International Journal of Engineering Education Vol. 35, No. 1(A), pp. 262–272, 2019
Printed in Great Britain

0949-149X/91 $3.00+0.00

Engineering Student Evaluation of Teaching Quality in Saudi Arabia*

ABDELHAKIM ABDELHADI
College of Engineering, Prince Sultan University, Rafha Street, PO Box—66833, Riyadh 11586, Saudi Arabia.
E-mail: abdelhadi@psu.edu.sa

MOHAMMAD NURUNNABI
College of Business Administration, Prince Sultan University, Rafha Street, PO Box—66833, Riyadh 11586, Saudi Arabia.
E-mail: mnurunnabi@psu.edu.sa
St Antony's College, University of Oxford, 62 Woodstock Road, Oxford OX2 6JF, United Kingdom.
E-mail: mohammad.nurunnabi@sant.ox.ac.uk

Prior research suggests that student evaluation systems are a valuable aid to effective teaching quality. Research on student evaluations of teaching (SET) in engineering programme is somewhat lacking in the Middle East, particularly in Saudi Arabia. This study therefore focuses on the factors that influence engineering students' evaluations of teaching. A survey of 63 students from an engineering programme was carried out. Multivariate statistical analysis was used to group students into clusters based on the closeness of their decisions in rating instructors according to selected factors. Using Vroom's model of "expectancy theory" to investigate students' decision-making processes with regard to "improving teaching" or "improving the course format", the study reveals that meeting students' needs in the classroom appears to be an important factor when evaluating instructors at the end of the semester. In addition, the study suggests that instructors should not average or compare the average of student scores when considering the results of SET surveys. The findings also suggest that instructors should use a balanced approach in the classroom, paying attention to a range of factors in order to improve their overall teaching performance. In an attempt to raise awareness of SET, several implications are drawn out, and some directions for future research are suggested.

Keywords: student evaluations of teaching (SET); expectancy theory; multivariate statistical analysis; Saudi Arabia

1. Introduction

> "Research conducted on students' evaluations of teaching (SET) over the last 30 years suggests that SETs, while they do have shortcomings, provide valuable information regarding teaching effectiveness . . . [and have] proven to be a popular, relatively efficient means of obtaining feedback on instruction in higher education in the USA, the UK, and in many other parts of the world. Given the large investment in SETs and the strong likelihood that they will continue to be used, it is important that we learn more about how best to design and administer SETs. We also need to take steps to improve the quality of student responses and assist educators in using the data for maximum benefit." [1, p. 31]

Historically, the rating of university and college teachers by students dates back to 1920–1925 when students at Harvard University, University of Washington and Purdue University used to rate teachers. The initial findings of empirical studies indicated that smaller class sizes have a positive impact on retention of learning [2]. Since the 1970s SETs (student evaluations of teaching) have been primarily based on summative evaluation: that is, they have "summed up" overall performance, and have been used to help make decisions about promotion and tenure [3–5]. In the 1990s, there was a focus on quality improvement, in response to the increased international competition for students and operational funds in higher education [6].

Globally, the student voice has been increasing in volume in various forms. For example, SETs are now used in the professional development of teachers. Prior research indicates that student evaluation can be a way of highlighting course and lecturer strengths and areas for improvement [7]. Additionally, SET may be used differently in different countries: for example, the higher education sector in Portugal operates as a competitive market, higher education institutions are obliged to compete for scarce resources, and the survival of some existing institutions is threatened [8].

The use of SET in higher education is an important and controversial concept. Its effectiveness is an important issue for faculty members and administrators. SET has evolved as the main indicator of teaching quality [5, 9, 10]. Researchers have also found that there is a strong correlation between overall positive course ratings and grades received by students enrolled on courses [11–13].

In addition, few studies have considered the theoretical applications of SET; rather they have focused on a rigorous and quantified methodology to assess teachers, teaching and courses across

* Accepted 10 October 2018.

educational institutions over time [14]. Notably, using expectancy theory [15], the study measures the motivational effect of and attractiveness of various SET outcomes to students [16]. Four outcomes were identified from the literature on SET as "the primary uses of teaching evaluations": (a) improving teaching; (b) influencing tenure, promotion, and salary rise; (c) improving course contents; and (d) making exam results available to students. However, the study failed to identify group evaluations and expectations of the group [16].

In response to the limited research in a Saudi context, the primary aim of this study is to analyse how different factors influence engineering student evaluations of teaching quality in Saudi Arabia. In other words, the study looks at whether there is any evidence of engineering student evaluations of teaching quality being influenced by these factors [17]. This study particularly focuses on educational factors that influence students' decisions in evaluating teaching such as calling on students in class and administering pop up quizzes. It also explores students' views on how to improve courses. According to Duque "In line with subjective approaches (based on perceptions), there are simple models trying to understand how different perceptions of quality areas affect student satisfaction" [18, p. 1]. Other factors and models based on the perceptions of groups of students are not explored in prior literature. Accordingly, in terms of this new approach, the study uses a group of students clustered according to the closeness of their evaluations (e.g., group rather than individual evaluation). Multivariate statistical analysis is used to achieve this. The clusters are ranked according to the similarities of students' perceptions (i.e. student perceptions are most similar in cluster 1).

The paper is organised into eight sections. The next section discusses a literature review related to student evaluations of teaching quality in higher education and identifies specific gaps. The third section explains the clustering algorithms based on similarity coefficients. The fourth section outlines the research methodology, and the findings are then presented. The fifth section provides the results and discussion whilst section six contains implications of the study; The final two sections discuss limitations of the study and avenues for future research, and conclusions and contributions.

2. Literature review

The use of student perceptions is the most common approach to evaluating teaching quality in universities [6–8, 14, 18–23] (see Table 1). Notably, though, there is no consensus in the literature on the definition of effective teaching [24]. This approach is primarily based on the completion of online or paper forms and is popular because of its simplicity.

