Statistical Analysis Based on Students' Evaluation of Teaching Data in an Application-Oriented University

Jing Zuo[1], Xing Wu[2], Geng-e Zhang[3], and Tongyao Huang[4(✉)]

[1] Quality Assessment Office, Nanning University, Nanning, Guangxi, China
[2] College of Artificial Intelligence and Information Engineering, Guangxi Electrical Polytechnic Institute, Nanning, Guangxi, China
[3] College of Transportation, Nanning University, Nanning, Guangxi, China
[4] College of General Education, Nanning University, Nanning, Guangxi, China
tongyaohuanghty@163.com

Abstract. This paper analyzed 184378 records of data of 2021–2022 fall students' evaluation of teaching (SET) in an application-oriented university. The research results demonstrate that the SET results have high reliability; SET is influenced by many different factors, but it is necessary to conduct further research to find out the variable with the greatest impact; SET results differ from supervisors' evaluation of teaching, and it is more importantly to design a SET scale with reasonability and high discrimination.

Keywords: Students' Evaluation of Teaching · Influence Factors · Consistency

1 Introduction

Student evaluation of teaching (SET) has become an important part of teaching quality monitoring system in colleges and universities [4]. With the advantages of objectivity, stability and operability [5], SET has been widely used in colleges and universities [6]. In recent years, the reliability and validity of SET have been questioned by researchers and writers. For example, some researchers claimed that due to the lack of evaluation ability, effective evaluation cannot be made by students in some aspects [11]; additionally, some argued that the evaluation score cannot fully reflect the real teaching effects [7]. Nowadays, colleges and universities have also carried out SET projects based on their own quality management practice. Does the result of SET have a certain degree of reliability and validity? Are SET scores affected by different factors? Is the result of SET consistent with that of supervisor evaluation of teaching? In order to answer these questions, this paper analyze the latest SET data of a university in Nanning concerning the reliability, influencing factors and consistency of supervisors' evaluation of teaching and SET.

2 Methodology

2.1 Data Collection and Sampling

Nanning University applies the internal quality management and diagnostic analysis system to carry out the supervisors', coworkers' and students' evaluation of teaching projects and store all the primary data. The author works in this university and able to download all the primary data about the supervisors' and students' evaluation of teaching. Nanning University is one of the first batch of pilot universities of applied technology in China and the pilot university for the transformation and development of newly-built undergraduate schools in Guangxi. Additionally, the reform of SET in Nanning university is representative in domestic application-oriented colleges and universities. Therefore, this paper may be a reference to the similar colleges and universities.

The SET of Nanning University has ten question (0.1 point each) from four indices 'teachers' ethics', 'teaching contents and methods', 'teaching management', and 'teaching effect, students' capacity and satisfaction'.

The author exported 184378 records of data of 2021–2022 fall SET, and set up database 1 with all the other essential data (relevant teachers' information, course information, students' grades). Database 1 includes fourteen variables (see Tables 1, 2, and 3). Besides, database 2 with the data of **supervisors'** evaluation of teaching had been set up to prepare for the research about the coherence of supervisors' evaluation of teaching and SET.

2.2 Data Analysis

This research applied the SPSS 19.0, and adopted the following analytical methods.

- Reliability and Validity Analysis: to check the reliability and validity of SET results;

Table 1. Variables utilized in research (student).

Variable	Grade				Major			Hometown	
Assignment	Freshman	Sophomore	Junior	Senior	Art	Science	Inter-discipline	Undergraduate	Upgrading from junior college to university
	1	2	3	4	1	2	3	1	2

Table 2. Variables utilized in research (course).

Variable	Type			Specialty		
Assignment	General education course	Professional course	Practical course	Compulsory course	General optional course	Profession optional course
	1	2	3	1	2	3

Table 3. Variables utilized in research (teacher).

Variable	Gender		Degree			Professional title				Type	
Assignment	Male	Female	Bachelor's degree	Master's degree	Doctor's degree	Professor	Associate professor	Lecturer	Teaching assistant	Full-time teachers	External part-time teachers
	1	2	1	2	3	1	2	3	4	1	2

- Correlation Analysis: to analyze the variables that may affect the results of SET. The variables include three types-- students, courses and teachers;
- Chi-Square Goodness of Fit Test: to analyze the consistency of SET and supervisors' evaluation of teaching.

