

【原著】

An Investigation into the Links between Course Evaluation Survey Results and Student Learning Outcomes in English Classes

Arthur Rutson-Griffiths

英語の授業における授業評価アンケートの
結果と学習成果の関係性

アーサー・ラットソングリフィス

Abstract

End-of-semester course evaluation surveys are a common method of collecting data for improving teaching at universities worldwide. The Bunkyo English Communication Centre at Hiroshima Bunkyo Women's University uses its survey as part of a Faculty Development system, but its validity has not previously been analysed. Data from the teaching evaluation section of the survey were compared to class grade point average and statistically-significant positive relationships were found. This suggests that the survey does produce meaningful data, but a number of caveats are identified and confirms the need to use multiple measurements of teaching effectiveness.

概 要

学期末の授業評価アンケートは、国内外の大学において教育の質向上のためのデータ収集方法として広く用いられている。広島文教女子大学の文教英語コミュニケーションセンターにおいてもFDシステムの一環としてアンケートが用いられているが、その有用性が分析されたことは今までない。アンケート内の教員評価に関する項目のデータとクラスごとのGPAの関連性を分析した結果、統計的に有意な正の相関が見つかった。これは、授業評価アンケートのデータが役に立つことを示唆するものである。しかしながら、使用にあたって注意しなければならない点もいくつか確認されたことから、教育効果を評価するには複数の方法が必要であるということが言える。

Introduction

Hiroshima Bunkyo Women's University is a small, private university in the suburbs of Hiroshima. It has approximately 1,200 students spread across five departments: Early Childhood Education, Welfare, Psychology, Nutrition and Global Communication. The Bunkyo English Communication Centre (BECC) was established in 2008 to provide compulsory General English classes to all Bunkyo students and subsequently, with the establishment of the Global Communication Department in 2010, skill- and domain-based English courses.

The BECC currently employs 11 full-time lecturers and provides 72 class periods per week across 18 courses per semester. An integral part of the BECC's mission is to provide high-quality classes; the first of the BECC's mission statements reads 'We are here to provide relevant, challenging and engaging courses teaching real world language skills.' Helping teachers develop the ability to deliver such courses is one of the main goals of the BECC Faculty Development (FD) system, which comprises class observations, a teaching portfolio and results from the BECC course evaluation survey.

Gathering feedback from students is a common method used by universities worldwide to evaluate teaching effectiveness (Berk, 2005). The BECC course evaluation survey was introduced from Kanda University of Foreign Studies, along with the BECC curriculum and teaching staff, when the centre opened in 2008. A number of updates have been made to the survey structure and question wording over the intervening period. As it is an integral part of the BECC FD system and contributes to the attainment of one of the BECC's mission statements it plays an important role in the evaluation of the BECC's performance but its relationship to student performance is unknown. If there is a positive relationship between course evaluation ratings and student achievement it could be considered a useful component of the toolkit for evaluating course effectiveness, and conversely its usefulness should be questioned if this relationship does not exist.

Methods

The data used for this analysis are course evaluation survey results and class grade point average (GPA).

Course evaluation survey

The BECC course evaluation survey is administered at the end of each semester in all classes to all students. It is an anonymous online survey consisting of a number of different sections (see Table 1) and is given to students before they receive their grades. Sections 1-3 are common to all courses whereas section 4 is created by the course teacher(s) to receive feedback on specific aspects of the course. There are also pages for identifying which course is being evaluated (not included in Table 1).

Table 1: BECC course evaluation survey sections

Section number	Section title	Number of questions
1	Student self-evaluation	2 closed
2	Teacher evaluation	12 closed, 1 open
3	Generic course evaluation questions	12 closed
4	Course-specific evaluation questions	Varies

This analysis uses data from section 2. The 12 statements shown in Table 2 are rated by students on a scale of 1–6 (strongly disagree to strongly agree) and a weighted average calculated. The

An Investigation into the Links between Course Evaluation Survey Results and Student Learning Outcomes in English Classes

average of these 12 weighted averages is used in this analysis as a measure of teacher performance as perceived by students.

Table 2: BECC course evaluation survey teacher evaluation statements

Number	Statement
Q1	The teacher was ready to start the class on time.
Q2	The teacher was well prepared for lessons.
Q3	The teacher was fair to all the students.
Q4	Appropriate action was taken against disruptive students.
Q5	The teacher clearly explained activities and materials.
Q6	The teacher was approachable and available to provide help during and after lessons.
Q7	The teacher responded well to questions and comments.
Q8	The teacher gave appropriate feedback about assigned work.
Q9	The teacher gave timely feedback that I could use to improve my next assignment.
Q10	The teacher was enthusiastic in teaching.
Q11	The teacher created motivation and interest in the course.
Q12	Overall, the teaching in the course was effective.