Additionally, Cook [6, p. 213] earlier argued that although a student's total educational experience is important for universities, the major factors that determine this experience are unknown, and how

Table 1. Prior research on SET and identified factors in evaluation of teacher performance

Author(s)	SET factors and findings
Centra and Creech [52]	A statistically significant relationship between student grade expectation and student evaluation on teaching quality. For instance, students expecting an A grade evaluate the instructor with a mean of 3.95 of a five-point scale while students expecting a D grade, evaluate the instructor with a mean of 3.02.
Haskell [53]	Faculty members' overemphasis on the numerical results may be contributing to an erosion of quality of teaching and scholarship, to a lower level of respect for teachers, and to weakening of faculty positions.
Emery et al. [19]	Punctuality in meeting class, availability outside class, fair and reasonable grading, preparation for class, and knowledge of the subject.
Yorke [54]	If an instructor inflates grades, they are more likely to receive a more positive evaluation. Several statistical techniques were used such as mean analysis, standard deviation analysis, analysis of variances, ANOVA or test of hypothesis.
van der Merwe [22]	Course content, knowledge, personality and attitude of a lecturer all play an important role in determining the rating of effectiveness of teaching in financial accounting.
Schumacher et al. [23]	Managing classrooms, organising for instructions, implementing instructions, monitoring student progress, and potentials.
Mohammed and Pandhiani [27]	Five dimensions of SET identified as factors (instructor's personality, knowledge, teaching ability, marking and grading policy, and course attributes and learning outcomes).
Hornstein [24]	Interpretations are questionable on conceptual and statistical grounds (SET equates to teaching competence).

Table 2. Relationship of academic staff job descriptions, key result areas and success factors identified by students

Academic staff job description, key result areas	Factors identified by students
Participate in course/programme development in order to ensure that teaching/learning activities are appropriate to the needs of consumers and customers.	Study timetabled appropriately. Class sessions timetabled appropriately. Time during the college day to study. Students informed of course/timetable changes.
Participate in theme/subtheme group development in order to ensure teaching/learning activities are appropriate to the needs of consumers and customers.	
Organise resources required for learning activities in order to ensure that all aspects of course/programme delivery are of high quality and comply with the curriculum/course document.	Information to guide students in their studies. Library facilities.
Facilitate and evaluate teaching/learning in order to sustain a high-quality educational service to maximise learning potential of students in accordance with the curriculum/course document.	Variety of teaching methods. Staff skilled in helping students to learn.
Provide appropriate academic supervision and support to students so as to monitor individual progress, ensure regular feedback and facilitate remedial action when necessary.	Advice provided on how to study. Feedback from staff. Personal support and Encouragement provided by staff. Help with course work.
Further develop and enhance positive links with staff in placement areas in order to strengthen and maintain relationships so as to optimise the teaching/learning environment.	Preparation prior to practice. Good placement supervision. Practical experience in a variety of placements.
Contribute effectively to college and personal development by participating fully and appropriately with colleagues in order to achieve the operational objectives.	
Participate in research and development initiatives in order to improve the quality of service provided to consumers and customers and promote the academic/scholarly standing of the college.	Knowledgeable staff.

Source: Cook (1997, p. 213).

this should be managed is not covered in the literature (see Table 2). Based on the feedback of nursing students, he suggests that students who receive high quality support from teachers give higher ratings. He also concludes that the conduct of academic staff—for example, turning up for classes, and providing students with accurate records of performance—is a key factor in student success.

A conceptual study by [8] who identify several factors which influence student satisfaction in higher education and the consequences of student satisfaction for their achievement it. Using quantitative structural equations, they found that the most influential variable was "image" of the university followed by perceived "value". The major finding of their study is that greater student satisfaction leads to increased student loyalty, and to the spread of positive messages about the university by word of mouth [8].

Similarly, [20] investigated expectations and perceptions of quality based on a sample of post-graduate Chinese students at a leading business and management school in the UK. Based on the five SERVQUAL dimensions, the research findings suggest that the "post-graduate Chinese students were not impressed by certain lecture theatres, the lack of study areas, class sizes, and insufficient media support. Hence, certain lecture theatres need to be re-vamped, and there is a genuine need to create study rooms for post-graduates, reduce class sizes, and provide more technical support facilities. Students also claimed that refreshments were expensive, and there was a strong desire for more budget-orientated offerings, which the institution could consider." [20, p. 328].

Again, two research questions were examined by [21]: (1) Are student ratings reliable and valid measures of teaching effectiveness for evaluating instructors? and (2) Are student ratings useful in improving the effectiveness of teaching? The first question deals with the validity of ratings (i.e. the relationship between students' ratings of teaching and measures of students' learning), and the extent to which the can be offer reliable feedback for the assessment of teaching and courses. The second question focuses on teaching improvement. Based on SET data from various departments of an Indian university, they find that ratings can be used as an aid for teaching improvement in an Indian context [21]. Investigating the evaluation techniques used to take on the challenges of quality improvement in

higher education, [25] found that there are typically three service areas in higher education: academic, administrative, and auxiliary functions. They argue that the service provided by educational institutions differs from that provided by the typical business, and future research could be directed at the development and application of quality techniques for each of these different areas.

In the case of student evaluations in Trinidad and Tobago, [7] aim to determine whether the student voice is being heard. The research focused on the following question: "How do you think this course could be improved?" Based on five purposefully selected courses from 2011/2012 to 2012/2013, they found little evidence that student evaluations actually led to any significant changes in lecturers' practices. On the other hand, [26] report that female instructors, part-time instructors, graduate classes, summer classes, morning and evening classes, and smaller classes received higher SET ratings. They conclude that students may form opinions about instructors in a very short time and that these opinions are largely unchanged at the end of the semester. This form of opinion is based on impression management.

It has been argued that students need to be engaged effectively to become active participants in improving teacher quality in both North America and the UK [1]. To measure teaching effectiveness and learning outcomes, it is important to maximise the practical information gained from student feedback. In another study, [27] aimed to determine the factors affecting students' perceptions of teaching effectiveness in a Saudi University, and found a significant relationship between students' ratings and teaching quality.

