Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia - Social and Behavioral Sciences 69 (2012) 990 – 996

Procedia
Social and Behavioral Sciences

International Conference on Education and Educational Psychology (ICEEPSY 2012)

Who participates (not)? A non-response analysis on students' evaluations of teaching

Pieter Spooren*[a], Francis Van Loon[a]

University of Antwerp, Sint Jacobstraat 2, Antwerp B-2000, Belgium

Abstract

Students' evaluation of teaching (SET) is the most common way of assessing teaching quality at universities. Recently, online SET (instead of the traditional paper-and-pencil questionnaires) has become the established means for gathering students' opinions. Obviously, online surveys have great advantages, but their greatest challenge remains the low response rates that might affect the representativeness of the sample. This paper contains a non-response analysis on online SET-data using multilevel logistic models. At the student level, course grade, program level, and the number of course evaluations a student was asked to complete are significant predictors for participation. Student's gender and study domain are not significant. The variance component of the course level is estimated to be 0. The implications for future research and SET-practice are discussed.

Selection and/or peer-review under responsibility of Dr. Zafer Bekirogullari of Cognitive – Counselling, Research & Conference Services C-crcs.

Keywords: students' evaluation of teaching; teacher evaluation; validity studies; higher education; teaching quality

1. Introduction

Students' evaluation of teaching (SET) is the most common and sometimes sole way of assessing teaching quality at universities. This widespread use has much to do with their (apparent) ease of collecting the data and presenting and interpreting the results (Penny, 2003). On top of that, students are considered to be important stakeholders when gaining insight into the quality of teaching in a course (Seldin, 1993). Whereas SET in the early days had a mainly formative character, in the 1970s they quickly became an important instrument in faculty personnel decisions as well (Galbraith, Merrill & Kline, 2011). More recently, SET-procedures are included as a key mechanism in internal quality assurance processes to prove an institution's performance in accounting and auditing practices (Johnson, 2000). The main purpose of SET is thus threefold: 1) improving teaching quality, 2) appraisal exercises (tenure/promotion decisions), and 3) institutional accountability (demonstrating adequate procedures for ensuring teaching quality) (Kember, Leung and Kwan, 2002). It is clear that, in most institutions, SET nowadays are used for both formative (i.e., students' feedback for the improvement of teaching) and summative (i.e, mapping teaching competence for administrative decision-making and institutional audits) reasons (Arthur, 2009; Burden, 2008; Emery, Kramer & Tian, 2003). The economy of these surveys thus has to be high (Braun & Leidner, 2009),

1877-0428
Selection and/or peer-review under responsibility of Dr. Zafer Bekirogullari of Cognitive – Counselling, Research & Conference Services C-crcs.
doi:10.1016/j.sbspro.2012.12.025

since institutions need (by preference comparable) information for different types of course as much and as quickly as possible.

Recently, online SET (instead of the traditional paper-and-pencil questionnaires) has become the established means for gathering students' opinions. The reasons for this shift onwards online SET are obvious: greater accessibility to students, no disturbance of class time, accurate analysis of the data, guaranteed student's anonymity, less susceptible to faculty influence, lower costs and less time consuming for administrators (Anderson, Cain & Bird, 2005; Ballantyne, 2003; Bothell & Henderson, 2003; Bullock, 2003). Still, some fear that SET results in this way are more traceable and can be consulted by almost everyone (Gamliel & Davidovitz, 2005). Concerning the question whether or not the shift onwards online SET affects SET-scores, several authors (Leung & Kember, 2005; Liu, 2006; Venette, Sellnow & McIntyre (2010) found no significant effect between SET-scores from a paper-based evaluation and an online evaluation. Barkhi & Williams (2010) in their turn noted that, at the aggregated level, online-based SET-scores are lower than SET-scores from a paper-based survey. However, when controlling for course and instructor variables, these differences disappeared. In sum, one can conclude that online evaluations do the job as well, and provide similar results compared to the traditional paper-and-pencil evaluation forms. Obviously, online surveys have great advantages, but their greatest challenge remains to increase the low response rates compared to the more traditional paper-and-pencil questionnaires (Dommeyer, Baum, Hanna & Chapman, 2004). Students are among a heavily surveyed group, and selection effects might bias SET-outcomes. In an online SET-environment, students are less accessible and have more freedom to decide whether or not to participate in an evaluation procedure. As a consequence, one could call into question the reliability and the validity of online SET, since self-selection bias might affect the representativeness of the sample. It is important to profile those students participating in digital SET, and compare them with those who decided not to co-operate. In this way, the representativeness of the respondents for the whole population can be defined and more insight can be gained in students' motives to (not) take part in SET.