Grade point average

The GPA for each class in the corresponding semester was calculated based on students' final letter grades. Numerical values are given in Table 3. Students who receive a D grade have not submitted satisfactory work but may apply to receive additional work to receive a passing grade in the next semester. Students who receive an E grade have attended fewer than 65% of classes and must retake the class. Students with an E grade are not included in the analysis.

Table 3: Grade point average weights at Hiroshima Bunkyo Women's University

Letter grade	Weight
S	4
A	3
B	2
C	1
D	0
E	0

GPA was chosen as it is a measurement of student achievement common to all courses and captures the results of all student assessments. The method of assessment differs across courses, but generally includes a combination of multiple-choice tests, reports, presentations, homework completion and class participation.

Results

Course evaluation results and GPA were compiled for two semesters: the second semester of the 2017–18 academic year and the first semester of the 2018–19 academic year. Semesters in different years were chosen to broaden the number of different students in the analysis.

Courses are provided by the BECC to students of all years. They can generally be divided into General English courses for students of all departments and skill- and domain-based English courses for students in the Global Communication department. Many courses are also divided into more than one class with either the same teacher for all classes or different teachers. General English classes are also usually streamed by ability, along with some Global Communication classes.

Before analysing the data further, a scatterplot was created, and two outlying data points were identified and removed. Classes with responses from two or fewer students were also removed. This left 69 classes and course evaluation survey data from 1,219 students. A detailed summary is not provided in order to preserve teacher anonymity.

Table 4: Overview of class data¹⁾

Year	General English classes		Global Communication classes	
	Classes	Responses	Classes	Responses
1	18	443	16	249
2	14	350	7	88
3	0	0	10	64
4	0	0	3	18
Total	32	793	36	419

The data were analysed in various combinations as shown in Table 5 and Figure 1.

Table 5: Correlation statistics for average course evaluation score and class GPA

Condition	Classes	Responses	R-squared	P-value
A) All data	69	1,219	0.23	<0.01
B) All Global Communication classes	36	419	0.12	0.04
C) All General English classes	32	793	0.18	0.01
D) All 1 st year classes	34	692	0.11	0.06
E) All 2 nd year classes	22	445	0.41	<0.01
F) Taken more than one BECC teacher's class	51	776	0.27	<0.01

Condition F) is an analysis of the data from all classes in which students have taken more than one BECC teacher's class. This includes first-year Global Communication students, who take four

1) General English classes taken only by Global Communication students are counted as Global Communication courses.

An Investigation into the Links between Course Evaluation Survey Results and Student Learning Outcomes in English Classes

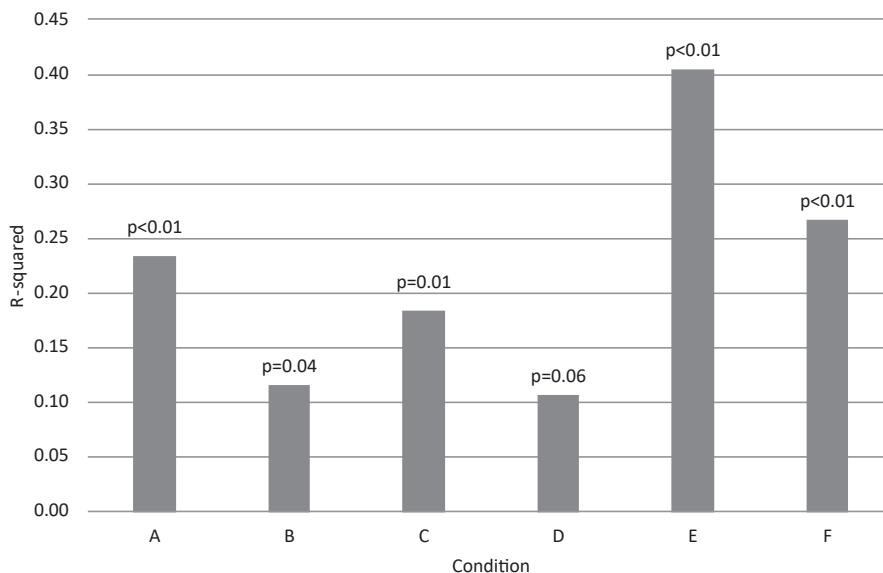


Figure 1: R-squared and p-values for conditions presented in Table 5

BECC classes per week, and all second- to fourth-year students.

