

Effects of Chinese University Students' Academic Achievement and Course Satisfaction on Their Ratings of Teachers' Instruction

Yun Mo

Tsinghua University

Xibin Han

Tsinghua University

Yingqun Liu

Tsinghua University

Abstract: *The purpose of this study was to investigate the relationships between students' perception of teachers' instruction, course satisfaction, and student academic achievement. The data were extracted from TsingHua Education Online (THEOL) Management Information System (MIS) and analyzed via structuring equation modeling. The researchers hypothesized student ratings of instruction were affected by their achievement and course satisfaction. The results confirmed a positive relationship between course satisfaction and student ratings of instruction, but did not find direct effect of student academic achievement on their ratings of teachers' instruction. This suggests that course satisfaction is a more important factor than academic achievement when students rate their teachers' instruction. The finding also suggests that the student ratings of instruction may be an objective and acceptable performance indicator for teachers' instruction in a course.*

Keywords: Student ratings of instruction, course satisfaction, evaluation, English achievement

1. Introduction

Teachers' instruction is a critical factor in improving student course performance. As the primary objective of teaching is student learning, evaluating the impact of teachers' instruction has been considered as a major indicator of teaching quality in higher education institutions (Zerihun, Beishuizen, & Van Os, 2012). The following questions should be asked when looking at teachers' instruction: How to evaluate teachers' instruction?;

Are students qualified to rate their teachers and their instruction they receive?; and Do students rate teachers on the basis of received grades? Student ratings of instruction (SRI) are debated by a large population of faculties. McKeachie (1997) summarized the research studies on the validity of student ratings by stating that "student ratings are the single most valid source of data on teaching effectiveness" (p. 1219). However, Cashin (1989) pointed out that students are not qualified to judge a number of factors that characterize exemplary

instruction. For example, the appropriateness of the instructors' objectives, the relevance of assignments or readings, the degree to which subject matter content was balanced and up-to-date, and the degree to which grading standards were unduly lax or severe can affect instructional quality.

The purpose of the study reported in this paper is to investigate the relationships between student ratings of teachers' instruction, course satisfaction, and student English achievement. The research hypothesized SRI were affected by students' achievement and course satisfaction. Furthermore, gender and origin-based differences in the student academic achievements were examined.

2. Literature Review

Evaluation of teaching and instruction based on quantitative student opinion surveys has been an accepted in the higher educational landscape internationally (Chalmers 2007; Harvey 2003; Knapper & Wright 2001). Today, student ratings of instruction are widely employed in China. Yet, student feedback-based evaluation remains largely an unwelcome fringe dweller in current academic life, often responded to with skepticism and unease (Darwin, 2012; Edstrom, 2008).

Students' descriptions of the characteristics of effective teachers are commonly used to develop measures of instructional quality. Zerihun et al. (2012) make two interrelated key assumptions on the evaluation of instruction effectiveness. One is that teachers are seen to have an influence on student learning. The second one is the students' capacity to provide teachers with feedback to be used for instruction improvement.

Some researchers mentioned the use and misuse of SRI. Relating students' evaluations of teaching to student achievement as an approach

to validate SRI has produced inconsistent results. A study in a major Canadian university investigates teachers' attitudes about SRI. Teachers tend to agree that the student rating practice is an acceptable means of assessing general teaching quality and is useful to administrators in making summative decisions on evaluating instruction. However, they consider student evaluations only marginally valuable in their impact on enhancing their instructional practices (Beran & Rokosh, 2009). Zabaleta (2007) indicates that there is a moderate correlation between low grades and low evaluations, but no correlation between high grades and high evaluations when all cases are considered together. Zabaleta (2007) suggests the results of SRI should not be used in critical personnel decisions such as retention, tenure, and promotion of faculty because the relationship between SRI and the actual merits of teaching performance had not been clearly identified, and explained with a sound theory of instruction effectiveness. Stehle, Spinath, and Kadmon (2012) test a hypothesis that the strength of association of SRI and student learning varies with the criteria used to indicate student achievement. Results show a strong positive association between SRI and the practical examination, but no significant correlation between SRI and multiple-choice test scores.