Prior research suggests that SET should not be used for summative evaluation of university faculties. For instance, student evaluation questionnaires tend to invite negative criticism [28], and low response rates could be a problem for SET exercises and suggest a number of ways to increase participation [29]. Recently, [24, p. 5] provided some recommendations for improving SET surveys, as reproduced below:

(a) Drop omnibus items about "overall teaching effectiveness" and "value of the course" from teaching evaluations: They are misleading.
(b) Do not average or compare averages of student rating scores: Such averages do not make sense statistically. Instead, report the distribution of scores, along with the number of responders and the response rate.
(c) Pay careful attention to student comments—but understand their scope and limitations. Students are the authorities on their experiences in class, but typically are not well situated to evaluate pedagogy generally.
(d) Use caution extrapolating student evaluations to the entire class. When response rates are low, extrapolation is unreliable.
(e) Avoid comparing teaching in courses of different types, levels, sizes, functions, or disciplines.
(f) Use teaching portfolios as part of the review process.
(g) Use classroom observation as part of milestone reviews.
(h) To improve teaching and evaluate teaching fairly and honestly, spend more time observing the teaching and looking at teaching materials.

Prior research show that student evaluation of teaching (SET) is affected by gender (both student and instructor) [30, 31]. [32] provides practical ideas to new professors to interpret student evaluations in order to get high evaluations from engineering students. Several studies investigate the perceptions of the usefulness of the teaching evaluation (TE) instrument and the rationale behind their responses and its effectiveness [33, 34]. They found that the faculty members receiving the best evaluations are not always the most effective teachers [33, 34]. It has also been argued that student satisfactions factors and the biases in SET results could be attributed to SET protocol, course content, and course delivery mode [35, 36]. In a recent study on assessing teaching evaluation within engineering departments in USA, it is found that:

> ". . . End-of-course student evaluation of teaching (SET) is the most common approach. In addition to SET, other approaches to teaching evaluation include classroom observation by peers or non-peers, evaluation of classroom materials, and student mid-course evaluations. It was also identified that the formative practices used mainly to gather student feedback or to improve faculty teaching are useful to the participants. Although there is substantial interest in improving teaching evaluation practices, generally current practices are still much different from identified best practices such as evaluating active learning approaches in the classroom, constructive alignment of outcomes, activities and assessments as well as the frequency and quality of feedback to students. The teaching evaluation system in engineering programs can be improved when educators become aware of and choose to adopt approaches that have been demonstrated to improve teaching and student learning." [37, p. 1317].

Since there is a lack of research regarding teaching evaluation practices that are used specifically in engineering programs, further research is still required in different regions [37]. To sum up, the objective of this research study is to characterize the engineering student evaluation of teach quality in Saudi Arabia.

3. Clustering algorithms based on similarity coefficients

The similarity coefficient is used to identify the relationships between parts with regard to certain characteristics under investigation. Based on these relationships, groups of items are identified. Among the algorithms used to identify and form part-families associated with the formation of machine cells are clustering algorithms based on the similarity-coefficient method [38]. These are used to find similarities between parts and machines and then to group them into part-family/machine cells. Pairwise similarity coefficients between machines and parts are calculated using specific similarity-coefficient formulas. These similarities are then organised into a matrix called the similarity-coefficient matrix. This matrix is used as an input to one of the clustering algorithms, such as complete linkage clustering (CLINK). Complete linkage clustering forms groups by merging nearest neighbours on the basis of the maximum distance/similarity between them. It works as follows:

- Start with M clusters and an $M \times M$ symmetric matrix of distances or similarities, $\boldsymbol{D} = \{d_{ik}\}$.
- Find the maximum distance/similarity in D = {dik}, let the distance between most similar cluster U and V be dUV.
- Merge the corresponding objects U and V to get the new cluster (UV).
- The distance/similarity between UV and any other cluster Q is computed by $d_{(UV)\,Q} = max\{d_{UQ,}\, d_{VQ}\}$.

The values d_{UQ} and d_{VQ} are the distance/similarity between clusters U and Q and clusters V and Q, respectively. The results are shown graphically in the form of tree diagrams (dendrograms).

This concept has been used in research focusing on preferred learning styles of students. [39] used clustering algorithms to group engineering students into cluster based on their learning styles. Using this approach in the present study, maximum similarity occurs amongst students who rate their course instructors more highly.

4. Research methodology

As demonstrated in prior research, the majority of the studies on SET have been based on a survey questionnaire [7, 21, 25, 40]. [41] highlight that, although SET plays a vital role in faculty promotion and tenure, and in assessing teaching effectiveness, there is growing controversy regarding the use of SET instruments. [13, p. 1190] listed methodological problems, including "(a) implying causation from correlation; (b) use of an inappropriate unit of analysis (the class average is usually appropriate, whereas the individual student is rarely appropriate); (c) negligence of the multivariate nature of SETs and potential biases; (d) inappropriate operational definitions of bias and potential biasing variables; and (e) inappropriate experimental manipulations".

4.1 Data

The data in this study have been collected in five randomly selected departments in a private university in Saudi Arabia. Undergraduate students enrolled in senior classes (Year 3 and Year 4) were invited to participate in the research. The survey was distributed to 110 students; however, only 63 students (37 students from Year 4 and 26 from Year 3) responded, giving a response rate of 57%. Students were informed that their response would be anonymised and that they should not include their names or any information which might allow them to be identified. Students were assured that data would be analysed collectively. Table 3 presents information on the characteristics of the students who participated in the study. [42] report average response rates of 70% for in-class surveys and 29% for online surveys. [43] raise concerns that lower response rates can lead to lower teaching evaluations.

