

Table 2 Initial factors and levels identified during brainstorming

Factor	Low Level	High Level
Number of speakers in the delivery of lectures	Instructor on his or her own	Instructor and guest speakers
Background of the instructor	Instructor has recently completed a PhD with little exposure to industry	Instructor with rich industrial and teaching experience
Method of assessment for the course	Examination on its own	Coursework on its own (this means students will be asked to write a featured article for a practitioner journal and there is no examination associated with the assessment)
Interaction during the delivey of lectures	No discussions or question and answer sessions	Healthy discussions and question and answer sessions
Content of the course	Heavy theoretical content with a few case studies	Less theoretical content with more practical case studies
Frequency of lectures	Three hours per week (this implies students attend this particular course every week)	Six hours every two weeks(this implies students do not need to be in the campus every week for this particular course)
Feedback	No feedback for coursework and other related work relevant to assessments	Feedback for coursework and other related works relevant to assessments
Students' background	Students with no previous knowledge and experience in the field	Students with some knowledge on the topic and experience in the field
Students' attitude towards learning	Negative (i.e., not motivated, poor attendance, and no responsibility towards their own learning)	Positive (i.e., motivated, good attendance, and responsibility for their own learning)
Types of exercises in the classroom	Individual exercise	Group exercise
Presentation style of the instructor	Lack of clarity in speaking; monotonous in tone and	Clarity of speech; passionate speaker with varied tone and

	pitch	pitch
Time of delivery of the lecture	Morning	Afternoon
Field trip associated with the course	Unavailable	Available
Facilities of the room where the teaching is delivered	Poor lighting, poor ergonomics, lack of seating, etc.	Good lighting, good layout, ample seating, etc.
Supporting materials	Insufficient information	Sufficient information
Coherence	No structure, random	Logical and structured
Professionalism	Unfamiliarity of materials, unprepared, not handling questions professionally	Familiarity of materials, topic, explaining complex things, handling questions professionally.
Support after lectures	No support (e.g., lacking appointment times with tutors and poor response to emails)	Good support (e.g., tutor giving dedicated time and responding to student communication)
Approachability	Students do not feel they can approach lecturer	Students feel they can approach lecturer with confidence
Behaviour of lecturer	Less caring, rude, not helpful	Caring, approachable, and helpful

Table 3 Factors and their levels chosen for the screening experiment

Factor name	**Label**	**Low level (-1)**	**High level (+1)**
Course content	A	Heavy theory	Less theory
Presentation style	B	Monotonous tone	Varied tone
Interaction	C	No (no discussion)	Yes (healthy discussion)
Feedback	D	No	Yes (Have feedback)
Instructor background	E	Inexperienced	Experienced
Frequency of lectures	F	Weekly	Fortnightly
Professionalism	G	Unprofessional (unfamiliarity of materials)	Professional (familiarity of materials)
Assessment method	H	Exam on its own	Coursework on its own
Types of exercise	I	Individual	Group
Facilities	J	Poor	Well equipped
Supporting materials	K	Insufficient	Sufficient

Table 4 Plackett-Burman experimental layout (Plackett and Burman, 1946)

Run	A	B	C	D	E	F	G	H	I	J	K
1	+1	-1	+1	-1	-1	-1	+1	+1	+1	-1	+1
2	+1	+1	-1	+1	-1	-1	-1	+1	+1	+1	-1
3	-1	+1	+1	-1	+1	-1	-1	-1	+1	+1	+1
4	+1	-1	+1	+1	-1	+1	-1	-1	-1	+1	+1
5	+1	+1	-1	+1	+1	-1	+1	-1	-1	-1	+1
6	+1	+1	+1	-1	+1	+1	-1	+1	-1	-1	-1
7	-1	+1	+1	+1	-1	+1	+1	-1	+1	-1	-1
8	-1	-1	+1	+1	+1	-1	+1	+1	-1	+1	-1
9	-1	-1	-1	+1	+1	+1	-1	+1	+1	-1	+1
10	+1	-1	-1	-1	+1	+1	+1	-1	+1	+1	-1
11	-1	+1	-1	-1	-1	+1	+1	+1	-1	+1	+1
12	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Table 5 Plackett-Burman experimental layout including results

Runs	A	B	C	D	E	F	G	H	I	J	K	Teaching Effectiveness
1	+1	-1	+1	-1	-1	-1	+1	+1	+1	-1	+1	6.0, 5.2
2	+1	+1	-1	+1	-1	-1	-1	+1	+1	+1	-1	4.7, 6.3
3	-1	+1	+1	-1	+1	-1	-1	-1	+1	+1	+1	5.8, 6.5
4	+1	-1	+1	+1	-1	+1	-1	-1	-1	+1	+1	4.3, 3.5
5	+1	+1	-1	+1	+1	-1	+1	-1	-1	-1	+1	5.7, 6.1
6	+1	+1	+1	-1	+1	+1	-1	+1	-1	-1	-1	5.5, 5.1
7	-1	+1	+1	+1	-1	+1	+1	-1	+1	-1	-1	5.3, 5.1
8	-1	-1	+1	+1	+1	-1	+1	+1	-1	+1	-1	6.7, 6.5
9	-1	-1	-1	+1	+1	+1	-1	+1	+1	-1	+1	4.3, 4.2
10	+1	-1	-1	-1	+1	+1	+1	-1	+1	+1	-1	5.3, 4.9
11	-1	+1	-1	-1	-1	+1	+1	+1	-1	+1	+1	5.0, 4.4
12	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1.5, 1.4

