

On the Significance of Mindless Phenomenon in Performing Course and Teaching Evaluation at a Technological University in Taiwan

S.J. Jerome Teng^{*} Ching-Yuan Chang^{**} Shu-Yin Tseng^{***}

Abstract

Course and Teaching Evaluations (CTE) performed by students have been widely used in higher education. The CTE method is now recognized as an important tool for accountability and quality assurance in education. The CTE can help teachers to improve their teaching effectiveness through students' feedback. In the mean time, CTEs have also encountered many criticisms of its shortcomings, downsides, and inaccuracy. The most common criticism is that high CTE ratings reflect lenient grading and grade inflation. Another related shortcoming is that CTEs tend to encourage instructors to make their course easier and less challenging in order to keep the students happy. Teachers usually can raise their CTE ratings by simply becoming accommodating, undemanding, and personable, because the ratings often reflect students' emotional experiences in the course. These effectively make the CTE a personality and popularity contest in an attempt to turn disgruntled students into friendly ones. As a consequence of these and other factors, students' rating behavior and pattern in CTE is an interesting area to investigate. Through examining actual CTE data at a technological university in northern Taiwan, we have found a previously unknown but significant phenomenon that undercuts the reliability and accuracy of CTE. Many students tend to give the same ratings to all questions as though it is a mindless operation. In this paper, we report the extent and significance of this so-called mindless phenomenon based on detailed CTE data. The findings of this mindless phenomenon are illustrated by using graphical bar charts. Its likely cause, implication, and potential remedy are discussed.

Keyword: Course and Teaching Evaluation (CTE), mindless phenomenon, higher education

^{*} Department of Industrial Engineering and Management National Taipei University of Technology, Taipei, Taiwan

^{**} Department of Industrial Engineering and Management National Taipei University of Technology, Taipei, Taiwan

^{***} Department of Applied English Yu Da University, Miaoli County, Taiwan



學生執行課程與教學評鑑不用心現象之重要性研究，以台灣某科技大學為例

鄧樹楨* 張清淵** 曾淑穎***

摘要

高等教育機構由學生執行的課程與教學評鑑(CTE)已被廣泛使用，此方法是目前公認教育質量保證和負責的一個重要工具。經由學生意見反饋，教學評鑑可幫助學校與教師提高教學成效，但在同時，它也遇到許多對其缺點與不準確性的批評。最常見的批評是教學評鑑造成寬鬆成績與分數膨脹，另一相關的缺點是，教學評鑑通常是鼓勵教師，把課程變的較容易及不具挑戰性，目的是使學生高興。教師只要簡單的能表現出包容性、不挑剔、易相處，通常就可以提高自己的教學評鑑分數，因為評鑑分數經常反映的是學生對於課程在情緒上的經驗。這有效的使教學評鑑成為個性與受歡迎程度的比賽，教師企圖把心懷不滿的學生轉成友好的學生。由於這些與其他因素，使得學生執行教學評鑑的行為模式成為一個有趣的調查領域。本論文經由研究台灣北部某科技大學的實際教學評鑑數據，我們發現一個以前未知但很重要的現象，這現象能減弱教學評鑑的可靠性和準確性。我們發現許多學生往往對所有的問題給予同樣的評分，就像是一種不用心的操作。本論文根據實際的教學評鑑數據，我們將報告此所謂「不用心現象」的發生情況與其重要性，此現象的揭示是使用柱狀圖來圖解說明。本論文對於「不用心現象」的可能造成原因、影響程度與可能的補救措施都有討論。

關鍵字：高等教育，課程與教學評鑑(CTE)，不用心現象

* 國立台北科技大學工業工程與管理學系助理教授

** 國立台北科技大學工業工程與管理學系學生

*** 育達科技大學通識教育中心講師



1. Introduction

Course and Teaching Evaluations (CTE), or Student Evaluation of Teaching (SET), are evaluations conducted by university students. They are used as an important but controversial tool in the improvement of education and teaching quality during the past few decades. Student ratings of course and teaching have a relatively long history. It was reported by Marsh [1] and Wachtel [2] that the first 'teacher rating scale' was published in 1915. Informal student evaluations of faculty were started in the 1960's by enterprising college students [3]. Today, web-based CTEs are convenient for students to perform and they are widely used in almost all institutions of higher education.

The CTEs are survey results of students' satisfaction and opinions about the course and instruction. This feedback mechanism can help the teachers improving their teaching effectiveness. It makes good contribution to the quality of teaching. The CTEs also help the school administrators monitoring the educational quality and teachers' job performance. In some educational institutions, the CTE results are also used for reviews of faculty, course assignments, teaching awards and honors, and/or promotional decisions.

To be more specific, there should be two kinds of CTE formats. Those used to evaluate teachers for the performance, salary and tenure purposes are called summative evaluation. For example, typical summative questions are "Overall, how do you rate this instructor's teaching ability compared to other instructors in this school?" "Overall, how do you rate this course compared to other courses you have now and have had in the school?" These questions are more related to teachers' teaching ability and performance. On the other hand, those used to diagnose teaching and allow teachers to improve their teaching are called formative evaluation. They are designed to provide teachers with feedback on various aspects of course and teaching. However, some formative CTEs can also be used in the summative category. There are plenty of indications that these two kinds of evaluation should be separated. But often for simplicity's sake, they are combined and practiced as one.

The CTE or SET questions take on different formats and address different aspects of the course, teaching, and teacher performance. They are probably the main source of information used for evaluating faculty teaching performance [4, 5]. Teachers' tenure, promotion, and salary decisions are often influenced by such results [6]. The premise that the CTE methods conducted by students are useful and reproducible has been generally accepted as valid. Researchers who have surveyed the published literature on the CTEs agree that their usefulness invariably outweigh their shortcoming [7, 8]. There is a good likelihood and general trend for highly rated teachers to be those whose students achieve well. Poor overall CTE ratings are signs of problem in teaching performance [9]. Research on the validity and accuracy of CTEs is enormous, but they are riddled with inconsistencies.