In the present study, the factors to be included were identified based on prior research and a focus group of 25 students. [20] argues that the comments made in focus group discussions offer some evidence that students reflect deeply in considering factors that impact on teaching quality. [20] also highlights the necessity of carrying out focus groups before finalising a questionnaire in order to avoid one of the previously identified limitations of SERVQUAL. For this study, 17 factors were initially identified. However, after considering the feedback from the focus group, these factors were reduced to 10 (see Table 3). The identification of factors which influence student evaluations of teaching is intended to promote deeper understanding of how SETs can be used to support teacher improvement (since such meanings may vary within student groups over time and may also be dependent on different country specific contexts). As indicated in Table 1, prior research provides some confounding factors including class size, class time, department, location, gender of instructor and/or student, grades, expected grade, class difficulty, instructors' academic rank or seniority, whether courses are required or elective, first impressions of the instructor, age of students, whether the students are graduates or undergraduates, and timing and method of SET administration [40, 44]. Importantly, [40] found that instructors had influence

Table 3. Factors influencing student evaluations of instructors

Factor Number	Factor
1	(The instructor) Asks embarrassing questions.
2	(The instructor) Grades hard.
3	(The instructor) Uses unannounced quizzes.
4	(The instructor) Sets large amounts of homework.
5	(The instructor) Uses humour.
6	(The instructor) Calls on students in class.
7	(The instructor) Punishes late arrival of students.
8	(The instructor) Extends the class duration beyond the scheduled time.
9	(The instructor) Delays exam and homework results.
10	Instructor experience (Rank).

over how SET exercises are administered and this might also influence the results.

4.2 Rating

The study is based on the feedback of 63 students from an engineering programme. The survey used a five-point Likert scale [45] with options ranging from "strongly disagree" (1) to "strongly agree" (5). Numerical coding of each category allowed statistical analyses to be conducted. [46] found that using a five-point Likert scale may be better than using a ten-point rating scale. [47, p. 440] report that "many measurement biases affect student evaluations of instruction (SEIs). However, two have been relatively understudied: halo effects and ceiling/floor effects." To examine the ceiling/floor effects, they extended the standard five-point rating to either seven or nine points for use across three universities in the US and 537 students responded. [24], however, argues that most institutions in North America use a seven-point rating scale: (1) unacceptable, (2) very poor, (3) poor, (4) satisfactory, (5) good, (6) very good, (7) outstanding. Of course, good judgment and understanding must be applied in the interpretation of any statistical analyses. [47, p. 440] suggest that "direct instruction regarding how to use the evaluation forms seems to be necessary to limit these biases on ratings".

4.3 Reliability and validity

Prior research has increasingly questioned the validity of the survey as an instrument for measuring teaching effectiveness [12]. There is a strong positive correlation between grade received and rating of teaching effectiveness rating. This study used a construct-validation approach in which SET ratings are posited to be positively related to a wide variety of indicators of effective teaching and posited to be most highly correlated with variables [13]. The following issues were addressed regarding the reliability and validity of the current study:

- Senior students were chosen to participate in the study. This is because prior studies are unclear about whether first year undergraduate students have the ability to assess instructor teaching competence and rate teaching effectiveness truthfully [24].
- Several instructors teaching various subjects gave permission for their students to participate in the study and complete a 30 minutes survey questionnaire in class. This is important because if the students do not get or take enough time to answer the questions carefully, the results will be compromised [24, 44].
- To minimise any survey bias, two research assistants conducted the survey and explained its purpose. This addresses the issue raised by Simpson and Siguaw who argue that if students are not prepared to make mature evaluations it may not be appropriate to use SETs [48].
- The students were informed that the purpose of the survey was to gather their opinions on teaching effectiveness and that the survey would not be used by university administrators. This assures the students that the survey is not linked with grading. Two studies point out that students may simply care about their grades while the faculty cares about student learning, and wants students to be dispassionate evaluators of instructor performance [1, 49].

5. Results and discussion

Using a traditional approach to analysing the data, it is clear that "(The instructor) Grades hard" with a score of 256 is the factor that appears to have greatest influence on students' evaluations of their instructors. This is followed by: "Uses unannounced quizzes", "Extends the class duration beyond the scheduled time", "Calls on students in class", "Sets large amounts of homework", "Instructor experience" (Rank), "Delays exam and homework results", "Uses humour", and "Asks embarrassing questions". Finally, the factor with least influence on student evaluations of instructors is "Punishes late arrival of students" with a score of 140. Traditionally, this can result in instructors focusing on the factors that appear to have greatest influence on students' evaluations when they are interacting with students.

In this research, a different approach was used to analyse the data which classifies the importance of the factors mentioned in a different way and forces instructors to consider the importance of these factors from a different perspective. Minitab is used to build groups according to the full bonding aggregation described before. A dendrogram figure was produced on the basis of simila-

Table 4. Allocation of students to clusters

Cluster	Number of students	Student IDs
1	31	1-60-2-3-6-8-23-5-20-47-56-34-40-45-9-44-13-41-36-28-61-4-14-15-19-26-17-18-21-25-24
2	4	33-46-53-63
3	8	7-45-29-62-48-27-43-55
4	8	22-49-30-35-39-42-50-51
5	12	52-58-38-37-57-31-32-16-59-12-11-10