Table 6. Main effects and their associated optimal settings

Main Effect	**Optimal Setting**	**Level Description**
Instructor background	High	Experienced tutor on the subject matter
Professionalism of the instructor	High	Professional tutor with familiarity of materials
Presentation style of the instructor	High	Varied tone during the duration of the lecture
Interaction between the students and instructor in the classroom	High	Good interaction with students by asking questions and engaging with them via exercises
Facilities	High	Well-equipped room with all the facilities for the delivery of lecture
Method of course assessment	High	Coursework on its own (i.e., students will be asked to write up a featured article for a practitioner journal)
Types of exercise set by the instructor	High	Group exercises (optimal size of 7 to 8)
Feedback	High	Effective feedback to course work
Course content	High	Less theory with more practical examples and real case studies
Frequency of lectures	Low	Weekly lectures (3 hours)
Supporting materials provided	High	Good case studies on different topics and other supporting materials such as white papers and practitioner-based viewpoint articles

Table 7 Comparison of students' perceptions with scientific experiment

Top 5 factors from students' perspective (use of brainstorming and multi-voting methods)	Top 5 factors from the screening experiment
Course content	Background of the instructor
Presentation style of the instructor	Professionalism of the instructor
Interaction in the classroom	Presentation style of the instructor
Professionalism of the instructor	Interaction in the classroom
Method of assessment	Facilities for delivery of the lecture

