



OPEN ACCESS

EURASIA Journal of Mathematics Science and Technology Education
ISSN: 1305-8223 (online) 1305-8215 (print)
2017 13(8):5837-5845
DOI: 10.12973/eurasia.2017.01033a



An Empirical Study on Student Evaluations of Teaching Based on Data Mining

Wei Zhang

Central China Normal University, China

Shiming Qin

Central China Normal University, China

Hanjun Jin

Central China Normal University, China

Jing Deng

Central China Normal University, China

Longkai Wu

Nanyang Technological University, Singapore

Received 9 May 2017 • Revised 26 June 2017 • Accepted 18 August 2017

ABSTRACT

Under the influence of big data, many fields have undergone tremendous changes. In the field of education, the data still contains a wealth of practical value, but the data mining and knowledge discovery is not enough, especially in the application of student evaluations of teaching (SET). In study, the K-means algorithm is used to cluster the data of three main teaching evaluation indexes (TEI) including individual background, course content, teaching method into high satisfaction degree (HSD), middle satisfaction degree (MSD), and low satisfaction degree (LSD). The logistic regression results showed that gender was a significant factor in students' evaluation of teachers and that there were potential connections between teaching evaluation and teachers' gender, age, and teaching content. In addition, the research shows that the effect of satisfaction degree on students' academic achievement is limited. The findings from this empirical study present a better understanding of reform of SET in higher education.

Keywords: student evaluation of teaching (SET), k-mean algorithm, the logistic regression, big data

INTRODUCTION

Student evaluations of teaching (SET) are the most commonly used tool in contemporary higher education (Knapper, 2001). Institutions generally use SET for the following three purposes: (a) to improve teaching quality through feedback, (b) to determine the promotion and tenure of the faculty, and (c) to demonstrate an institution's accountability (Kember, Leung, & Kwan, 2002). Big data analysis is a popular research topic of many fields. In the field of education, big data will rebuild the teaching evaluation method (Picciano, 2012) from the original experiential evaluation into scientific and quantitative evaluation and from single dimension evaluation to multi-dimension evaluation. With the support of data mining technology, SET based on educational big data will be more scientific and accurate.

© **Authors.** Terms and conditions of Creative Commons Attribution 4.0 International (CC BY 4.0) apply.

Correspondence: Shiming Qin, Master Degree Candidate, *National Engineering Research Center for E-Learning, Central China Normal University, Wuhan 430079, China.*

✉ shiming_qin@126.com

State of the literature

- Most of papers focus on the effect of SET on students' learning.
- Realizing that the quantity and quality of evaluation indexes are of vital importance to evaluate the effect of learning, researchers begin to pay attention to the selection of evaluation indexes.
- In order to make the conclusion more scientific and practical, researchers combine algorithms with SET in recent years.

Contribution of this paper to the literature

- The paper builds the system of indicators from multi-aspects and multi-dimensions in order to ensure an objective and scientific outcome.
- The paper analyses the effectiveness of the evaluation indexes.
- In this paper, the interaction of evaluation indexes is analyzed.

Many indexes influence students' evaluations of teachers in education, researchers lack specific selection process of effective teaching evaluation indexes (TEI). In the traditional teaching evaluation, the data source is single, and not comprehensive enough, which will affect the fairness of student evaluations of teaching. In this paper, we summarize three main indexes and ten sub-indexes combining the whole semester, and analyze the effectiveness of the evaluation indexes, as well as the interaction between indexes.