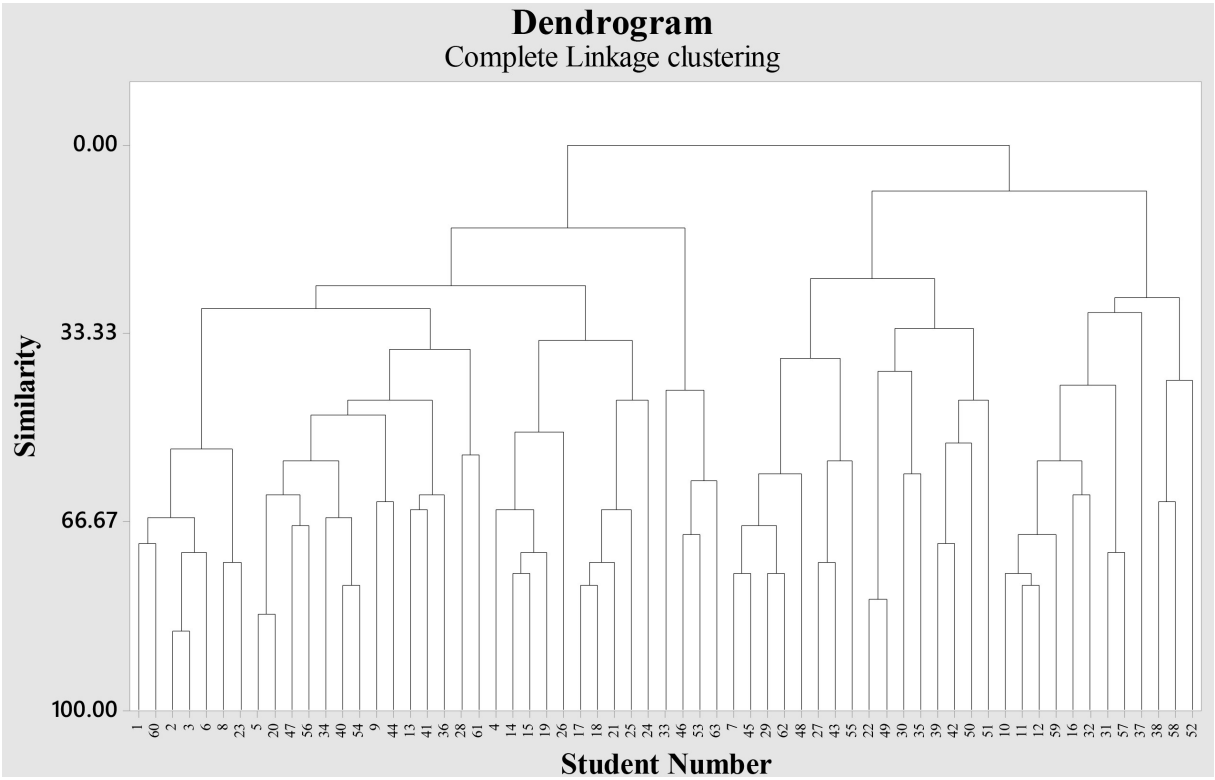

Fig. 1. Cluster formation based on similarity of students' perceptions.

rities between students' responses as shown in Figs. 1 and 2.

In this study five groups of students were created based on their responses to the factors influencing teacher evaluation, with the students in each group sharing common interests. The results are shown in Fig. 1. Group 1 consists of 31 students, Group 2 consists of 4 students, Group 3 and 4 consist of 8 students each, while Group 5 has 12 students. Table 4 presents student IDs for each cluster.

Table 5, part of the Minitab output, shows the similarities between the variables (factors influencing students' evaluations) and how they join with each other to form the five identified clusters. Variable 9 (''Delays exam and homework results'') joins variable 7 (''Punishes late arrival of students'') at about the 70% level of similarity to form cluster 7, while factor 4 (''Sets large amounts of homework'') and factor 5 (''Uses humour'') join each other at about the 60% level of similarity to form cluster 4. Variable 2 (''Grades hard'') joins cluster 7, and so on.

A decision was then taken to form five clusters from the considered variables (factors influencing students' ratings). And also, on the similarity between variables as perceived from the students' point of view. Results are shown in Table 6.

Cluster 1, shown in Table 6 comprises ''Asks embarrassing questions'' and ''Uses unannounced quizzes'' and is formed as a result of the collected students voting at about the 55% similarity level. Cluster 2 is formed by joining ''Punishes late arrival of students'' and ''Delays exam and homework results'' with ''Grades hard'' at 62% and ''Extends the class duration beyond the scheduled time'' at 53%.

In contrast, the traditional approach is the most common to analysing the data. It is clear that ''Grades hard'' with a score of 256 is the factor that appears to have the greatest influence on

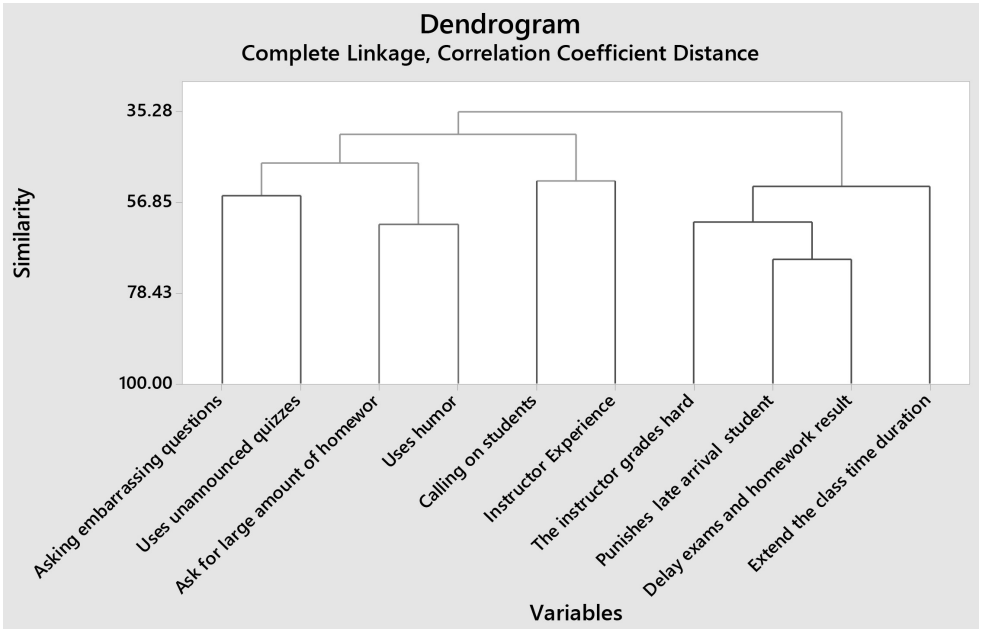

Fig. 2. Similarities between students based on perceptions of different aspects of teaching.