3   Findings

This part, first, will provide the analysis on reliability and validity of SET results. Second, the relevant factors of SET scores will be analyzed. The last part will be the analysis on the consistency of SET and supervisors' evaluation of teaching.

3.1   Analysis on Reliability and Validity of SET Results

DeVellis (1991) suggest that Cronbach's alpha between 0.8 and 0.9 may be to be recognized as satisfactory. After calculating Cronbach's alpha coefficients of SET, it is obvious to see that the reliability and internal consistency SET of Nanning University is good ($\alpha = 0.9$).

According to the results of KMO test and Bartlett's test, the data of SET is suitable for factor analysis (KMO $= 0.953 > 0.900$, p $= 0.000 < 0.001$). After factor analysis, it is easy to find out that there is only one factor with an eigenvalue ≥ 1, and it explains 53.142% of 10 questions in SET; which illustrate that the discrimination of this scale is low.

3.2   Analysis on Relevant Factors of SET Scores

3.2.1   Analysis of the Relationship Between SET Scores and Students' Course Grades

Due to the lack of some students' course grades (absent or delay), the analysis of the relationship between SET scores and students' course grade will be discussed separately.

There is a significant positive correlation between SET scores and students' course grade as the Pearson's r $= 0.064$ and P $= 0.000$, which differs from Li (2008)'s research on H University with no correlation between SET scores and students' course grades. The possible reason for the different research results is that the previous research analyzed the correlation between the average score of SET and the average score of total courses grades of all the courses taught by the same teacher. The course grade difference and SET score difference between different students have been neutralized when calculating

the average score of them. However, this research is to analyze the correlation between the individual SET score of each student and the individual course grade, so the result of this research could more reliable.

In order to better analyze the correlation between SET scores and courses grades, the courses grades have been divided into four groups: A (90–100), B (80–89), C (70–79), D (less than 70). One way ANOVA has been used to analyze the difference within and between groups. According to the average score of SET scores of each group (A 90.03, B 89.90, C 89.65 and D 88.94), it is obvious to see that the average values of SET scores of the four groups rank from highest to lowest in the order of A, B, C, D. Through multiple comparisons of one-way ANOVA, the following results can be obtained: there is no significant difference between group A and group B, while significant differences are able to be witnessed between group A and group C, between group A and group D, and between group C and group D as well.

Students may not grade SET according to their courses grades as they can only check the grades after finishing SET process. Therefore, the significant positive correlation between SET and course grade is most likely because students gain from teachers' classes (got A or B), and they prefer to score high in SET. On the contrary, students with low grades (got C or D) may not satisfy with teachers' teaching methods, teaching management, teaching effect, or self-achievement, so they score low in SET.

3.2.2 Analysis of the Relationship Between SET and Other Variables

The analysis of the correlation between SET and three types of variables (student: grade, major, hometown; course: type, specialty, period, credit; teacher: gender, degree, professional title, age, type) is shown in Table 4.

Significantly correlation can be seen between SET score and most of the variables, except specialty (compulsory or optional course) and credit of course.

In terms of student factors, the higher grade the students in, the higher the SET score; students studying interdiscipline or science score higher than those studying art. These two findings are similar to the previous studies [9]. It is noteworthy that students upgrading from junior college to university score higher than undergraduate students, which may because they enter university in a different way, so they have relatively strong learning motivation, and a higher sense of self-acquisition and course satisfaction.

In terms of course factors, professional courses and practical courses have higher SET scores; courses with higher credits have higher SET scores. These two findings echo the previous studies [9].

In terms of teacher factors, female teachers get higher SET scores than male teachers; teachers with higher degree get higher scores in SET. However, the situation is different when it comes to professional title and age, teachers with lower titles (lecturer or teaching assistant) get higher SET scores; the younger teachers get higher scores than older ones. Full-time teachers score higher than external part-time teachers. The reason why these findings differ from the previous studies may be teachers with high professional title (professor or associate professor) spend more time on research instead of teaching, while those ones with lower title focus on teaching and perform well in SET; the younger teachers are able to integrate with students, so they may be able to get higher scores in SET.