The CTE result when used summatively as a terminal indicator of mastery is a contentious issue among academics [10, 11], many of whom consider it a superfluous and detrimental process with no real role in higher education [12]. Part of the controversy and lack of respect for CTEs have been attributed to a notion that students rate highly only those faculty who are easy graders and are personable [13], which is also called "halo effect". The CTEs are not always accurate because their validity and usefulness have been doubted by research [14, 15]. For example, Slade and McConville (2006) [14] have found that the survey instrument was flawed and all compulsory questions collapsed into one dimension. That dimension was determined to be the extent of popularity of the lecturer. Isley and Singh (2005) [15] have concluded from empirical results that if an instructor has some classes in



which students expect higher grades, a more favorable average CTE is obtained in these classes.

The CTE survey has become part of the culture and routine in higher education today. Other than the teachers, the students are also stakeholders. For CTE to be truthful and useful, student cooperation in the process is vital, yet often it can be undermined [16, 17]. Students who do not understand the use of teaching evaluations or feel that the evaluation will be used for purposes that they do not appreciate will stop giving truthful input [18]. Without truthful input, the CTE results may do harm than good. There are potentially many reasons for students not to give truthful or objective opinions. One may be due to ignorance. If the students do not think highly of the CTE result, then they are less likely to take it seriously. To make the CTE a useful instrument and gauge, the first step is to have the students take the CTE survey seriously and objectively. Truthfulness in CTE is the bottom line of reliability and accuracy.

In this paper, we report a previously unknown factor which is called mindless phenomenon in the CTE process. Many students are not paying attention to individual questions in the CTE and they tend to give the same rating to most, if not all, questions. As though, they have decided the overall rating beforehand. If this phenomenon dominates in the CTE process, then the design and contents of the questionnaire become irrelevant. By analyzing actual CTE data collected at a technological university in northern Taiwan, we have found that this mindless phenomenon does exist in a significant proportion which can undercut the design and accuracy of CTE. We will attempt to describe the reasons why students tend to give the same ratings to all questions in the CTE. In Section 2, we describe the finding and properties of this phenomenon. In Section 3, the actual CTE questions and data used are described. In Section 4, we present the results and discussion of this study. The findings of this mindless phenomenon in CTE are illustrated by using graphical bar charts. The likely cause, implication, and potential remedy are also discussed. Finally, a brief conclusion is made in Section 5.

2. Mindless Phenomenon in CTE

Other than those commonly known factors that cause inaccuracy or bias in CTE [19], in this paper we present a previously unknown factor called mindless or mind-free phenomenon. The CTE event is usually held in a period before the end of a semester. The teachers are not allowed to see the result until the semester is over. Thus the CTE result has no impact on the course grading. Due to lack of incentive to fill the questionnaire, many students do not pay attention to the CTE event. Some may even skip the event. This could make the response rate too low for evaluation purpose. In order to boost the response rate, many universities in Taiwan require the students to complete their online CTE as scheduled, otherwise they would automatically lose some privileges such as early course registration for the next semester. These procedures are all done online. As a result, many students see the CTE as a required chore near the end of a semester. A student usually has many courses and thus many questionnaires to fill. They tend to finish them quickly. Many students may not have the patience to read and rate each question carefully, especially when a very long list of questions is implemented in a questionnaire. These are likely CTE scenarios that would happen at universities in Taiwan.

Based on study of actual CTE data, general patterns in survey results can be delineated. We need to know to what extent that the students do not read and judge carefully on each question. A careful rater would produce ratings that show some fluctuations among the



On the Significance of Mindless Phenomenon in Performing Course and Teaching Evaluation at a Technological University in Taiwan

questions. On the other hand, if a student completes the online CTE quickly or mindlessly, the likely outcome is that his/her ratings would remain the same or show little fluctuation. In this study, the actual CTE data were obtained from a questionnaire that had a long list of 20 questions. Such a long list of questions could easily add to the challenge for the students to read and judge them carefully. Our study shows that many students actually tended to rate the same number for all questions. Only then can the ratings be finished quickly. Irrespective of the questions asked, such students perform their rating job mindlessly. The same ratings given are likely based on the students' general feeling toward the course and the teacher.

There may be additional reasons for the students to give a fixed rating number. For example, if the student thinks highly of the teacher or is happy about the course, the ratings could be straight five's for all questions. The students may regard this top rating as a reward or compliment to the teacher. This is the so-called friendly students, especially when they are acquainted with the teacher. In the other extreme, if the students are highly dissatisfied or discontented with the course or the teacher, the ratings could be straight 1's or 2's for all questions. This is the so-called hostile students. The disgruntled students have an ax to grind with the teacher and they may take this low rating as a revenge or punishment to the teacher. The mood may also come from the feeling of being unfairly treated. Overall, these students are in the minority.

Both the friendly and hostile students are emotional raters. Their ratings are unfairly and subjectively given. They are most likely to perform the CTE without careful reading of the questions. In actuality, they have inadvertently joined the ranks of mindless students in performing the CTE. Their behaviors undercut the validity and accuracy of CTE even more, because they tend to give extreme scores. The extent of damage to the accuracy of CTE depends on the number of mindless raters. It will be shown later in this paper that hostile students would do more damage to the CTE rating than friendly students could help.