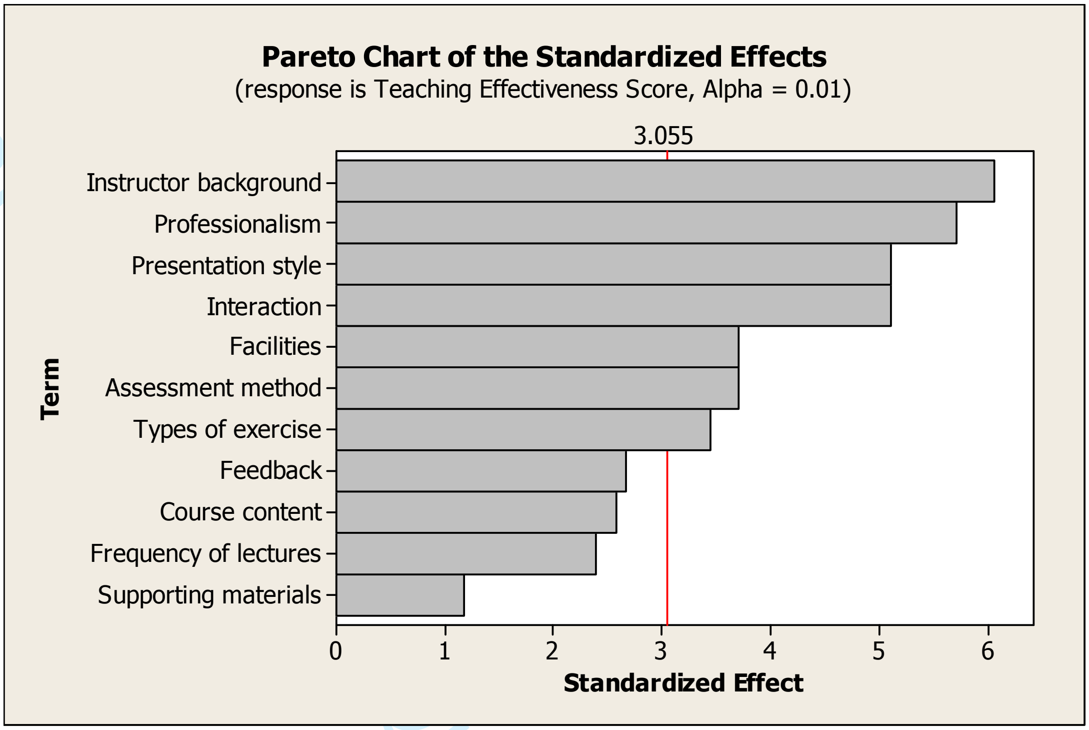

Figure 1 Pareto plot of the effects for the screening experiment

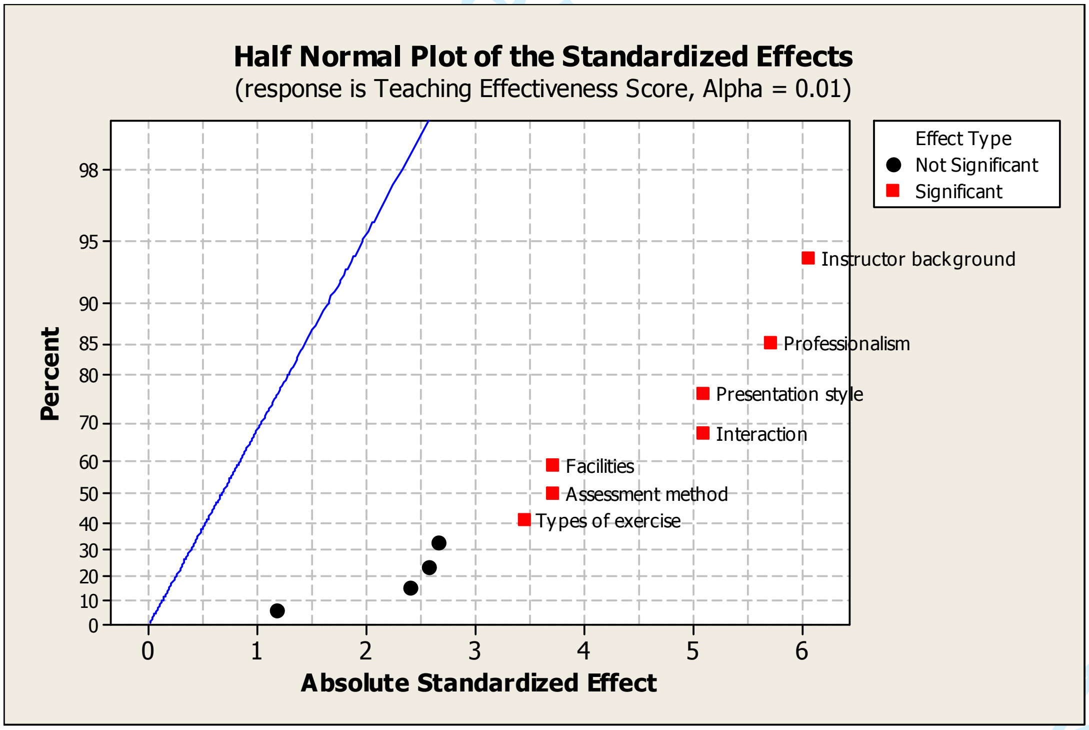

Figure 2 Half-normal probability plot of the effects for the screening experiment

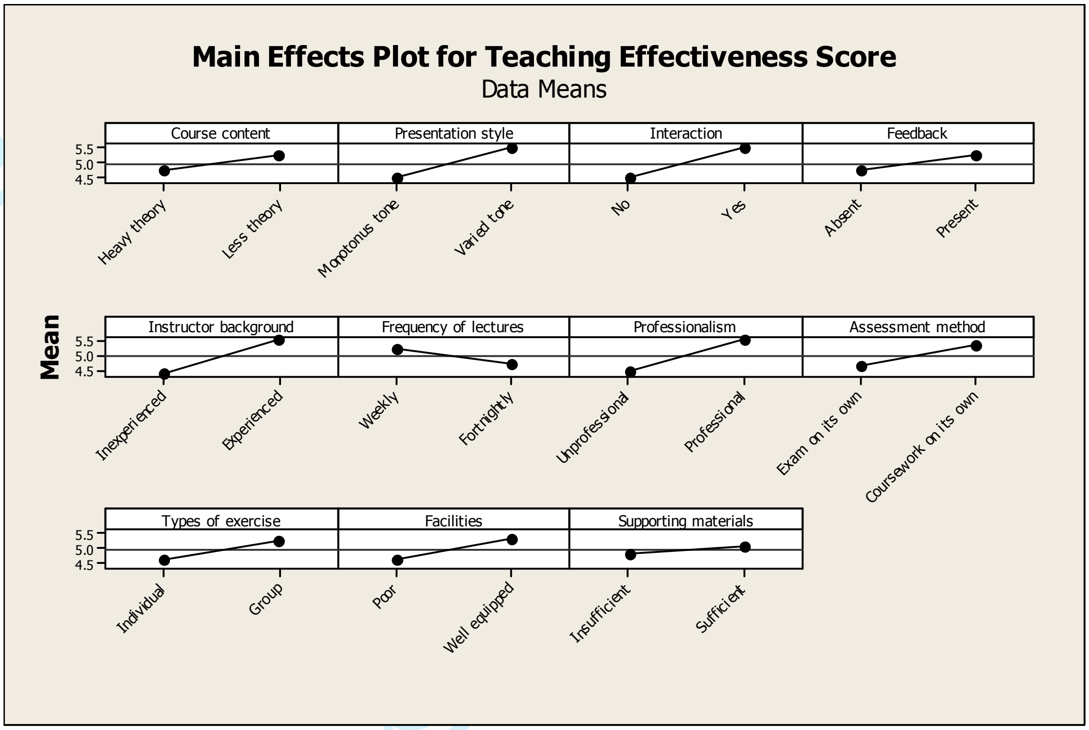

Figure 3 Main effects plot for the screening experiment