Table 5. Similarity level between the joined clusters

Step	Similarity level	Cluster joined		New cluster	Number of obs. in new cluster
1	70%	7: Punishes late arrival of students	9: Delays exam and homework results	7	2
2	62%	4: Sets large amounts of homework	5: Uses humour	4	2
3	62%	2: Grades hard	7	2	3
4	55%	1: Asks embarrassing questions	3: Uses unannounced quizzes	1	2
5	53%	2	8: Extends the class duration beyond the scheduled time	2	4
6	52%	6: Calls on students in class	10: Instructor experience (Rank)	6	2
7	47%	1	4	1	4
8	41%	1	6	1	6
9	35%	1	2	1	10

student evaluations of instructors. This is followed by: ''Uses unannounced quizzes'', ''Extends the class duration beyond the scheduled time'', ''Calls on students in class'', ''Sets large amounts of homework'', ''Instructor experience'' (Rank), ''Delays exam and homework results'', ''Uses humour'', and ''Asks embarrassing questions''. Finally, the factor with least influence on student evaluations of instructors is ''Punishes late arrival of students'' with the lowest score of 140 points.

Table 6. Cluster formations based on their similarity as precieved from students' point of view

Clusters	Variables
Cluster 1	Asks embarrassing questions; Uses unannounced quizzes.
Cluster 2	Grades hard; Punishes late arrival of students; Extends the class duration beyond the scheduled time; Delays exam and homework results.
Cluster 3	Sets large amounts of homework; Uses humour.
Cluster 4	Calls on students in class.
Cluster 5	Instructor experience.

It is obvious that results obtained using this approach give a different result than that obtained using traditional statistical analysis. As shown, cluster 4 indicates that students classify four factors—''Punishes late arrival of students'', ''Extends the class duration beyond the scheduled time'', ''Delays exam and homework results'' and ''Grades hard''—in one category which they use to evaluate instructors. In contrast, the traditional approach does not reflect the factors discussed in the present study.

6. Implications

This study has implications for the higher education policy agenda in Saudi Arabia. First of all, the

approach to SET should be improved and reviewed on a periodic basis by various stakeholders [50]. Secondly, increasing awareness of the importance of SET is essential, and this could be done by educating students. This is particularly important for the freshman or first year undergraduate or post-graduate students. Thirdly, there should be clear guidance for students since prior studies suggest that SETs are generally conducted with minimal instructions [47]. For instance, students should be told how important SETs are for improving the future performance of instructors. As per expectancy theory, Vroom argues that this is a ''cognitive model'' in which ''motivation [is] based on the belief that people actively weigh potential outcomes and make decisions based on those perceptions'' [15, p. 7]. The key aspect of theoretical aspiration is the decision-making process. In this case, students should be told that SET outcomes are used for ''improving teaching'' or ''improving the course format''. It is also recommended that to improve students' perceptions of overall educational experience, the factors included in SET surveys should be reviewed on a periodic basis to ensure they reflect the factors that have the greatest effects on perceived learning outcomes, and satisfaction.

7. Limitations and future research

The current study is subject to several limitations. The study uses the traditional definition of effective teaching. Future research could explore the meaning of ''effective teaching'' or teaching competence, and engage various stakeholders in this debate [5].

The study does explore the extent to which gender is a factor in ''effective teaching'' ratings based on categorising male or female instructors. Future studies could consider teaching effectiveness and the difference between male and female instructors across a broad range of courses [51–54]. Future studies could also consider the impact of other variables such as part-time instructors, graduate classes, summer classes, morning and evening classes, and smaller classes on perceptions [26].

There is a concern that an overemphasis on SET may be contributing to an erosion of teaching effectiveness. This present study is based on students' perceptions, but subjectivity must be complemented with objective performance measures. Therefore, future studies could use qualitative interviews as part of a case study approach to explore the views of students in more depth.

Undertaking research at one institution does not provide generalisable results. Hence, the findings reported here should be treated with caution. Thus, a future study could apply the questionnaire used in this study more widely to make comparisons between institutions and programmes, and to track changes in students' perceptions of teaching quality over time.

Future studies could explore differences in perceptions of teaching quality between semesters, local and international students, and students of different genders. The perception scores could be associated with high or low satisfaction over time. More established student groups could be examined in future, as the gender structure of some groups with strong preferences towards specific factors may influence the rating.

Finally, in terms of methodology, this survey was conducted in the middle of the semester. It could be argued that students' reflections may change at busy periods of the semester and this may also affect the response rate. Future studies could be conducted on the timing of the survey and the impact on students' perceptions, locally, regionally, and globally.

8. Conclusions

The study examined whether there was any evidence of student evaluations of teaching quality being based on other factors in higher education in Saudi Arabia. This research provides some new findings that meeting engineering students' needs in the classroom appears to be an important factor from the students' perspective when evaluating course instructors at the end of the semester. Factors influencing engineering student evaluations were formed into groups based on their importance using multivariate statistical analysis. Groups of students were then formed using the same concept based on their assessment of the importance of the factors used in this study. Importantly, the findings of the study suggest that instructors should not average or compare the average of students score; instead the research shows that instructors should look at the bigger picture by 'by comparing the way groups of students rate the elements they base their evaluations on. Students reflect on the potential outcomes of their actions in participating in SET surveys and, according to Vroom's expectancy theory, this reflection has two components: valence, and instrumentality. Students' have expectations about their overall educational experience at university and are able to assess this, but they lack the judgment to evaluate the pedagogy. The findings of the study also suggest that instructors should use some sort of balanced approach in the classroom. In terms of policy, the Saudi Ministry of Education could establish a national survey covering all universities similar to the UK National Student Survey (NSS) which is given to students in their final year as undergraduates. The NSS has grown to be a remarkably widely used SET tool. More than

300,000 students from 350 institutions completed the survey in 2015. A survey of this type in Saudi Arabia would help policymakers to understand and compare student satisfaction across universities.