Summarizing from the rating results, the students rating behavior may be described as follows. For those students who do not have a strong opinion or particular liking toward a course or a teacher, they tend to give average or above-average ratings in the CTE. Some students may read the individual questions and give ratings between 3 and 5. Some impatient raters may fall into the mindless mode and give straight ratings, such as all 4's, 5's, or 3's. They are less likely to give ratings of 2's or 1's, unless the teacher's performance is particularly unsatisfactory. This scenario is usually true because the school overall ratings are in the range around 4. Rarely have the ratings of 1's or 2's appeared in the CTE data. Under normal circumstance, if raters do rate the questions carefully, their rating scores would show some fluctuation among the questions. Statistically, the distribution or standard deviation will also increase to reflect the rating variations. For mindless raters, the distribution or standard deviation of ratings is small to reflect little or no fluctuation. In the next two sections, we report the detailed study of the CTE data, the analyzed results, and related discussions.

3. CTE Method and Data

The CTE data analyzed in this paper are collected at a technological university in northern Taiwan. Web-based CTE method has been used in this university for many years. A single set of questionnaire was designed and used for all students. The CTE event was scheduled near the end of a semester, not long before the final exam period. Teachers can access their own overall CTE results after the grades are given. The CTE questionnaire used has a total of 20 questions, however the last two questions do not count in CTE rating. A



five-point Likert scale is used, for 1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree. The ratings of 5 and 1 represent the best and worst rating, respectively. One final averaged score weighted evenly over the 18 questions is used for the CTE rating. No standard deviation data is given in the CTE report.

The questionnaire consists of mostly formative questions and only two summative questions. The questions include course content, teaching method, teacher attitude and enthusiasm, skills, grading, and overall evaluation. The 18 questions are listed below.

- Q.1 The course contents are consistent with the course title.
- Q.2 The course materials are moderate in difficulty.
- Q.3 The teacher can follow students' understanding and then adjust the instruction.
- Q.4 The teacher has provided teaching plan or syllabus.
- Q.5 The teacher delivers the course material in a clear fashion.
- Q.6 The teacher uses teaching aids to assist teaching.
- Q.7 The teacher adjusts teaching method based on student needs.
- Q.8 The teacher is able to maintain order in the classroom.
- Q.9 The teacher shows enthusiasm in teaching.
- Q.10 The teacher has not been frequently changing class hours, absent or late for class, or leaving early.
- Q.11 The teacher cares about students' learning problem and makes timely response.
- Q.12 The teacher explains or presents the course content with patience.
- Q.13 The teacher informs in advance about the grading method and standard.
- Q.14 The teacher has emphasized or checked on students' in-class participation.
- Q.15 The teacher's exam/test can reflect or focus on the course contents.
- Q.16 The teacher pays attention to the importance of student attendance.
- Q.17 My interest in this course has been gradually increasing.
- Q.18 I have been profited from learning after taking this course.

The range and coverage of the above questions appear to be comprehensive. The contents are similar to CTE questionnaires used by other universities in Taiwan. However, the CTE has a total of 20 questions which is higher than other schools. Many minor issues are being asked, such as maintaining order in the classroom and paying attention to student attendance. One would argue that many of the questions have little to do with teaching performance. Good examples are Questions 1, 2, 4, 8, 10, 13, 15, and 16. There are only two summative-type questions as in Questions 17 and 18. It is easily seen that too many trivial questions are being asked. To take all these 18 ratings equally into account, it doesn't seem to do justice on fair evaluation of teaching performance. It is likely that when students encounter such a long list of 20 questions, many of them may easily lose patience in reading and judging carefully. Raters without patience and finishing the rating in a rush are mindless raters. Uniform ratings for all questions are thus produced.

The overall averaged CTE rating of the whole university in 2012 Fall semester is 4.2 which reflects a good performance in CTE. Courses having rating below 3.5 would be very



On the Significance of Mindless Phenomenon in Performing Course and Teaching Evaluation at a Technological University in Taiwan

poor performance and near the bottom in ranking. There are also written comments made by the students. If there were hostile students in the ratings, they are likely to leave a few words to express their feelings. Based on the comments made, such as pleading for passing grades, it seems that some students are unaware that the teacher is not able to see them before the grades are given.

4. Results and Discussion

Each teacher is able to see their own CTE final scores. However, the final score doesn't reveal what happened inside the rating. The composition of rating components would tell the story of how the ratings are performed. In order to illustrate the mindless phenomenon that exists in CTE, the rating results are analyzed both on an individual basis and from a larger perspective. The rating composition and fluctuation will reveal the extent that the mindless phenomenon happens.

For a typical example that has a good CTE rating, Figure 1(a) shows a bar diagram that displays the averaged CTE ratings for each of the 18 questions. It is a junior-level course that has 28 students of which 27 have responded. The response rate is high due to the CTE is required. The overall rating or mean value is 4.4218 which is somewhat higher than the school average of 4.2. The standard deviation of the 18 separate ratings is 0.0384. In other words, as shown in Fig.1(a), small deviation means very similar ratings were achieved for the 18 questions. Consider that many of these 18 questions are totally different in nature, nonetheless similar ratings were achieved among them. In order for this to happen, most of these 27 students were likely to give their fixed ratings to all the questions. This pattern of little or no fluctuation meets the characteristics of mindless operation. To further illustrate the existence of mindless phenomenon, Fig.1(b) shows the stacked bar diagram that displays the composition of accumulated ratings for each question. There is no ratings of 1's or 2's. The accumulated rating numbers, with respect to 5, 4, and 3, given for each question are relatively evenly distributed. This shows that there are mainly three groups of raters. Each group more or less gives the same rating to each question. This example and many other similar results have revealed and confirmed that mindless operations do exist to a large extent. It is a frequent and significant phenomenon in the CTEs of this university.