RELATED WORK

In the past few years, many researchers have put forward many methods for SET in universities, SET have provoked heated discussions in the following 4 aspects. (a) Transformation of SET paradigm: by expanding current criteria(e.g., educational scholarship, academic papers), schools could better inform the selection process, as well as promote evidence-based teaching practices, career promotion, and innovations in education (Kiersma et al., 2016); Data were collected using an internet survey designed to measure students' conceptions regarding five teaching dimensions referring to goals to be achieved, long-term student development, teaching methods, relations with students, and assessment (Alhija, 2016), It is found that long-term student development is the most important dimension for students to evaluate their teaching ability; Some researchers proposed that teaching should move toward standardization, professional teaching evaluation and teaching evaluation oriented (Appleton, Christenson, & Furlong, 2008), and institutions should pay attention to the use of basic data. (b) The influence of SET: through interviewed with students, researchers found that students were generally positive about SET, and they agreed that SET can provide accountability for teaching quality. They also believed that SET might influence future teaching practice when they perceived that the results would be used by the teachers (Spooren & Christiaens, 2017). Burden (2008, 2010)interviewed several award-winning teachers in higher education, and observed that they all attached great importance to SET, but only a few teachers considered that educational evaluation was helpful for them to improve their teaching quality. (c) Combining algorithms with SET: in the literature, (Vasconcelos, 2012) puts forward the key rule mining method based on data mining technology to find the meaningful association between data, and gives the framework of teaching evaluation data mining system. In order to study more scientifically and accurately, a study results generalized ordered logistic regression analyses show that male students express a bias in favour of male professors (Boring, 2017). (d) Indicators selection of SET: student ratings are an influential measure of teaching effectiveness, active participation by and meaningful input from students can be critical in the success of such teaching evaluation systems (Sojka, Gupta, & Deeter-Schmelz, 2002). Also, the differen5656t teaching dimensions that students value in male and female professors tend to match gender stereotypes (Eagly & Karau, 2002). (Spooren & Christiaens, 2017) recognize that based on their conceptions of their roles as evaluators of university professors, students might differ in their perceptions of teaching competence.

At present, the evaluation model of teaching evaluation focuses on the study of the whole teaching in our country, mainly focusing on theoretical research. It has an important guiding significance for the comprehensive development of education, however, it lacks of practical guidance causing the scope of the analytics is abstract. Aiming at this problem, this paper presents an empirical study on SET based on data mining.

Table 1. Teaching evaluation indexes

Construct	Substructure	Operation definition	Code	
Individual background	Gender	Male and Female		
	Age	30-60		
	Professional title	Professor		I_PRO
		Associated professor		I_APR
Lecturer			I_LEC	
Course content	Language skills	Explain clearly, highlight key knowledge and difficult knowledge	C_LS	
		Novel example, easy to understand		
		Pay attention to language learning methods		
	Cultural knowledge	Penetrate national and foreign cultural knowledge	C_CK	
		Expand international perspective		
		Cultivate cultural exchange strategy		
	Moral education	Guide students to establish a correct outlook on life	C_ME	
		Cultivate patriotism and national pride		
		Cultivate independent thinking ability		
	Assignment	Clear evaluation criteria, no bias	C_A	
Arrange the suitable amount of work assignments				
Teaching method	Classroom organization	Include written assignments and practical test	T_CO	
		Create pleasant classroom teaching atmosphere		
		Time allocation and reasonable teaching process		
	Activities	The link between teaching and learning is natural and smooth	T_A	
		Effective design of classroom teaching activities		
		Teachers and students' active participation, good interaction		
	Interaction	Considering individual differences	T_I	
		Online and offline learning		
		Students cooperative learning		
		Timely feedback		

METHODOLOGY

On the basis of the literature review, we focus on the potential indexes that may influence students' evaluation toward teachers, to validate whether they have an influence on teaching evaluation in our own research context. Moreover, we construct the three models based on logistic regression algorithm to verify the interaction between the various indexes.

Teaching Evaluation Indexes (TEI)

Teaching evaluation prediction is carried out under the premise of students' evaluation analytics, before that, selecting which teaching evaluation indexes for analyzing is critical. The research's object is a subjective feeling of the students to the teacher for the whole term, occurred on not only face-to-face learning in the classroom, but also other learning areas, such as Hstar cloud platform(a system that we designed for integration of education resources) .The items were worked to focus on satisfaction degree in education specifically. All survey items were answered using a five-point Likert scale. The course content included four questions on teachers' language skills, cultural knowledge, moral education and assignment. The teaching method comprised three questions each for classroom organization, activities and interaction. In addition, the teacher's personal background information is also an evaluation index, including gender, grade and professional title. In this paper, we have constructed three main indexes and 10 sub-indexes, with a total of 21 questions (except for the teacher's background), which are shown in [Table 1](#).