2. Objectives

This paper contains a non-response analysis on online SET-data from the University of Antwerp (Faculty of Social and Political Sciences, 2478 questionnaires, 895 students, 24 courses) and takes into account several course, teacher and student characteristics that might influence a student's decision to participate in an online SET-procedure. This makes it possible to sketch a profile of respondents and non-respondents and to draw up some suggestions for future research and practice on this topic.

3. Methods

Instrument. In the present study, the SET37-questionnaire (Mortelmans & Spooren, 2009) which represents 12 quasi-balanced scales (37 items) was administered by means of an online survey. Students received an email (including a login code) that invited them to participate in the evaluation procedure for each particular course in which they were enrolled. In the next three weeks students received two more email reminders.

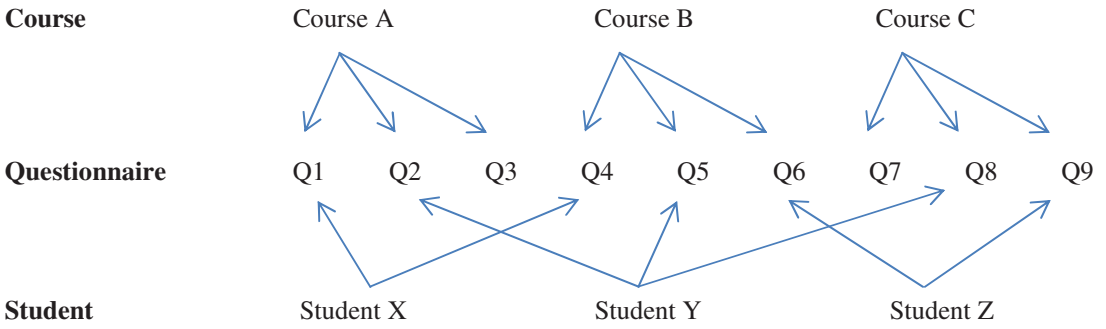
Participants. SET were administered at the Faculty of Social and Political Sciences of the University of Antwerp during the spring semester of the 2010-2011 academic year. 895 students (both graduates and undergraduates) enrolled in 24 courses were asked to participate in the evaluation surveys. Table 1 contains some descriptive statistics concerning the variables (i.e. student characteristics and course characteristics) that were included in the present study. Since a number of students were enrolled in two or more courses (see 'number of evaluations' in Table 1), the total number of surveys was 2478. The overall response rate was 26.8%.

Table 1: Descriptive statistics for the variables included in the study (2478 questionnaires, 895 students, 24 courses)

Variable		**N**		**%**	
Response		2478		100	
	response		*663*		*26.8*
	no response		*1815*		*73.2*
Course Type		2478		100	
	required course		*1778*		*71.8*
	elective course		*700*		*28.2*
Course grade		2478		100	
	no grade		*344*		*13.9*
	fail		*320*		*12.9*
	pass		*830*		*33.5*
	pass – good grade		*585*		*23.6*
	pass- excellent grade		*399*		*16.1*
Student's gender		895		100	
	male		*372*		*41.6*
	female		*523*		*58.4*
Student's study domain		895		100	
	Social Sciences		*785*		*87.7*
	Other		*110*		*12.3*
Number of evaluations		895		100	
	1		*222*		*24.8*
	2		*232*		*25.9*
	3		*151*		*16.9*
	4		*142*		*15.9*
	5		*112*		*12.5*
	6		*34*		*3.8*
	7		*2*		*0.2*
Student's study program		895		100	
	master		*170*		*19.0*
	pre-Master		*179*		*20.0*
	bachelor		*546*		*61.0*
Teacher's gender		24		100	
	male		*22*		*91.7*
	female		*2*		*8.3*
Teacher's rank		24		100	
	lecturer		*9*		*37.5*
	assistant professor		*5*		*20.8*
	professor		*5*		*20.8*
	full professor		*5*		*20.8*
Class size		24		100	
	small (< 50)		*5*		*20.8*
	medium (<100)		*10*		*41.7*
	large (≥100)		*9*		*37.5*