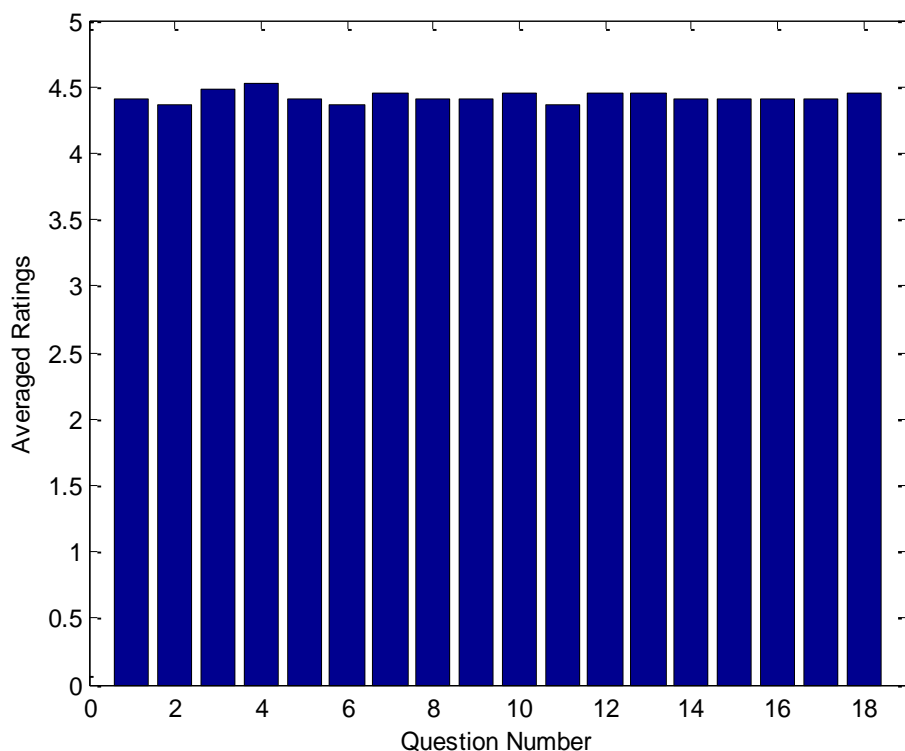


Figure 1(a). This bar diagram shows the averaged CTE ratings for each of the 18 questions. (27 respondents. Mean value = 4.4218, Standard deviation = 0.0384, Variance = 0.0014)

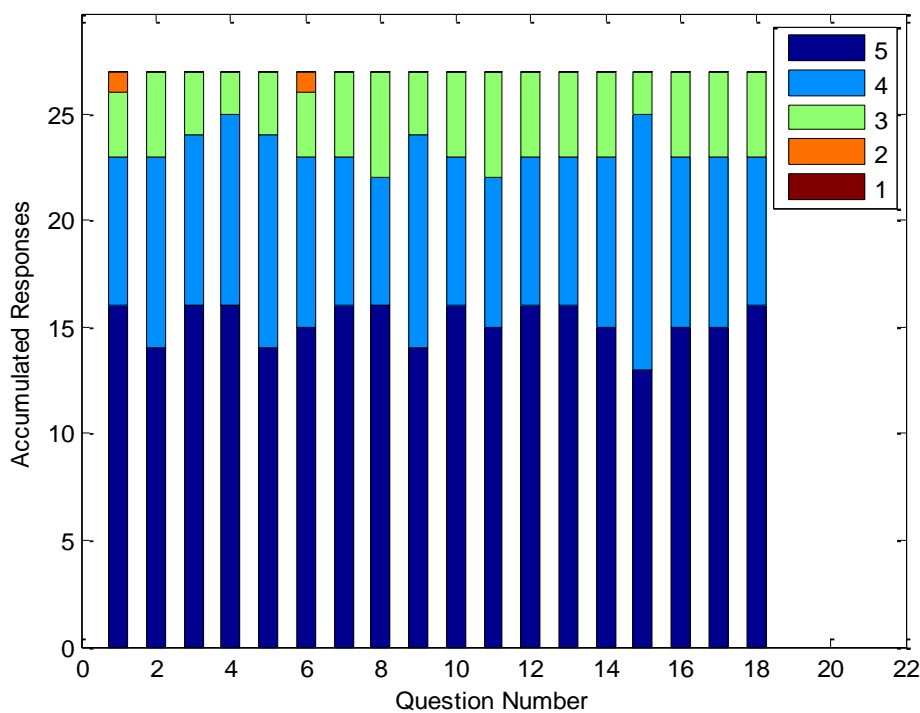


Figure 1(b). This stacked bar diagram shows the composition of ratings in Figure 1(a).



On the Significance of Mindless Phenomenon in Performing Course and Teaching Evaluation at a Technological University in Taiwan

To show an example that has a low CTE rating, Fig.2(a) is a bar diagram that plots the CTE ratings for the 18 questions. It is a freshman-level course. A total of 36 out of 38 students have responded to the CTE. This course received a poor overall rating of 3.3904. The standard deviation of the 18 ratings is 0.1440 which indicates the fluctuation is higher than that of Fig.1(a). The composition of accumulated ratings for each question is shown in a stacked bar chart in Fig.2(b). All five ratings exist for each question. As shown in Fig.2(b), it is likely that two hostile or discontented students have given ratings of 1's to all the questions. Low rating of 1's effectively drags down the overall rating because most average ratings are in the range of 4. There are more rooms downside than upside. This shows that a teacher cannot afford to have many unhappy or disgruntled students, his/her rating can easily go down as a result. Also noticed in Fig.2(b), there are 6 to 7 friendly students in the class who have given high ratings to all questions. The mindless phenomenon still exists in the example of Fig.2, only somewhat less prominent than that of Fig.1.