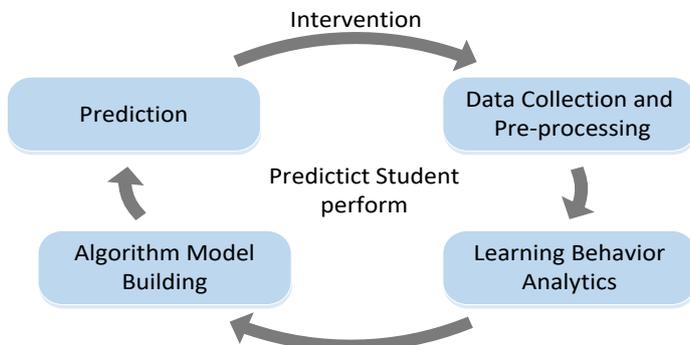


Figure 1. Research conceptual induction

Teaching Evaluation Model

Based on the extensive literature research, we construct our teaching evaluation model from the students' evaluation scores, as shown in [Figure 1](#).

Data Collection and Pre-processing

This phase needs to complete two tasks: data collection and data pre-processing, the data on Hstar cloud platform are stored in different data types in different types of databases, database type includes relational database and non-relational database.

In our university, administrators ask all freshmen to fill in teaching evaluation before final exam. Students who do not complete their teaching evaluation are not allowed to enter Hstar cloud platform, check their scores cannot register for the next semester, cannot print a diploma. We need to pre-process the data, for some reason, the data stored in the database has a small amount of incomplete and abnormal data. For example, some students may forget to register account, to fill in, or fill in the wrong way. According to the above indexes, we combine the data of the students' actual evaluation scores to pre-process the data.

Satisfaction Degree Analytics

After pre-processing, the next thing is analyzing data further. We need classify students according to certain criteria through clustering analysis, then compare teachers of different types of satisfaction degree, analyze the TEI of different teachers.

Clustering is an unsupervised data mining technique whose main task is to group the data objects into different clusters such that objects within a group are more similar than the objects in other clusters. K-means algorithm is very popular clustering technique for numerical data. We can use the k-means algorithm which can be used to divide teachers into different groups. The k-means algorithm depends on the given value for k when grouping data. In study, we use Gap statistic which is one of cluster selection criteria to identify the number of groups.

A description of Gap statistic is given below: Consider a data set $D_{ij}, i = 1, 2, \dots, m; j = 1, 2, \dots, n$, consisting of m data objects with values of n attributes. Assuming d_{xy} is the squared Euclidean distance between objects X and Y given by $d_{xy} = \sum (X_j - Y_j)^2$. If the data set has been clustered into k clusters, c_1, c_2, \dots, c_k , where c_i indicates the i th cluster, then $n_i = |c_i|$.

Let $D_i = \sum d_{xy}$, (where $x, y \in c_i$) is the sum of pair-wise distances for all points in cluster i and W_k is the collective within cluster sum of squares around the cluster means and is given by Eq. (1). $Gap_n(k)$ can be defined as the difference between expected and observed values of $\log(W_k)$ and given in Eq. (2). K can be taken for the value maximizing $Gap_n(k)$.

$$W_k = \sum_{i=1}^k \left(\frac{1}{2n_i}\right) D_i \quad (1)$$

$$Gap_n(K) = E_n^*\{\log(W_k)\} - \log(W_k) \quad (2)$$

Where E_n^* denotes the expectation under a sample size n from the reference distribution ion.

A brief formal description of k-means algorithm is given below:

Input: Data set consists of data objects, Number of clusters

Output: k clusters

Method:

1. Choose the objects at random from, as initial cluster centers.
2. **Repeat**
3. Assign each data object to the cluster to which its distance centers.
4. Update the cluster means, i.e. calculate the mean value of the objects for each cluster.
5. **Until** no data object changes its cluster membership or any other convergence criteria is met.

Three Model Building

In order to distinguish whether these differences of TEI were significant, we conducted a statistical analysis by using logistic regression.

This process contains three parts: build Model 1, analysis of teachers' individual background, including gender, age and professional title; build Model 2, analysis of teachers' personal background and course content that contains language skills, cultural knowledge, moral education and assignment; Model 3, analysis of all indexes.