Overall, according to reviews of the literature conducted by Aleamoni (1999) and Arreola (1995) well-developed, tested, student rating forms of teaching effectiveness exhibited both reliability and validity. Other more critical studies, however, highlighted the shortcomings of student evaluations of faculty and questioned their validity and usefulness (Zabaleta, 2007). Evident is that research findings are inconsistent and there is need for more empirical research to fully understand the relationship of SRI and student academic achievement.

3. Methodology

3.1. Data Sources

Traditional evaluation is based on time, cost, and quality. In this digital era, online evaluation has replaced traditional evaluation in most higher education institutions in China. Online evaluation is a type of evaluation where the medium is through Information and Communication Technology (ICT).

The data for this study were extracted from TsingHua Education Online (THEOL) Management Information System (MIS). THEOL MIS is developed by the Education Technology Institute of Tsinghua University (THETI). THETI has investigated the e-educational management model; dedicated to promoting the high quality, effective management level with the information technology for more than 10 years. Till now, THETI has been collaborated with more than 30 universities in China to serve to E-management support in Higher Education.

3.2. Sample

In this study, the researchers selected one sample university in Beijing because the questionnaire of that university to measure student ratings of instruction was revised

from Marsh's (1982) Students Evaluation of Educational Quality (SEEQ) questionnaire based on their school's characteristics by institutional researchers. The SEEQ has been extensively tested and used in more than 50,000 courses with over one million students at both the graduate and undergraduate levels (Marsh & Hocevar, 1991). Because the SEEQ has been widely used, these measures valid representation of the students' perceptions of their teachers' instruction.

This particular university co-constructed by the Chinese Central Government and Beijing Municipal Government, and administrated mainly by Beijing Municipal Government, has nearly half of its students from Beijing. Students in the sample were enrolled in one public course entitled "College English, listening and speaking" in the 2012 fall semester and finished a student ratings of instruction survey (N=235). There were more male students than female students (60.9% vs. 39.1%) in the sample. An equal balance of Beijing and Non-Beijing students, at 53.6% and 46.4% respectively, existed in the sample (see Table 1). The balanced sample size has its significance and the research findings can be generalized to similar higher institutions in Beijing metropolitan area.

Table 1. Frequency of gender and origin of student

<i>Variable</i>		<i>N</i>	<i>Percentage (%)</i>
Gender	Male	143	60.9
	Female	92	39.1
Origin of Student	Beijing	126	53.6
	Non-Beijing	109	46.4
Total		235	100

3.3. Methods

This study used the structural equation modeling (SEM) to test the hypothesis that students' English achievement was affected by students' perception of their teachers' instruction and students' course satisfaction. Structural equation modeling (SEM) is a statistical method that grows out of multiple regression, and takes a confirmatory approach to analyze a structural theory of the relationships of some variables of interest (Byrne, 1998). Causal relationships among theoretical variables are represented by a series of structural equations (regression equations), and the relationships are represented in a model. Then, the goodness of fit between the theory-based model and the data (usually variance-covariance matrix) is tested statistically. If the goodness of fit is adequate, then the postulated model is plausible and consistent with the data. If the goodness of fit is poor, then the model is not plausible and needs to be rejected or re-specified (Byrne, 1998).

The standard approach to estimate a SEM distinguishes two steps (Anderson & Gerbing, 1988). In the first phase of the two-step approach, measurement models for all latent variables in the model are estimated. In SEM, researchers are interested in the causal relationship between latent variables and observed variables, which is called the measurement model. Measurement model is tested by confirmatory factor analysis, which is specified based on theory or empirical studies that generate a statistical representation about the relationships among latent and observed variables. In the measurement model, both dependent and independent latent variables are specified (Mustafa, 1999). The second step is the structural part of the SEM. This structural part specifies the relationships between the exogenous and endogenous variables.

In the present study, the researchers estimate a latent variable structural equation model using LISREL 8.8 computer program (Jöreskog & Sörbom, 2005). Structural equation modeling is an especially appropriate method for analyzing non-experimental data. In addition to parameter estimates, the program provides fit indices to assess how well the model fits the data. Such fit indices make it possible to evaluate the adequacy of the theoretical model in explaining the data (Bollen, 1988; Schumacker & Lomax, 2004).