References

1. F. Hammonds, G. J. Mariano, G. Ammons and S. Chambers, Student evaluations of teaching: improving teaching quality in higher education, *Perspectives: Policy and Practice in Higher Education*, **21**(1), 2017, pp. 26–33.
2. W. J. McKeachie, Research on college teaching: The historical background, *Journal of Education Psychology*, **82**(2), 1990, pp. 189–200.
3. S. J. DeCanio, Student evaluations of teaching: A multinomial logit approach, *Journal of Economic Education*, **17**, 1986, pp. 165–176.
4. R. A. Berk, Survey of 12 strategies to measure teaching effectiveness, *International Journal of Teaching and Learning in Higher Education*, **17**(1), 2005, pp. 48–62.
5. P. Spooren, B. Brockx and D. Mortelmans, D., On the validity of student evaluation of teaching: The state of the art, *Review of Educational Research*, **83**(4), 2013, pp. 598–642.
6. M. J. Cook, A student's perspective of service quality in education, *Total Quality Management*, **8**(2–3), 1997, pp. 120–125.
7. E. Blair and K. V. Noel, Improving higher education practice through student evaluation systems: Is the student voice being heard?, *Assessment & Evaluation in Higher Education*, **39**(7), 2014, pp. 879–894.
8. H. Alves and M. Raposo, Conceptual model of student satisfaction in higher education, *Total Quality Management & Business Excellence*, **18**(5), 2007, pp. 571–588.
9. J. J. Wallace and W. A. Wallace, Why the costs of student evaluations have long since exceeded their value, *Issues in Accounting Education*, **13**(2), 1998, pp. 443–448.
10. C. S. Galbraith, G. B. Merrill and D. M. Kline, Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business related classes? A neural network and bayesian analyses, *Research in Higher Education*, **53**(3), 2012, pp. 353–374.
11. D. L. Crumbley, Forensic accountants appearing in the literature, *New Accountant*, **10**(7), 1995, pp. 23–25.
12. H. H. Bauer, The new generations: Students who don't study, *Paper presented at The Technological Society at Risk Symposium*, Orlando, FL, 10 September, 1996.
13. H. W. Marsh and L. A. Roche, Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility, *American Psychologist*, **52**(11), 1997, pp. 1187–1197.
14. D. Ernst, Expectancy theory outcomes and student evaluations of teaching, *Educational Research and Evaluation*, **20**(7–8), 2014, pp. 536–556.
15. V. H. Vroom, *Work and Motivation*, Jossey-Bass, San Francisco, CA, 1964.
16. Y. Chen and L. B. Hoshower, Student evaluation of teaching effectiveness: An assessment of student perception and motivation, *Assessment & Evaluation in Higher Education*, **28**(1), 2003, pp. 71–88.
17. L. Crumbley, B. K. Henry and S. H. Kratchman, Students' perceptions of the evaluation of college teaching, *Quality Assurance in Education*, **9**(4), 2001, pp. 197–207.
18. L. C. Duque, A framework for analysing higher education performance: students' satisfaction, perceived learning outcomes, and dropout intentions, *Total Quality Management & Business Excellence*, **25**(1–2), 2014, pp. 1–21.
19. C. R. Emery, T. R. Kramer and R. G. Tian, Return to academic standards: A critique of student evaluations of teaching effectiveness, *Quality Assurance in Education*, **11**(1), 2003, pp. 37–46.
20. B. R. Barnes, Analysing service quality: The case of post-graduate Chinese students, *Total Quality Management & Business Excellence*, **18**(3), 2007, pp. 313–331.
21. A. Chatterjee, C. Ghosh and S. Bandyopadhyay, Assessing students' rating in higher education: A SERVQUAL approach, *Total Quality Management & Business Excellence*, **20**(10), 2009, pp. 1095–1109.
22. D. C van der Merwe, The usefulness of student evaluations for enhancing the effectiveness of teaching of financial accounting students at a South African university, in V. Tauringana and M. Mangena (Eds.), *Accounting in Africa (Research in Accounting in Emerging Economies*, **12A**, Emerald Group Publishing Limited, UK, 2012, pp.107–126.
23. G. Schumacher, B. Grigsby and W. Vesey, Determining effective teaching behaviors through the hiring process, *International Journal of Educational Management*, **29**(1), 2015, pp. 139–155.
24. H. A. Hornstein, Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance, *Cogent Education*, **4**(1), 2017, pp. 1–8.
25. A. Quinn, G. Lemay, P. Larsen and D. M. Johnson, Service quality in higher education, *Total Quality Management & Business Excellence*, **20**(2), 2009, pp. 139–152.
26. S. Nargundkar and M. Shrikhande, Norming of student evaluations of instruction: Impact of noninstructional factors, *Decision Sciences Journal of Innovative Education*, **12**(1), 2014, pp. 55–72.
27. T. A. Mohammed and S. M. Pandhiani, Analysis of factors affecting student evaluation of teaching effectiveness in Saudi higher education: The case of Jubail University College, *American Journal of Educational Research*, **5**(5), 2017, pp. 464–475.
28. T. Tricker, M. Rangecroft, P. Long and P. Gilroy, Evaluating distance education courses: The student perception, *Assessment & Evaluation in Higher Education*, **26**(2), 2001, pp. 165–177.
29. C. Weng, A. Weng and K. Tsai, Online teaching evaluation for higher quality education: Strategies to increase university students' participation, *Turkish Online Journal of Educational Technology – TOJET*, **13**(4), 2014, pp. 105–114.
30. A. G. d'Entremont and H. Gustafson, PANEL: Gender bias in student evaluations of teaching, *2017 ASEE Annual Conference & Exposition*, Columbus, Ohio, 24 June 2017, https://peer.asee.org/28731, Accessed 10 January 2018.
31. B. Hempel, K. Kiehlbaugh and P. Blowers, Student Evaluation of Teaching in an Engineering Class and Comparison of Results Based on Instructor Gender, *2018 ASEE Annual Conference & Exposition*, Salt Lake City, Utah, 23 June 2018, https://peer.asee.org/31006, Accessed 27 September 2018.
32. W. W. Schilling, J. K. Estell and F. C. Berry, Practical Interpretation of Student Evaluations for Starting Professors, *2011 ASEE Annual Conference*, Vancouver BC, 26 June 2011, https://peer.asee.org/practical-interpretation-of-student-evaluations-for-starting-professors, Accessed 27 September 2018.
33. I. Al-Abbadi, F. Alkhateeb, N. Khanfar, B. Mujtaba and D. Latif, Pharmacy students' perceptions of the teaching evaluation process in Jordan, *Education, Business and Society: Contemporary Middle Eastern Issues*, **2**(3), 2009, pp. 181–190.
34. D. L. Crumbley and K. J. Reichelt, Teaching effectiveness, impression management, and dysfunctional behavior: Student evaluation of teaching control data, *Quality Assurance in Education*, **17**(4), 2009, pp. 377–392.
35. C. R. Dennison, R. Butz, R. S. Fuhrer and J. P. Carey, Comparison of Student Evaluation of Teaching Results when Stratified by Protocol, Course Content, and Course Structure, *The International Journal of Engineering Education*, **31**(6A), 2015, pp. 1476–1490.
36. A. B. González-Rogado, M. J. Rodríguez-Conde, S. Olmos-Migueláñez, M. Borham and F. J. García-Peñalvo, Key Factors for Determining Student Satisfaction in Engineering: A Regression Study, *International Journal of Engineering Education*, **30**(3), 2014, pp. 576–584.
37. K. A. Villanueva, S. Brown, N. Pitterson, D. Hurwitz and A. Sitomer. Teaching Evaluation Practices in Engineering Programs: Current Approaches and Usefulness, *International Journal of Engineering Education*, **33**(4), 2017, pp. 1317–1334.