For comparison purpose, Fig.1 and Fig.2 are from two different courses taught by the same male teacher in the same semester. However, the achieved CTE ratings are drastically different at 4.4218 and 3.3904. The score of 3.3904 is more one point below 4.4218, yet the courses were taught by the same teacher. If the school takes the CTE rating seriously, then this teacher would be rated very poor in performance. One might ask what did he do differently in that course such that he had two hostile students and many unfriendly or dissatisfied students. Two hostile students have given straight 1's to all the CTE questions! They were determined to give the lowest scores, to say they are mindless-type raters may be inappropriate.

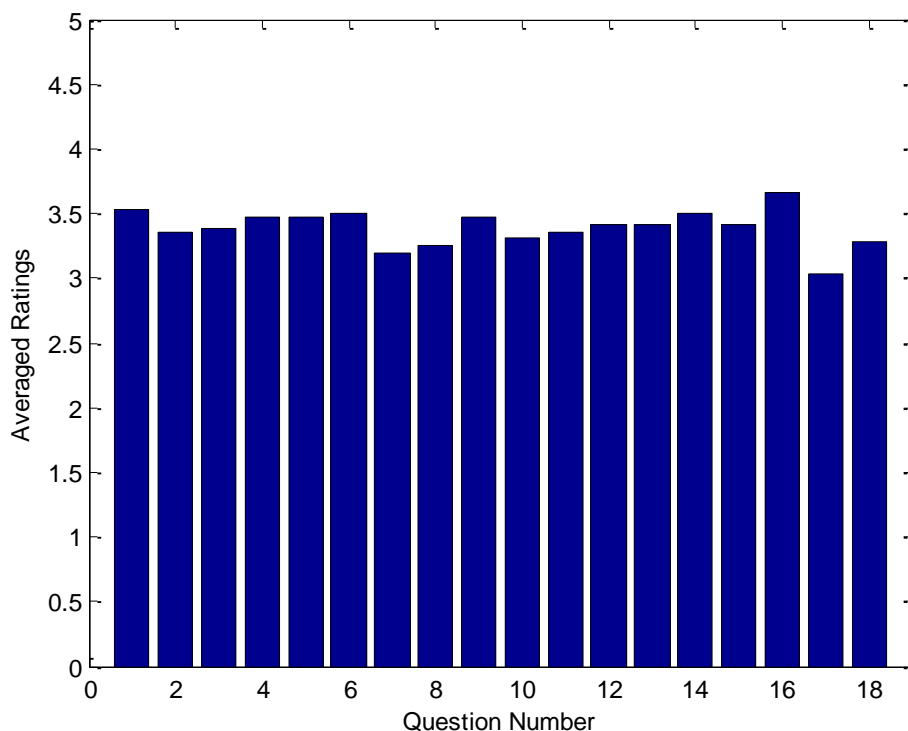


Figure 2(a). This bar diagram shows the averaged CTE ratings for each of the 18 questions. (36 respondents. Mean value = 3.3904, Standard deviation = 0.1440, Variance = 0.0196)



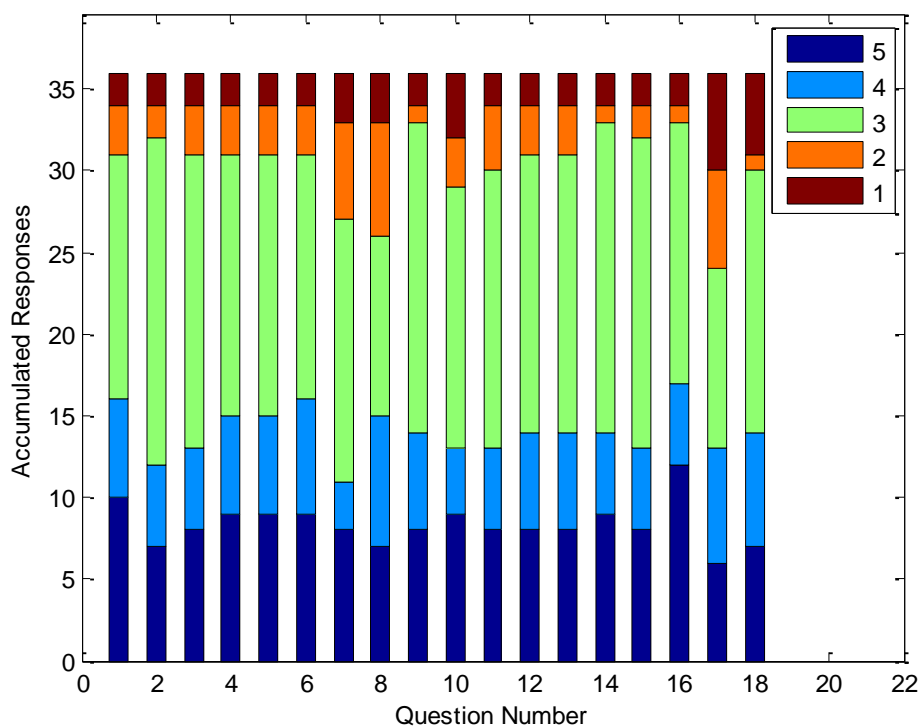


Figure 2(b). This stacked bar diagram shows the composition of ratings in Figure 2(a).

Several examples of across-the-board mindless operation have been observed for smaller classes having less than ten students who responded. Their rating compositions are simplified versions of Fig.1(a) and 1(b) but with the individual components evenly distributed. The students voted their own same number in all 18 questions without any astray. They are certainly efficient and fast performers in the CTE survey. These may be considered as special cases, because only several students are involved.