Analysis. Because of the cross-classified hierarchical nature of the dataset, which suggests that evaluations are nested in both students and courses (Figure 1), cross-classified multilevel analysis using a binary outcome variable (0 = no response; 1 = response) is the best placed method to evaluate which characteristics at the questionnaire level, the student level and the teacher/course level are significant predictors of a student's decision to participate in a particular evaluation. All models were run via the SAS GLIMMIX procedure (SAS Institute, 2008).

Figure 1: Schematic of Cross-classification with Questionnaires, Students, and Courses



4. Results

First, a cross-classified intercept only model was fitted to test for sufficient between-class variance, which is needed for multilevel analysis (Model 0 in Table 2). The results show that the intercept equals -1.2133, which suggests that, across all students and all courses, the overall chance of participating in a course evaluation is e-1.2233 = .297 or almost 30%. The model also reports a not positive definite estimated G matrix and suggests that the random component at the course level is estimated to be 0. In other words, there almost no variation to be explained at the course level. Therefore, this level (and the corresponding teacher and course characteristics) should be removed from the model (Searle, Casella & McCulloch, 1992). An intercept only model which includes only the questionnaire level (first level) and the student level (second level) (Model 1), suggests that the intraclass correlation coefficient (i.e. the variability that is due to the student level) equals .489 (Snijders & Bosker, 1999).

In Model 2, the variables at the questionnaire level were added. The estimates for this model show that a student's course grade is a statistically significant predictor of student's participation in a course evaluation. Compared to students who passed for the examinations with a satisfactory grade, students who received no course grade (because they did not take examinations) or failed for the course are less willingly to participate in an evaluation of that course (the estimated odds are about 63% and 43% of the odds for a satisfactory course grade, respectively). The effects of course type and very good/excellent course grades did not reach statistical significance. Model 3 includes the variables at the student level as well. The parameters show that both the number of evaluations a student was asked to complete and his/her educational level are statistically significant indicators for their participation in a course evaluation. The higher the number of evaluations, the less likely a student will be to participate in a course evaluation. Besides, the odds of students attending a pre-Master's or a Master's program are almost three times higher than student's attending a Bachelor's program. It thus seems to be the case that more mature students are more willingly to participate in course evaluations.

Table 2: Estimates for Multilevel Logistic Regression of Response as a Function of Student and Course Characteristics

	Model 0		Model 1		Model 2		Model 3	
Fixed Effect	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**	**Estimate**	**Standard Error**
Intercept	-1.2133***	.08466	-1.2133***	.08466	-1.1131***	.1272	-1.0898***	.2811
Course Type					0.0559	.1429	0.0148	.1566
Course grade: none					-0.8537***	.2292	-1.0020***	.2377
Course grade: Fail					-0.4583*	.2125	-0.4805*	.2154
Course grade: very good					0.0387	.1645	0.0582	.1678
Course grade: excellent					0.2226	.1931	0.2839	.1971
Gender							0.1752	.1739
Study domain							0.3792	.3240
Master student							1.1248***	.2245
Pre-Master student							1.0179***	.2240
N Evaluations							-0.1904**	.0638
Teacher's gender			.	.	.	.	.	.
Teacher's rank			.	.	.	.	.	.
Class size			.	.	.	.	.	.
Random Effect	**Variance Component**	**Standard Deviation**	**Variance Component**	**Standard Deviation**	**Variance Component**	**Standard Deviation**	**Variance Component**	**Standard Deviation**
$\tau_{00} = var(U_{0j0})$ intercept variance	3.1535	.2778	3.1535	.2778	3.1703	.2811	3.0200	.2703
$\tau_{00} = var(U_{00j})$ intercept variance	0 (G-matrix not positive definite)	/						