38. J. McAuley, Machine grouping for efficient production, *Production Engineer*, **51**(2), 1972, pp. 53–57.
39. A. Abdelhadi, Selecting teaching style based on group technology philosophy, *Journal of Applied Research in Higher Education*, **9**(4), 2017, pp. 521–529.
40. O. L. Liu, Student evaluation of instruction: In the new paradigm of distance education, *Research in Higher Education*, **53**(4), 2012, pp. 471–486.
41. P. R. Olds and D. L. Crumbley, Higher grades = higher evaluations: Impression management of students, *Quality Assurance in Education*, **11**(3), 2003, pp. 172–177.
42. C. J. Dommeyer, P. Baum, R. W. Hanna and K. S. Chapman, Gathering faculty teaching evaluations by in-class and online surveys: Their effects on response rates and evaluations, *Assessment & Evaluation in Higher Education*, **29**(5), 2004, pp. 611–623.
43. T. Ling, J. Phillips and S. Weihrich, Online evaluations vs in-class paper teaching evaluations: A paired comparison, *Journal of the Academy of Business Education*, **12**, 2012, pp. 150–161.
44. S. Nargundkar and M. Shrikhande, An empirical investigation of student evaluations of instruction—The relative importance of factors, *Decision Sciences Journal of Innovative Education*, **10**(1), 2012, pp. 117–135.
45. R. Likert, A technique for the measurement of attitudes, *Archives of Psychology*, **140**, 1932, pp. 1–55.
46. R. E. Landrum and K. A. Braitman, The effect of decreasing response options on students' evaluation of instruction, *College Teaching*, **56**(4), 2008, pp. 215–218.
47. J. W. Keeley, T. English, J. G. Irons and A. M. Henslee, Investigating halo and ceiling effects in student evaluations of instruction, *Educational and Psychological Measurement*, **73**(3), 2013, pp. 440–457.
48. P. M. Simpson and J. A. Siguaw, Student evaluations of teaching: An exploratory study of the faculty response, *Journal of Marketing Education*, **22**, 2000, pp. 199–213.
49. M. Braga, M. Paccagnella and M. Pellizzari, Evaluating students' evaluations of professors, *Economics of Education Review*, **41**, 2014, pp. 71–88.
50. D. Spiller and P. Ferguson, Student evaluations: Do lecturers value them and use them to engage with student learning needs?, *Education in a Changing Environment 6th International Conference: Creativity & Engagement in Higher Education, University of Salford, UK*, 2011.
51. A. Boring, K. Ottoboni and P. B. Stark, Student evaluations of teaching (mostly) do not measure teaching effectiveness, http://littleofficeofintegrity.org/yahoo_site_admin/assets/docs/attachment_-03.19291430.pdf, Accessed 10 January 2018.
52. J. A. Centra and F. R. Creech, *The Relationship Between Student, Teacher, and Course Characteristics and Student Ratings of Teacher Effectiveness*, ETS Program Report (PR-76-01), Educational Testing Service (ETS), Princeton, NJ., 1976.
53. R. E. Haskell, Academic freedom, tenure, and student evaluation of faculty: Galloping polls in the twentieth century, *Education Policy Analysis Archives*, **5**(6), 1997, pp. 1–32.
54. M. Yorke, Student experience surveys: Some methodological considerations and an empirical investigation, *Assessment and Evaluation in Higher Education*, **34**(6), 2009, pp. 721–739.

Abdelhakim Abdelhadi, PhD, EIT, FHEA, is an Associate Professor of the Department of Engineering Management, and College Research Director, College of Engineering at Prince Sultan University, Saudi Arabia. His research interests include Lean Manufacturing, Group Technology and Engineering Education. He has published articles in international journals.

Mohammad Nurunnabi, PhD, CMA, SFHEA FRSA FAIA(Acad), is the Aide to the Rector on Research and Internationalization, Chair of the Department of Accounting, and an Associate Professor in Accounting at Prince Sultan University, Saudi Arabia. He is an Academic Visitor of St Antony's College, University of Oxford, Oxford, United Kingdom. He holds a PhD with outright pass (no corrections) from the University of Edinburgh, UK. He is a s senior fellow of Higher Education Academy (SFHEA), UK, a Chartered Management Accountant (CMA), an academic fellow of the Association of International Accountants (FAIA-Acad), and fellow of the Royal Society of Arts (FRSA). His most recent book is *The Role of the State and Accounting Transparency* (Routledge: London and New York; 2016). He currently serves as Editor-in-Chief of *PSU Research Review: An International Journal*, Emerald, and Senior Associate Editor of *Journal of the Knowledge Economy*, Springer (SCOPUS Indexed). He serves on the editorial boards of twelve journals. He has received numerous prestigious research and teaching awards including IMA Faculty Leadership Award from Institute of Management Accountants (IMA), USA in 2017.