To see and compare more compositions of CTE ratings, Figures 3 and 4 show another set of examples. Again the two examples are taken from two different courses taught by the same female teacher in the same semester. Figure 3 is a freshman-level course. Figure 3(a) shows the bar chart of CTE ratings of the 18 questions. A total of 39 out of 49 students have responded. The overall CTE rating is 3.9017 which is not far below the school average of 4.2. The standard deviation of the 18 ratings is 0.0881. This indicates that similar ratings were given among the different questions. Figure 3(b) shows the stacked bar chart that displays the composition of accumulated ratings for each question. Again, three distinct groups of raters of 5, 4, and 3 appear in the chart. They are slightly less evenly distributed than that of Fig.1(b). This is reflected by the difference in standard deviations which is 0.0384 versus 0.0881. Another difference is that there are more group 5 'voters' in Fig.1(b) compared to that of Fig.3(b). It shows that, without the presence of hostile students, more friendly students will make some improvement in the rating outcome.



On the Significance of Mindless Phenomenon in Performing Course and Teaching Evaluation at a Technological University in Taiwan

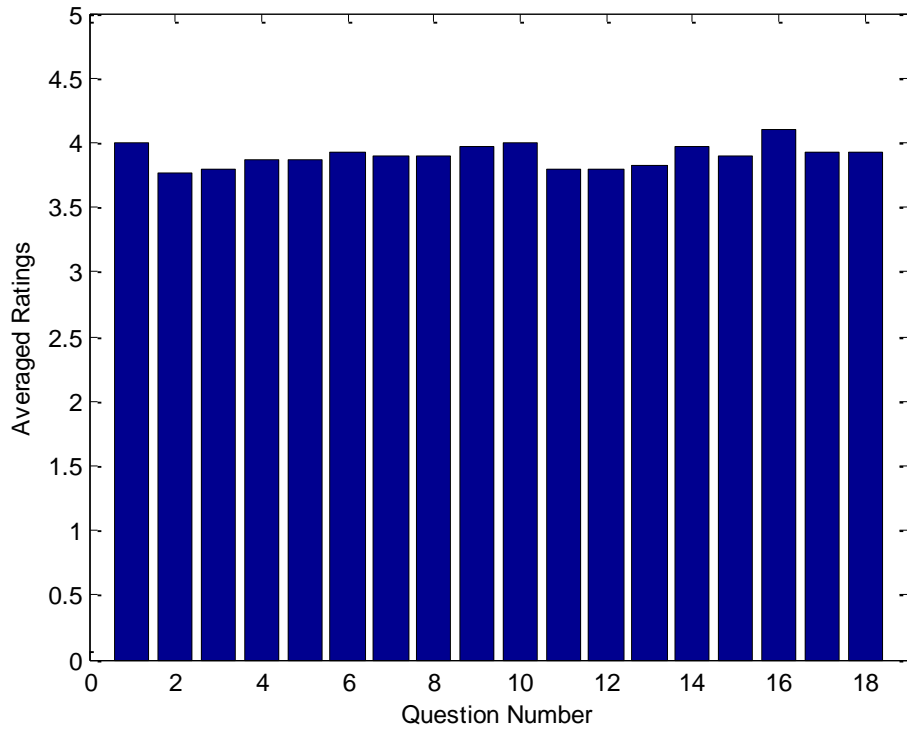


Figure 3(a). This bar diagram shows the averaged CTE ratings for each of the 18 questions. (39 respondents. Mean value = 3.9017, Standard deviation = 0.0881, Variance = 0.0073)

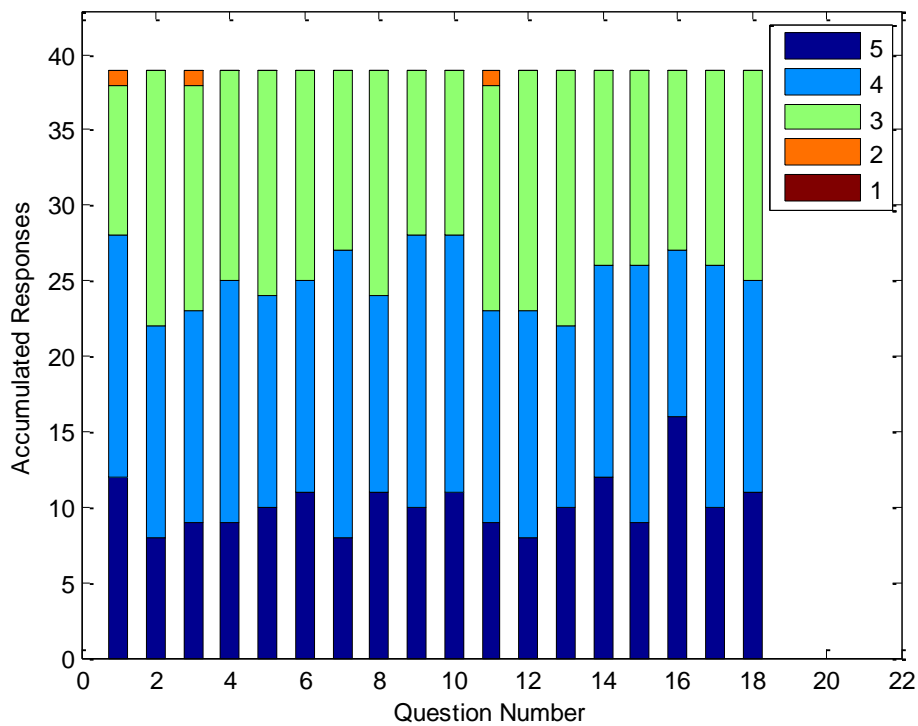


Figure 3(b). This stacked bar diagram shows the composition of ratings in Figure 3(a).