3.4. Measured Indicators

3.4.1. English achievement. English is the most popular foreign language in China, however, English is composed of an alphabet while Chinese characters are logograms. The difference makes many Chinese students have difficulties in learning English, thus teaching and learning English in China are challenges for both Chinese teachers and students. In this study, student English achievement was measured by their final test scores which were extracted from the achievement module of MIS. The scores ranged from 0 to 100 where the passing grade was 60. The mean score of 235 students was 73.37 (SD=10.57).

3.4.2. Students satisfaction and ratings of instruction. At the end of the semester, students were asked to rate teachers' instruction. The survey total included 10 questions chosen from the SEEQ (Marsh, 1982). Responses were in the form of Likert scale for agreement with each statement such as 1 = "strongly disagree" through 5 = "strongly agree." Items and descriptive statistics are shown in Table 2.

Table 2. Descriptive statistics

<i>Items</i>	<i>Mean</i>	<i>SD</i>
1 My language ability was improved in the class	2.54	.649
2 Course materials and instruction methods are well prepared and carefully explained	2.61	.577
3 The instructor's presentations were clear and understandable; complex or important points were summarized or emphasized in class.	2.59	.617
4 The instructor spoke with expressiveness and variety in tone of voice and spoke at an appropriate pace	2.59	.644
5 The instructor focused on classroom management; class sections were well organized	2.53	.655
6 The instructor's teaching and instruction methods were appropriate, which inspired students passion of learning	2.48	.675
7 The instructor was knowledgeable; course content was fulfilled and innovative	2.56	.667
8 The course is instructive, help us relate theories and concepts to practical issues	2.55	.628
9 The amount of information covered in this course and the pace were reasonable	2.53	.642
10 Readings, homework, feedback, etc., contribute to appreciation and understanding of the subject	2.52	.662

4. Results

4.1. Measurement Model

After preliminary analyses and empirical research, a confirmatory factor analysis was conducted to test the measurement model and assessed the validity of the constructs. The researchers hypothesized that 10 items had two dimensions. The first factor was composed by the first item: My language ability was improved in the class. The other nine items were significantly loaded on the second factor.

The fit indices for the measurement model were high, indicating a well-fitting model in which data fit into the hypothesized model, $\chi^2(35) = 65.37$, $p < .01$. The goodness-of-fit index (GFI) is .95, and the adjusted goodness-of-fit (AGFI) index is .92. The comparative fit index (CFI) was .99. The root-mean-

square error of approximation (RMSEA) is .06. Overall, these fit indices indicate a theoretically sound measurement model that explained the data well. All models were estimated using the maximum likelihood method. The measurement model had a good fit and high to moderate loadings (see Figure 1). Two latent factors represent these 10 items well. The first latent factor reflected students' course satisfaction and the second latent factor reflected student ratings of teacher's instruction.

4.2. Structural Model

To further investigate the relationships between student English achievement and their course satisfaction and ratings of instruction, the researchers hypothesized a structural model (see Figure 2).

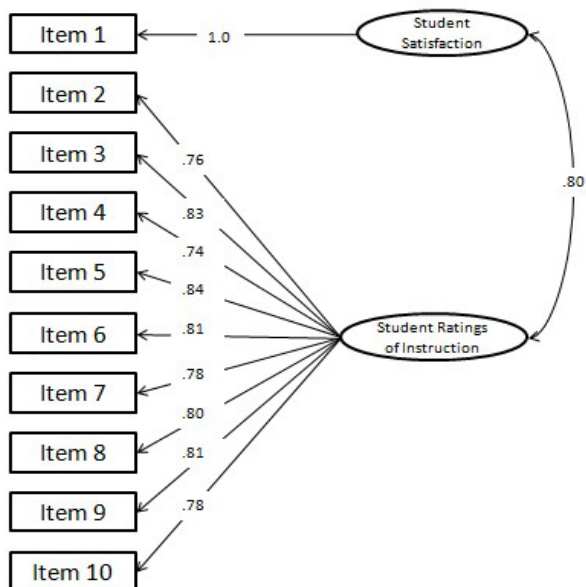


Figure 1. Loadings for measure model.