R-squared is “a statistical measure of how close the data are to the fitted regression line” (Minitab Blog Editor, 2013) that is always between 0 and 1. The higher the value, the more of the variation in the data is explained by the model. P-values are also reported; values of 0.05 or less are commonly regarded as statistically significant (Minitab Blog Editor, 2015).

Remarks

The analysis conducted showed that there is a correlation between class GPA and BECC course evaluation survey teacher evaluation section average scores. The correlation was weakest for first-year classes (D) and strongest for second-year classes (E). This could be due to the fact that all non-Global Communication students have only one BECC teacher in their first year and so do not have any other teachers with whom to make a comparison when rating their teacher, in contrast to second-year students.

The result for all Global Communication students (B), who have the most experience of BECC classes, does not fit this hypothesis. However, this result could be affected by the number of elective third- and fourth-year classes included in the analysis. Although classes with two or fewer students were removed, class numbers are still typically small and thus individual students' grades and survey responses can make a big difference to the averaged data, weakening the correlation between them. This explanation may be supported by the fact that the R-squared value for students who have experience of more than one BECC teacher's class (F), which includes a number of large, second-year General English classes, is more than double that of Global

Communication only.

This could be an important point to bear in mind when considering using student course evaluation data for evaluating teacher performance. If students do not have much experience of the subject matter or of the style of teaching employed, they may not be able to provide reliable data that teachers can use to improve their courses.

Although this analysis produced statistically-significant correlations between GPA and course evaluation results, there are many caveats to be aware of. The R-squared statistic shows how much of the variation in GPA is explained by course evaluation results. This means that even for the largest value of 0.41, or 41%, the remaining 59% of variation in GPA is still explained by other factors. Similarly, though the results are not reported here, a preliminary investigation into differences between the two semesters of data showed that correlations in the 2018–19 semester 1 data were generally stronger than in the 2017–18 semester 2 data, the reasons for which are not clear. This should be considered another note of caution when considering using course evaluation scores to evaluate teacher performance, particularly from only one course or semester; a pattern of data across time and courses is likely to be more meaningful.

There is also the caveat that correlation does not equal causation. It is possible that, for example, motivation is a lurking variable, i.e. it is an unmeasured variable that is affecting the relationship (Minitab, 2016). Highly-motivated students may be more likely to both attain a high grade and view their teacher favourably, giving them high ratings on the course evaluation survey with the reverse also being true. The fact that many classes in the analysis are already streamed by ability is also likely to affect the results. Ideally, course evaluation responses would also be linked to individual students in order to examine the relationship more closely, but students may be less likely to give honest responses if the survey were not anonymous.

Similarly, GPA may not be the best measure of student outcomes. On the positive side, it summarises student performance across a whole semester and is based on a number of assessments and observations of students' English ability. On the negative side, it loses detail that could be useful for this kind of analysis by categorising students. For example, a student whose total semester score is 80% and a student whose score is 89% both get an A and contribute a GPA of 3 to the class average. With GPA being averaged by class, there is also no way to measure individual student progress, which is ultimately what many teachers and other stakeholders would like to know about their effectiveness.

Conclusion

This paper has described an analysis of the relationship between course evaluation survey scores and student achievement as measured by GPA. A statistically-significant, positive relationship was established between the two measures that supports the validity of the survey, but there are a number of caveats that should caution teachers or administrators against using individual class

An Investigation into the Links between Course Evaluation Survey Results and Student Learning Outcomes in English Classes

results in isolation for evaluation of teachers or courses. Used as one part of a system of evaluation, as described by Berk (2005), course evaluation surveys can be a source of valid data for improving English classes at the university level.

References

- Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17(1), 48–62.
- Minitab. (2016). What is a lurking variable? Retrieved from <http://support.minitab.com/en-us/minitab/17/topic-library/modeling-statistics/regression-and-correlation/regression-models/what-is-a-lurking-variable/>
- Minitab Blog Editor. (2013, May 30). Regression analysis: How do I interpret R-squared and assess the goodness-of-fit? [Blog post]. Retrieved from <http://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit>
- Minitab Blog Editor. (2015, March 19). Understanding hypothesis tests: Significance levels (alpha) and P values in statistics [Blog post]. Retrieved from <http://blog.minitab.com/blog/adventures-in-statistics-2/understanding-hypothesis-tests-significance-levels-alpha-and-p-values-in-statistics>

—平成31年 1 月25日 受理—