Evaluation

Through the analysis of TEI, we can enlighten administrator to build a more important and meaningful evaluation system for teachers to conduct a more fair and scientific evaluation. Teaching evaluation is not a static, but a dynamic cycle process, school administrators, teachers, students, statisticians and other teams can analyze the evaluation process to improve the evaluation rules, enhance their ability and promote the teaching effect.

EXPERIMENTS AND DISCUSSION

We got related data generated from CCNU's Hstar cloud platform, used Gap statistic and k-means to cluster teachers and used Logistic Regression to explore the relationship between the indexes of teaching evaluation.

Data Collection and Pre-Processing

Higher Math, Moral Education, College English are all freshmen' compulsory courses in their first term at university. The data came from the Hstar Cloud Platform of 5402 freshmen who evaluated 149 teachers of the relevant courses in the fall semester of 2016. We used the eclipse tool to tally data in the database referring to 24 items (including teachers' individual background), then dealt with missed values and outliers. By data pre-processing, the study collect the effective data of 5204 freshmen and 142 teachers.

Data Analysis

Number of clusters to be made is identified using gap statistic as discussed above. The graphs clearly indicate that there are three clusters for teachers. **Figure 2** shows that gap value at cluster 3 that maximizes the $Gap_n(k)$

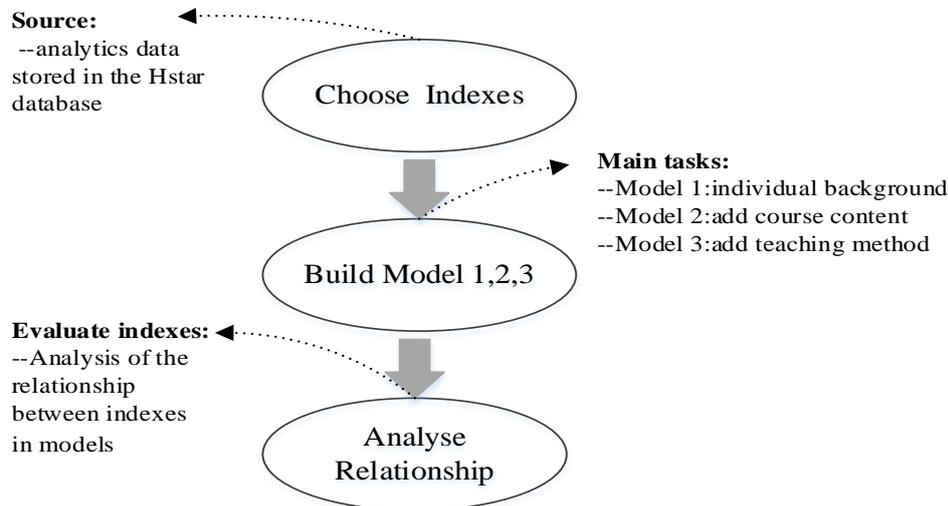


Figure 2. Three models building process

Table 2 presents descriptive statistics for the variables measured. Of the 42 teachers who reported that they received a high satisfaction degree evaluated by students, 58% were male. In contrast, 57% of the teachers who reported that they received a low satisfaction degree evaluated by students were female. With respect to age, the results showed an age difference favoring younger teachers—for example, more teachers under 50 reported that they received a high satisfaction degree than teachers over 50, and 31% of those who reported that they received a middle satisfaction degree and 35% of those who reported that they received a low satisfaction degree were in their 50s. Finally, teachers who reported to receive a high satisfaction degree showed higher scores on language skills (C_LS), cultural knowledge (C_CK), activities (T_A), and interaction (T_I), but in terms of moral education (C_ME), classroom organization (T_CO) and assignment (C_A), the scores are not too obvious different.

In sum, descriptive statistics showed a greater tendency of high scores among younger, male teachers, and those who showed higher scores on language skills, cultural knowledge, activities, and interaction. As for academic achievement, those who reported to receive a high satisfaction degree achieved similar, or somewhat higher, scores on usual achievement (U_A) and final exams (F_E) than others.

In order to distinguish whether these differences were significant, we conducted a statistical analysis by using logistic regression. In order to study how to get high evaluation, for simplicity, we combine MSD and LSD into the new group compared to HSD.