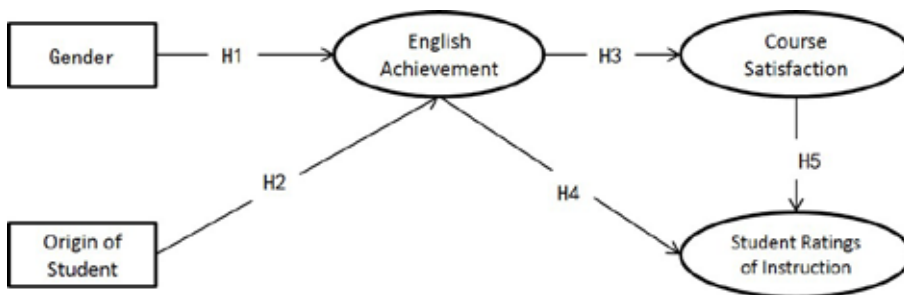


Figure 2. Hypothesized model.

Five null hypotheses were:

H1₀: There is no difference of student English achievement between male and female students.

H2₀: There is no difference of student English achievement between Beijing and non-Beijing local students.

H3₀: There is no effect of test score teacher gave to students on student course satisfaction.

H4₀: There is no effect of test score teacher gave to students on student ratings of teachers' instruction.

H5₀: There is no effect of student course satisfaction on SRI.

Table 3. Overall item correlation matrix.

	Score	Item 1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10	Gender
Score	1											
Item1	.134*	1										
Item2	.074	.620**	1									
Item3	.123	.652**	.649**	1								
Item4	.093	.543**	.587**	.635**	1							
Item5	.139*	.633**	.621**	.704**	.644**	1						
Item6	.117	.551**	.617**	.684**	.586**	.723**	1					
Item7	.100	.650**	.570**	.635**	.578**	.688**	.655**	1				
Item8	.106	.649**	.619**	.610**	.627**	.669**	.645**	.624**	1			
Item9	.236**	.677**	.564**	.686**	.575**	.665**	.620**	.592**	.693**	1		
Item10	.118	.637**	.594**	.646**	.589**	.636**	.601**	.575**	.632**	.679**	1	
Gender	.260**	.049	.121	.014	.027	-.012	.010	.009	.035	.123	.011	1
Origin of Student	-.308**	-.021	.020	.014	.027	.039	.043	.035	.052	-.040	.039	.029

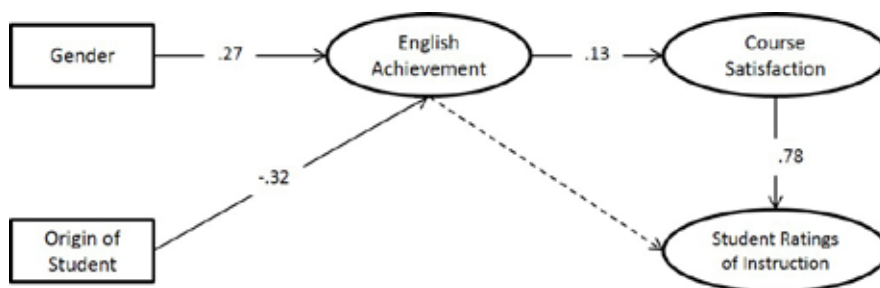


Figure 3. Final structural model.

An overall correlation matrix of all scales is listed in Table 3. The fit indices for the final model were high, indicating a well-fitting model in which data fit well to the final model, $\chi^2(64) = 94.09, p < .001$. The goodness-of-fit index (GFI) was .94, and the adjusted goodness-of-fit (AGFI) index was .92. The comparative fit index (CFI) was .99. The standardized root mean square error of

approximation (RMSEA) was .045. Overall, these fit indices indicate a theoretically sound model that explained the data well. Then, the researchers examined direct and indirect effects for significance and magnitude (see Table 4 and Figure 3).