Table 4. Correlation coefficient between variables and SET results

Types of variables	Name of variables	Correlation coefficient	P
Student	Grade	0.050^{**}	0.000
	Major	0.010^{**}	0.000
	Type	0.036^{**}	0.000
Course	Type	0.027^{**}	0.000
	Specialty	0.003	0.243
	Period	−0.010	0.722
	Credit	0.016^{**}	0.000
Teacher	Gender	0.007**	0.005
	Degree	0.024**	0.000
	Professional Title	0.022**	0.000
	Age	−0.019**	0.000
	Type	−0.020**	0.000

Table 5. Correlation analysis between SET and the different indices of SET questions

Index	Teachers' ethics	Teaching content and method	Teaching management	Teaching effect, students' capacity and satisfaction
Pearson	0.712^{**}	0.945^{**}	0.677^{**}	0.904^{**}
P	0.000	0.000	0.000	0.000

3.2.3 The Correlation Analysis Between SET and the Different Indices of SET Questions

According to the correlation analysis between SET and the different indices of SET questions (see Table 5), the correlation coefficient between teaching content and method and SET is the highest, followed by teaching effect, students' capacity and satisfaction, then teachers' ethics, and finally teaching management. It shows that in order to accredit by students, teachers should mainly improve the teaching content and methods, and then improve the teaching effect, students' sense of capacity and satisfaction with the courses.

3.3 Analysis on the Consistency of SET and Supervisors' Evaluation of Teaching

In order to analyze the consistency of SET and supervisors' evaluation of teaching, Chi-Square Goodness of Fit Test has been utilized in this research. The analysis idea of chi square goodness of fit test is as follows: first, assume that the proportion of students with

Table 6. Distribution ratio of SET and supervisors' evaluation of teaching.

Grade	SET		Supervisors' evaluation of teaching	
	Number of teachers	Percentage	Number of teachers	Percentage
A (90–100)	509	86.56	125	21.3
B (80–89)	77	13.10	448	76.2
C (70–79)	1	0.17	11	1.9
D (<70)	1	0.17	4	0.7
Total	588	1	588	1

different course grades (A, B, C, D) in supervisors' evaluation of teaching is our expected proportion; second, count the proportion of students with different course grades (A, B, C, D) in SET; finally, compare the proportion of SET with the proportion of supervisors' evaluation of teaching and evaluate its consistency. If these two proportion fit well, which indicates that the consistency of these two teaching evaluation is high, and vice versa.

Chi-square Goodness of Fit Test needs a categorical variable when conducting it. So, the scores of SET and supervisors' evaluation of teaching have been divided into four classification variables A (90–100), B (80–89), C (70–79), D (<70). Because the supervisors' evaluation of teaching failed to achieve full coverage, only 588 teachers covered by both SET and supervisors' evaluation of teaching were selected. Through frequency statistics, the proportion of students with different course grades (A, B, C, D) in supervisors' evaluation of teaching is A 21.3%, B 76.2%, C1.9% and D 0.7% (see Table 6). Weight cases have been utilized to give the supervisors' evaluation of teaching data different weights. The SET data have been processed (click analyze → nonparametric tests → legacy dialogues → chi-square), and then combined with the weights of supervisors' evaluation of teaching to get the analysis results ($\chi 2 = 11206.308$, P = 0.000). The results indicate that the data of SET is inconsistent with the data distribution of supervisors' evaluation of teaching. A lack of consistency within these two data can be witnessed.

The consistency between SET and supervisors' evaluation of teaching is low, and the score of SET is significantly higher than that of supervisors. The reasons for this are twofold: (1) the perspective of evaluation is inconsistent as students and supervisors use different indicators and scales to evaluate teaching; (2) teachers and students are interest-concerned parties, teachers determine students' course grades, so students are more inclined to give teachers higher score in SET.

4   Discussion and Recommendations

After the statistical analysis of SET in Nanning University, some conclusion and recommendation will be provided in this part.

4.1 The Results of SET Affect by Many Factors, but it is Necessary to Conduct Further Research to Find Out the Variable with the Greatest Impact

According to the findings, the results of SET were highly related to grade, major, hometown and course grade of students; credit and type of courses; gender, age, degree, professional title and type of teachers. However, from the value of Pearson correlation coefficient, the values are not high (all less than 0.1). This indicates that even if these variables are significantly correlated to SET, it is not enough to predict the teaching evaluation results according to these variables. Therefore, there must be variables that have a greater influence on the results of SET.