Table 2. Descriptive statistics

Indexes	Satisfaction Degree					
	HSD		MSD		LSD	
	Mean	SE	Mean	SE	Mean	SE
Female**	0.42		0.45		0.57	
Age**						
30-39	0.42		0.37		0.32	
40-49	0.31		0.32		0.33	
50-60	0.27		0.31		0.35	
Title***						
T_APR	0.35		0.32		0.30	
T_PRO	0.31		0.33		0.36	
T_LEC	0.34		0.35		0.34	
C_LS***	11.21	2.53	10.01	2.43	9.56	2.81
C_CK***	10.98	2.57	9.89	2.42	9.76	2.68
C_ME***	10.04	2.42	10.12	2.45	10.01	2.55
C_A***	9.95	2.45	9.82	2.31	9.71	2.32
T_CO***	10.21	2.21	9.69	2.32	9.81	2.43
T_A***	11.14	2.54	10.30	2.25	10.10	2.53
T_J***	10.65	3.09	9.98	2.84	9.91	2.14
U_A***	23.24	4.15	23.06	3.87	22.94	4.21
F_E***	78.87	7.89	77.99	6.34	77.14	8.26
N	42		72		28	

** , *** denote significant differences at $p < 0.01$ and $p < 0.001$ (two-tailed tests), respectively

In **Table 3**, the analysis examined TEI influencing the likelihood students in teaching evaluation would evaluate teachers with HSD. The first model consisted of a set of individual background indexes, among which only gender and age were significant predictors of the teachers’ likelihood of receiving high satisfaction degree. Regarding gender, male teachers were 23% (according to the value of OR) more likely to receive HSD than female teachers. Moreover, the odds of teachers in their 30s receiving HSD were 29% higher than teachers in their 50s or above, respectively. In terms of professional title, it is not a significant predictors from the model. The pseudo- R^2 of the first model was 0.065.

In Model 2, in which the course content of teachers were taken into account, significant differences in the likelihood of receiving HSD were observed between male and female gender. However, when comparing Model 2 with Model 1, the differences the 30s and 50s or older age range became non-statistically significant after controlling these indexes of course content. Among the course content, language skills and cultural knowledge were significant predictors of the teachers’ likelihood of receiving HSD by positively influencing the students’ decision. Model 2 increased the pseudo- R^2 from 0.065 to 0.136.

In Model 3, the teaching method were also included. Among them, the activities and interaction were found to be significant predictors of the teachers’ likelihood of receiving HSD. Even after controlling the teaching method and when comparing Model 2 with Model 3, the observed differences between male and female gender, remained significant in Model 3, though the effect is narrowing. So were the effects of language skills and cultural knowledge. Furthermore, one-unit changes in the course content indexes of language skills and cultural knowledge increased the odds of receiving HSD by 13% and 16%, respectively. In addition, the assignment was found to be a lightly significant predictor that increased the odds by 8% with a one-unit change. Finally, a one-unit change in

Table 3. Indexes influencing the likelihood of receiving HSD

Indexes	Model 1		Model 2		Model 3	
	B	OR	B	OR	B	OR
Female	-0.25*	0.77	-0.22*	0.80	-0.17*	0.84
Age						
30-39 (reference)						
40-49	0.14	1.15	0.09	1.09	0.08	1.08
50-60	-0.25***	0.79	-0.10***	0.90	-0.09**	0.91
Title						
T_APR (reference)						
T_APRO	0.01	1.01	0.00	1.00	-0.02	0.98
T_LEC	-0.03	0.97	-0.02	0.98	0.01	1.01
C_LS			0.12**	1.13	0.12**	1.13
C_CK			0.14***	1.15	0.15***	1.16
C_ME			0.06	1.06	0.09**	1.09
C_A			-0.01	0.99	0.01	1.01
T_CO					-0.06	0.94
T_A					0.10*	1.11
T_I					0.11*	1.12
Constant	1.23		-3.36***		-3.23***	
Log likelihood	-961.759		-896.341		-856.156	
Pseudo R^2	0.065		0.136		0.147	

***p < 0.001, **p < 0.01, *p < 0.05 (two-tailed tests)

interaction and activities, two teaching method indexes, increased the odds of receiving HSD by 12% and 11%. Overall, Model 3 increased the pseudo- R^2 from 0.136 to 0.147.