Hypothesis 1: The researchers rejected $H1_0$ and concluded that there was difference

of student English achievement between male and female students ($t=4.51$, $\beta=0.27$, $p<0.01$); female students' English achievement was better than male students.

Hypothesis 2: The researchers rejected H_{2_0} and concluded there is difference of student English achievement between Beijing and non-Beijing local students ($t=-5.29$, $\beta=-0.32$, $p<0.01$); non-Beijing students' English achievement was better than Beijing students.

Hypothesis 3: The researchers rejected H_{3_0} and concluded there is effect of test score teachers gave to students on student

course satisfaction ($t=2.07$, $\gamma=0.13$, $p<0.05$); high achievement students tend to have higher course satisfaction than low achievement students.

Hypothesis 4: The researchers accepted H_{4_0} and concluded there is no direct effect of test score teachers gave to students on student ratings of teachers' instruction ($t<1.96$, $p>0.05$).

Hypothesis 5: The researchers rejected H_{5_0} and concluded there is effect of student course satisfaction of SRI ($t=12.66$, $\gamma=0.78$, $p<.001$); course satisfaction was a strong predictor on SRI.

Table 4. Direct, indirect, and total effects ($n=235$) on SRI

	<i>English Achievement</i>			<i>Student Satisfaction</i>			<i>Student Ratings of Instruction</i>		
	<i>Direct</i>	<i>Indirect</i>	<i>Total</i>	<i>Direct</i>	<i>Indirect</i>	<i>Total</i>	<i>Direct</i>	<i>Indirect</i>	<i>Total</i>
Gender	.27		.27						
Origin of Student	-.32		-.32						
English Achievement				.13		.13		.11	.11
Student Satisfaction							.78		.78

Figure 3 displays the final structural model of relationships between student academic achievement, course satisfaction, and the results of SRI. Controlling for gender and origin of students, the results of SRI were strongly significantly predicted by student course satisfaction, but was not directly influenced by student achievement. Only an indirect effect ($\gamma=0.11$, $p<.05$) of achievement on the results of SRI was found by the mediator of course satisfaction. Hence, course satisfaction played an important role on measuring SRI.

5. Conclusion and Discussions

The study examined an important topic of relationship of SRI and student academic achievement and confirmed that well-developed, tested questionnaires were good measures of teaching effectiveness (Aleamoni, 1999; Arreola, 1995; Marsh, 1997; Zerihun et al., 2012). This study also reinforced the importance of student course satisfaction. Despite the common myths about students highly rating teachers when teachers grading students gave higher test scores, the research

results found that student course satisfaction was an important mediator between SRI and academic achievement. There was no direct effect of students' academic achievement on SRI, but an indirect relationship between ratings and grades occurred because effective teaching leads to learning that leads to student achievement and satisfaction. Students reported that teacher behaviors, classroom management, course content, the amount of work required, how much they felt they have learned, and the difficulty of the material have had an impact. They can reliably answer questions about the quality of instruction, the value of readings and assignments, the clarity of the instructor's explanations, the instructor's availability and helpfulness, and many other aspects of the teaching and learning process. The study has implications for practice and supports the structures that would facilitate SRI and teaching effectiveness in higher education in China.

6. Limitations and Future Research

SRI should be a multidimensional perspective and no single criterion of effective teaching is sufficient (Marsh, 1983; 1987; Marsh & Dunkin, 1997; McKeachie, 1997). The full SEEQ comprises of items grouped into nine dimensions of teaching: learning value, instructor enthusiasm, organization/clarity, group interaction, individual rapport, breadth of coverage, examinations/grading, assignments/readings, and workload/difficulty. Marsh (1982) suggested that each of these categories contained three or four questions. The survey used in this study only selected 10 questions, so the multidimensionality of teaching and SRI was not examined. This study focused on the effects of students' academic achievement and course satisfaction on their ratings of teachers' instruction. The results showed that SRI were not directly related to students' received grades and

concluded that student ratings were very reliable. Zabaleta (2007) suggested that the results of SRI should not be used in critical personnel decisions such as retention, tenure, and promotion of faculty, unless they were properly interpreted within a sound theory of teaching effectiveness. This study did not address these issues, but these ideas should be further studied to provide solid empirical results. The researchers suggest future studies using a long version of SEEQ and testing the multidimensionality of effectiveness of instruction to explore a deeper understanding of teaching effectiveness and SRI.