In previous research, Wu (2010) suggested that the determinants of SET scores are students' gains and satisfaction; Mao (2009) believed that the SET scores mainly depend on the teachers' abilities; additionally, Clayson (2022)'s recent research claimed that the results of SET are mainly related to whether students like the teacher or not. According to the correlation analysis between SET and the different indices of SET questions, the correlation coefficient between teaching content and method and SET is the highest, with teaching effect, students' capacity and satisfaction follow by it. Are these factors also affecting SET scores? The discrimination of SET scale is not high, these factors have not been subdivided into relevant dimensions and have not been analyzed in depth. Further research should be done to go deeper into this problem.

4.2 SET Differs from Supervisors' Evaluation of Teaching, and it is More Importantly to Design a SET Scale with High Discrimination

Under the influence of the concept of multiple quality evaluation, most of the colleges and universities have been conducting supervisors', coworkers' and students' evaluation of teaching projects. In previous research, Wu (2010) used correlation analysis and regression model to test the difference between SET and experts' evaluation of teaching, and suggested that there was a significant difference between experts' evaluation of teaching and students'; Sun & Zhai (2009) adopted Kappa Statistic to analyze the consistency of evaluation guiding team's, experts' and students' evaluation of teaching, and considered that the results of the evaluations were inconsistent. This research creatively adopts the Chi-square Goodness of Fit Test to test the consistency between SET and supervisors' evaluation of teaching, and the test result echoes previous studies.

However, in terms of the practical implications, the author believes that the inconsistency is appropriate. If the results of students', supervisors' and coworkers' evaluation of teaching, are consistent, it is no need for the colleges and universities to conduct multiple quality evaluation. The multiple quality evaluation can help to evaluate the teaching quality from different perspectives, especially help to find out different problems from different perspectives and to figure out proper solutions. Therefore, instead of improving the consistency of multi-party evaluation, the future focus of teaching evaluation is going to develop a reliable multi-party evaluation scale, which can carry out teaching evaluation from a differentiated dimension and perspective.

5   Conclusion

This study confirmed that the variables such as grade, major, credit, teacher's gender, teacher's age and so on were significantly correlated with SET. Moreover, the results of SET are not consistent with those of supervision. The key point of the reform is to set up more reasonable and more differentiated scales to carry out SET from different perspectives. This provides a basis for the subsequent research on the influencing factors of SET and the development of SET scale.

References

1. Clayson, D. 2022. The Student Evaluation of Teaching and Likability: What the Evaluations Actually Measure. *Assessment & Evaluation in Higher Education* 47 (2): 313–326.
2. DeVellis, R.F., and C. T. Thorpe. 1991. Scale development: Theory and applications. Sage Publications.
3. Li, Z., and F.W. Jiang. 2018. Research on students' evaluation of teaching in colleges and universities based on the data analysis. *China University Teaching* 04: 76–81.
4. Ma, X.L., K.D. Zhong, and L.C. Liu. 2014. Research of the validity about SETQ by big data analysis. *China Educational Technology* 10: 78–84.
5. Mao, F.F. 2009. On discrepancy of students' evaluation of teaching-based on data analysis. *Joural of Higher Education* 03: 87–91.
6. Pounder, J.S. 2007. Is Student Evaluation of Teaching Worthwhile? An Analytical Framework for Answering the Question. *Quality Assurance in Education* 15 (2): 178–191.
7. Shevlin, M., P. Banyard, M. Davies, and M. Griffiths. 2000. The validity of student evaluation of teaching in higher education: Love me, love my lectures? *Assessment & Evaluation in Higher Education* 25 (4): 397–405.
8. Sun, C.Y., and J.J. Zhai. 2009. The Kappa Statistics on the Consistency of Evaluation Guiding Team's, Experts' and Students' Evaluation of Teaching. *Heilongjiang Research on Higher Education* 3: 41–43.
9. Wu, P. Q. and X. H. Chen 2010. The Statistical Analysis of Undergraduate Students' Evaluation of Teaching and its Reform Approach—Based on the Statistical Analysis of Students' Evaluation of Teaching Scores in a University in Beijing. *Higher Education Exploration* 3:78–81+91.
10. Wu, P.Q. 2010. An Empirical Analysis of the Difference between Experts' and Students' Evaluation of Teaching in Colleges and Universities. *Journal of National Academy of Education Administration* 10: 71–77.
11. Yang, L.P., and Y.T. Zhang. 2009. What Affects the Reliability and Validity of Students' Evaluation of Teaching in Colleges and Universities. *China University Teaching* 3: 72–74.