Figure 4 shows an example of low CTE rating achieved by the same teacher as in Fig.3. It is a freshman-level course also. Figure 4(a) shows a bar diagram of CTE ratings of the 18 questions. A total of 41 out of 51 students have participated in the CTE. This course has received a very low overall rating of 2.8482. The standard deviation of the 18 ratings is 0.1760 which indicates the fluctuation is similar to that of Fig.2(a). The composition of accumulated ratings for each question is shown in a stacked bar chart in Fig.4(b). All five ratings exist for each question. The difference is that there are more hostile or disgruntled students than friendly students. As shown in Fig.4(b), there are about seven hostile students who have likely ‘voted’ all 1’s to all questions. Only two friendly or loyal students have voted all 5’s. This example shows that seven hostile students versus two friendly students out of 41 respondents could drag the CTE rating down to 2.8482, nearly the lowest in school. It is seen that friendly students may improve the rating, but it is not as drastic as the ‘damage’ done by hostile students. To stay away from CTE rating disaster, it is more important for teachers to avoid hostile students than to have friendly students. As a result, it is not unfair to say that personality contest plays an important role in the CTE process.

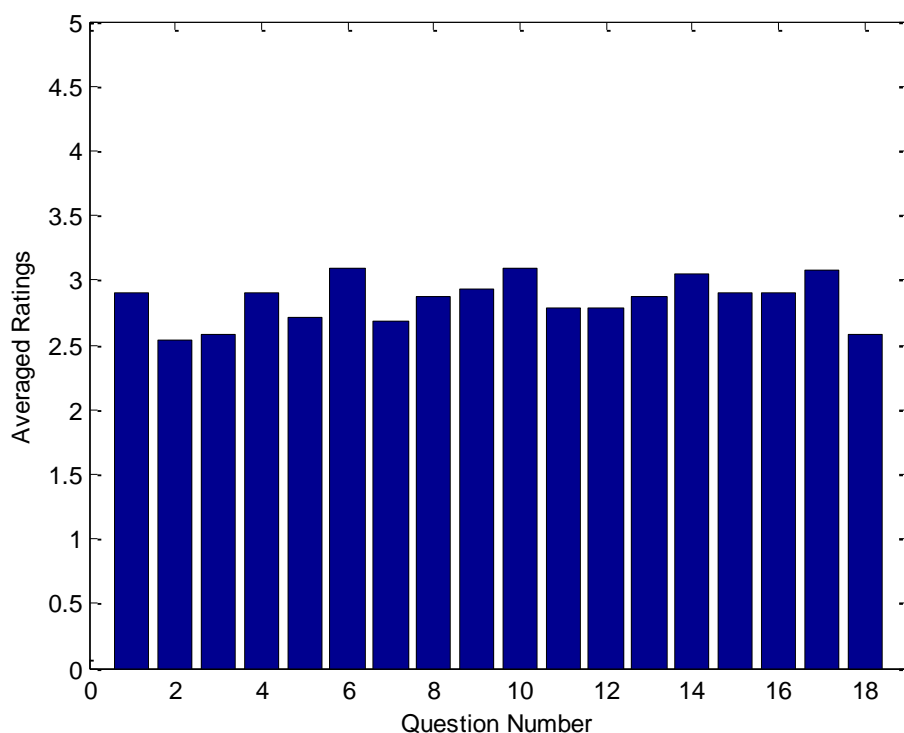


Figure 4(a). This bar diagram shows the averaged CTE ratings for each of the 18 questions. (41 respondents. Mean value = 2.8482, Standard deviation = 0.1760, Variance = 0.0293)

On the Significance of Mindless Phenomenon in Performing Course and Teaching Evaluation at a Technological University in Taiwan

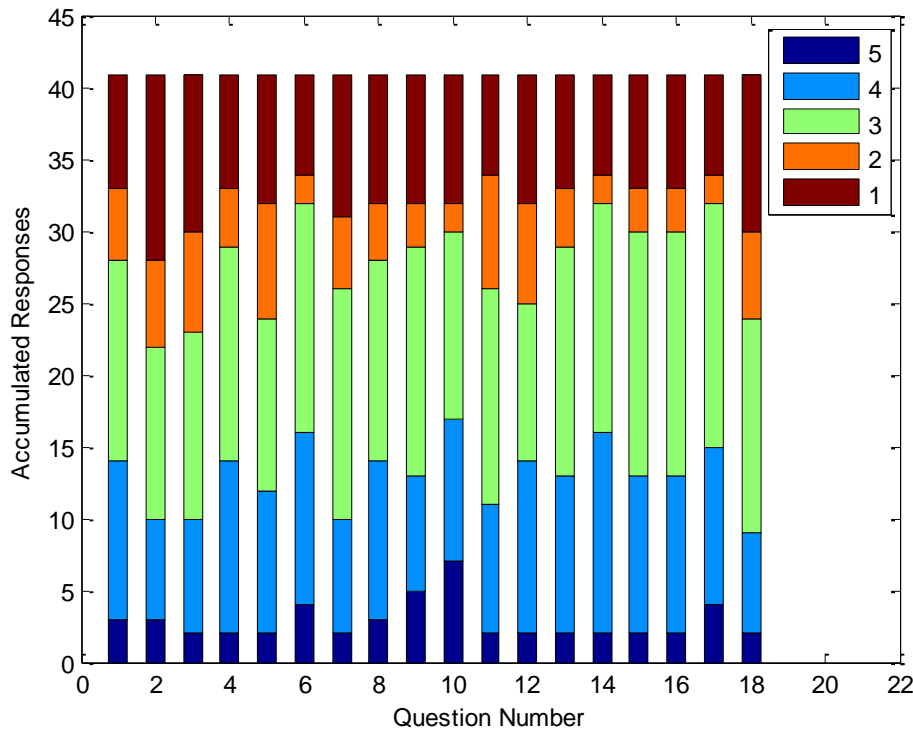


Figure 4(b). This stacked bar diagram shows the composition of ratings in Figure 4(a).