References

- Aleamoni, L.M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*, 13(2), 153-166.
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103, 411-423.
- Arreola, R. A. (1995). *Developing a comprehensive faculty evaluation system: A handbook for college faculty and administrators on designing and operating a comprehensive faculty evaluation system*. Bolton, MA: Anker Publishing Co.
- Beran, T. N., & Rokosh, J. L. (2009). Instructors' perspectives on the utility of student ratings of instruction - Springer. *Instructional Science*, 37(2), 171-184.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.
- Byrne, B. M. (1998). *Structural equation modeling with lisrel, prelis, and simplis*. Mahwah, NJ: Erlbaum.
- Chalmers, D. (2007). *A review of Australian and international quality systems and indicators of learning and teaching*. Strawberry Hills: The Carrick Institute for Teaching and Learning in Higher Education.
- Cashin, W. E. (1989). *Defining and evaluating college teaching*. IDEA Paper No. 21. Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.
- Darwin, S.(2012). Moving beyond face value: re-envisioning higher education evaluation as a generator of professional knowledge. *Assessment and Evaluation in Higher Education*, 37(6), 733-745.
- Edstrom, K.(2008). Doing course evaluation as if learning matters most. *Higher Education Research and Development*, 27(2), 95-106. doi:10.1080/07294360701805234
- Harvey, L.(2003). Student Feedback [1]. *Quality in Higher Education*, 9(1), 3-20.
- Jöreskog, K. G., & Sörbom, D. (2006). LISREL 8.8. Chicago, IL: Scientific Software International.
- Knapper, C., & Wright, W. A. (2001). Using portfolios to document good teaching: Premises, purposes, practices. *New Directions for Teaching and Learning*, 88, 19-29.
- Marsh, H. W. (1982). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology*, 52, 77-95.
- Marsh, H. W. (1983). Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics. *Journal of Educational Psychology*, 75, 150-166.
- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11, 253-388.
- Marsh, H. W., & Dunkin, M. J. (1997). Students' evaluations of university teaching: A multidimensional perspective. In R. P. Perry & J. C. Smart (Ed.), *Effective teaching in higher education: Research and practice* (pp. 241-320). New York: Agathon.
- Marsh, H. W., & Hocevar, D. (1991). Students evaluations of teaching effectiveness: The stability of mean ratings of the same teachers over a 13-year period. *Teaching & Teacher Education*, 7, 303-314.
- McKeachie, W. (1997). Student ratings; the validity of use. *American Psychologist*, 52(11), 1218-1225.
- Mustafa, B. (1999). A brief introduction to

- structural equation modeling techniques:
Theory and application. *ERIC Document
Reproduction Service No. 438 300.*
- Schumacker, R. E., & Lomax, R. G. (2004).
*A beginner's guide to structural equation
modeling.* Mahwah, NJ: L. Erlbaum
Associates.
- Stehle, S., Spinath, B., & Kadmon, M.
(2012). Measuring teaching effectiveness:
Correspondence between students'
evaluations of teaching and different
measures of student learning. *Research in
Higher Education*, 53(8), 888-904.
- Zabaleta, F. (2007). The use and misuse of
student evaluations of teaching. *Teaching
in Higher Education*, 12(1), 55-76.
doi:10.1080/13562510601102131
- Zerihun, Z., Beishuizen, J., & Van Os, W.
(2012). Student learning experiences
as indicator of teaching quality.
*Educational Assessment, Evaluation and
Accountability*, 24, 99-111.

Acknowledgements

This study was funded by the project
of Tsinghua University for heritage and
innovation of culture, Technology Enhanced
Teaching Quality Control in Universities:
Problems and Strategies numbered with
2012WHQN016.

Contact the Authors

Yun Mo

Tsinghua University
Email: moyun2012@gmail.com

Xibin Han
Tsinghua University
Email: hanxb@tsinghua.edu.cn

Yingqun Liu
Tsinghua University
Email: liu-yq@tsinghua.edu.cn