CONCLUSIONS

The current study provides empirical evidence that enables a better understanding of students' evaluation of teachers. The study not only found an effective way to cluster analysis, but also attempted to investigate the relationships among indexes pertaining to individual background, course content, and teaching method.

By analyzing the relationships among indexes more closely, we can gain insights into how teachers can improve their teaching quality and how administrators can make more scientific and fair teaching evaluation criteria. For example, the reasons why satisfaction degree of male teachers are higher than female teachers is probably due to students' impression of male stereotypes (of authoritativeness and knowledgeability). The teachers could also pay attention to language skills and interact with students in the teaching process, maybe, humorous teaching makes student easier to accept difficult knowledge. Teachers should also improve their cultural knowledge, so as to enable students to concentrate on learning and reduce the probability of skipping classes. More activities effective activities should be carried out as well.

SET is impacted by many factors. Although we analyzed three main indexes including ten sub-indexes, it is only a part of learning evaluation. The more indexes in the learning evaluation should be considered. Besides, it may be worth researching to further explore the relationship between satisfaction degree and students' academic performance, and we need to continue to explore.

ACKNOWLEDGEMENT

This research was partially funded by the University-Industry-Institute Base of Educational Information Software (2014-01), the Key Technologies and Standards of Public Service System for Teacher Training in Primary and Secondary Schools (2014BAH22F01).

REFERENCES

- Alhija, F. N.-A. (2016). Teaching in higher education: Good teaching through students' lens. *Studies in Educational Evaluation*. doi:10.1016/j.stueduc.2016.10.006
- Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools*, 45(5), 369-386. doi:10.1002/pits.20303
- Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145, 27-41. doi:10.1016/j.jpubeco.2016.11.006
- Burden, P. (2008). Does the use of end of semester evaluation forms represent teachers' views of teaching in a tertiary education context in Japan? *Teaching and Teacher Education*, 24(6), 1463-1475. doi:10.1016/j.tate.2007.11.012
- Burden, P. (2010). Creating confusion or creative evaluation? The use of student evaluation of teaching surveys in Japanese tertiary education. *Educational Assessment, Evaluation and Accountability*, 22(2), 97-117. doi:10.1007/s11092-010-9093-z
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological review*, 109(3), 573. doi:10.1037/0033-295X.109.3.573
- Kember, D., Leung, D. Y., & Kwan, K. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment & Evaluation in Higher Education*, 27(5), 411-425. doi:10.1080/0260293022000009294
- Kiersma, M. E., Chen, A. M., Kleppinger, E. L., Blake, E. W., Fusco, N. M., Mody, V., & Zavod, R. M. (2016). Evaluation of criteria utilized in the recognition of teaching excellence awards. *Currents in Pharmacy Teaching and Learning*, 8(4), 477-484. doi:10.1016/j.cptl.2016.03.003
- Knapper, C. (2001). Broadening our approach to teaching evaluation. *New Directions for Teaching and Learning*, (88), 3-9. doi:10.1002/tl.32
- Picciano, A. G. (2012). The Evolution of Big Data and Learning Analytics in American Higher Education. *Journal of Asynchronous Learning Networks*, 16(3), 9-20. doi:10.24059/olj.v16i3.2679
- Sojka, J., Gupta, A. K., & Deeter-Schmelz, D. R. (2002). Student and faculty perceptions of student evaluations of teaching: A study of similarities and differences. *College Teaching*, 50(2), 44-49. doi:10.1080/87567550209595873
- Spooren, P., & Christiaens, W. (2017). I liked your course because I believe in (the power of) student evaluations of teaching (SET). Students' perceptions of a teaching evaluation process and their relationships with SET scores. *Studies in Educational Evaluation*. doi:10.1016/j.stueduc.2016.12.003
- Vasconcelos, C. (2012). Teaching environmental education through PBL: Evaluation of a teaching intervention program. *Research in Science Education*, 42(2), 219-232. doi:10.1007/s11165-010-9192-3

<http://www.ejmste.com>