We have revealed and discussed the mindless phenomenon in actual CTE data. After gaining some insight and understanding, there are certainly ways to remedy the drawbacks and shortcoming. First of all, ratings with extreme outliers, such as all 1's and all 5's, can be removed or downgraded from the data. This will eliminate the unfair influence from both the mindless and emotional raters. This strategy is not new. It has been officially used to remove extreme scores in many athletic events. Another method is to use the computer to detect and downgrade raters with indiscriminate straight ratings, because they are likely done by mindless operator. To hamper the fastness of mindless operation, online CTE may present the questions on computer screen one at a time. This will force the raters to go slow.

The mindless phenomenon can be created by various kinds of students. The types may include: attention-deficit, in-a-hurry, lazy, emotional, does-not-care, or even too-dumb-to-read type. Their motives could be different too. Other than the corrective methods, the CTE questionnaire and process can be redesigned to counteract this mindless behavior by increasing the raters' attention or awareness. For example, the number of questions should be kept as concise as possible. Long list of questions, such as 20 questions in this study, may challenge the raters' patience to read and judge. Student education on the issues of CTE may help to increase their understanding. The improved CTE process should involve the students so that they may feel ownership. The CTE design can also notify the raters that '5' may not always be the best performance and '1' may not always be the worst performance. This may create confusion, but it will get many mindless raters' attention. These methods may help to reduce the extent of mindless phenomenon in CTE. In short, the CTE design needs to ensure that reliable data are gathered and thoughtful student feedback is facilitated.



5. Conclusion

Course and Teaching Evaluations (CTE) are now widely used in many universities internationally. In this paper, we have reported a previously unknown but significant phenomenon that can undercut the validity and accuracy of CTE. Based on actual CTE data collected at a technological university in northern Taiwan, we have shown that the great extent and significance of the mindless phenomenon do exist in the CTE process. For various reasons, mindless operation is that some students just voted the same numbers for all questions in the questionnaire. The findings are illustrated by using composition of accumulated ratings and graphical bar charts. We have also shown the voting patterns of the so-called hostile and friendly students who are also mindless operators in their behavior. To counteract the mindless phenomenon, some corrective measures such as computer data screening/correction have been suggested. Strategies such as CTE questionnaire design and student education and participation may be used to reduce the mindless phenomenon.



REFERENCES

- [1] Marsh, H. W. (1987) Students' evaluations of university teaching: research findings, methodological issues, and directions for further research, *International Journal of Educational Research*, 11(3), 253–388.
- [2] Wachtel, H. K. (1998) Student evaluation of college teaching effectiveness: a brief review, *Assessment and Evaluation in Higher Education*, 23(2), 191–210.
- [3] Cahn, Steven M. (1986) *Saints and Scamps: Ethics in Academia* (Totowa, NJ: Rowman & Littlefield, 1986).
- [4] Martin, C. et al. (1997) *The Use of Performance Indicators in Higher Education: The Challenge of the Quality Movement*, 3rd ed. (London: Jessica Kingsley Publishers, 1997).
- [5] Wilson, Robin (1998) New Research Casts Doubt on Value of Student Evaluations of Professors, *Chronicle of Higher Education* (Jan. 16, 1998): A12.
- [6] Yao, Y., Weissinger, E. and Grady, M. (2003). Faculty use of student evaluation feedback. *Practical Assessment, Research & Evaluation*, 8(21).
- [7] Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction: *American Psychologist*, v. 52, 1182-1186.
- [8] d'Appolonia, S. and Abrami, P. C. (1997). Navigating student ratings of instruction: *American Psychologist*, v. 52, 1198-1208.
- [9] Theall, M., Abrami, P.C., and Mets, L. M., (eds.) (2001). The student ratings debate: Are they valid? How can we best use them?: San Francisco, Jossey-Bass, *New Directions for Institutional Research*, n. 109.
- [10] Kealey, E. (2010). Assessment and evaluation in social work education: Formative and summative approaches. *Journal of Teaching in Social Work* 30, no. 1: 64–74.
- [11] Beran, T.N. and Rokosh, J.L. (2009). Instructors' perspectives on the utility of student ratings of instruction. *Instructional Science* 37, no. 2: 171–84.
- [12] Sproule, R. (2002). The underdetermination of instructor performance by data from the student evaluation of teaching. *Economics of Education Review* 21: 287–94.
- [13] Aleamoni, L.M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*. 13(2), 153-166.
- [14] Slade, P. and McConville, C. (2006) Student evaluation of teaching, *International Journal for Educational Integrity*, Vol.2, No.2, Dec. 2006, 43-59.
- [15] Isley, P. & Singh, H. (2005). Do higher grades lead to favorable student evaluations? *Journal of Economic Education*, 36(1), 29-43.
- [16] Chonko, L. B., Tanner, J. F., & Davis, R. (2002). What are they thinking? Students' expectations and self-assessment. *Journal of Education for Business*, 77, 271–281.
- [17] Svinicki, M. (2001). Encouraging your students to give feedback. In K. G. Lewis (Ed.), *Techniques and strategies for interpreting student evaluations* (pp. 17–24). San Francisco, CA: Jossey-Bass.
- [18] Chen, Y., Gupta, A., & Howshower, I. (2004). Marketing students' perception of teaching evaluations: An application of expectancy theory. *Marketing Education Review*, 14(2), 23–36.



- [19] Jones, J., Gaffney-Rhys, R., & Jones, E. (2012). Handle with care! An exploration of the potential risks associated with the publication and summative usage of student evaluation of teaching (SET) results. *Journal of Further and Higher Education*, (ahead-of-print), 1-20.