Note. Dummy codes: Course Type (0 = required, 1 = elective), Course grade (ref. = pass), Gender (0 = male, 1 = female), Study domain (0= Social Sciences, 1 = other), Student's study program (ref. = Bachelor's program). N Evaluations = number of evaluations.
* p < .05 ** p < .01, ** p < .001

5. Discussion and Conclusion

This contribution contains an exploratory study on the non-response issue in the field of SET. By means of multilevel binary regression models, we found that several characteristics at both the questionnaire level (student's course grade) and the student level (number of evaluations, educational level) are statistically significant predictors of a student's decision to participate in a SET-survey. It was shown that students who passed the examinations for a course are more likely to complete an evaluation questionnaire concerning teaching quality in that particular course. Since several authors report a moderate, but statistically significant correlation between course grade and SET-scores (Brockx, Spooren & Mortelmans, 2011; Feldman, 1997; Marsh, 2007; Spooren, 2010), the working hypothesis then is that SET-scores from a self-selected sample are slightly higher than the scores one could expect from the total population (i.e. all students who were enrolled in a course). Besides, the present study reveals that mature students are more likely to participate in course evaluation surveys. This might be due to the fact that they are more familiar and more involved with systems of quality assurance in higher education. Moreover, it is likely that they, during their study career, already sensed that their opinions are taken seriously by both teachers and educational boards. The results also show that students who are overburdened with invitations to participate in SET, are less likely to seize the opportunity to share their experiences concerning the organization and teaching in a course. SET-administrators should be aware of this important finding, and monitor a fair staggering of evaluations during the academic year/ the educational program to avoid over-surveying. An alternative would be to use sampling strategies (Kreiter & Lakhsman, 2005).

An important limitation of this study is that only 'administrative' variables were included in the analysis. Further research should take into account, for instance, student's feelings towards SET and motivational variables as well. There is a fair chance that these variables might have an even stronger effect on a student's decision to (not) participate in SET-surveys.

In sum, the present study might have important implications for the use of (online) SET. Stakeholders should take into account that student's performance, student's level and the number of evaluations play an important role in the decision to engage in SET-procedures. Therefore, the sample of students responding to the evaluations might not be representative for the complete population.

References

Anderson, H. M., Cain, J.C. & Bird, E. (2005). Online student course evaluations: review of literature and a pilot study. *American Journal of Pharmaceutical Education*, *69*, 34-43. Retrieved from http://archive.ajpe.org/view.asp?art=aj690105&pdf=yes.

Arthur, L. (2009). From performativity to professionalism: lecturer's responses to student feedback. *Teaching in Higher Education*, *14*, 441-454. doi:10.1080/1356251090305022

Ballantyne, C. (2003). Online Evaluations of Teaching: An Examination of Current Practice and Considerations for the Future. *New Directions for Teaching and Learning*, *96*, 103-112. doi: 10.1002/tl.127

Barkhi, R. & Williams, P. (2010). The impact of electronic media on faculty evaluation. *Assessment and Evaluation in Higher Education, 35*, 241-262. doi:10.1080/02602930902795927

Bothell, T. W. & Henderson, T. (2003). Do online ratings of instruction make sense? *New Directions for Teaching and Learning, 96*, 69-79. doi:10.1002/tl.124

Braun, E. & Leidner, B. (2009). Academic course evaluation. Theoretical and empirical distinctions between self-rated gain in competences and satisfaction with teaching behavior. *European Psychologist, 14*, 297-306. doi:10.1027/1016-9040.14.4.297.

Brockx, B., Spooren, P., & Mortelmans, D. (2011). Taking the 'grading leniency' story to the edge. The influence of student, teacher, and course characteristics on student evaluations of teaching in higher education. *Educational Assessment, Evaluation and Accountability, 23*, 289-306. doi:10.1007/s11092-011-9126-2

Bullock, C.D. (2003). Online Collection of Midterm Student Feedback. *New Directions for Teaching and Learning, 96*, 95-102. doi:10.1002/tl.126

Burden, P. (2008). Does the end of semester evaluation forms represent teacher's views of teaching in a tertiary education context in Japan? *Teaching and Teacher Education, 24*, 1463-1475. doi:10.1016/j.tate.2007.11.012

Dommeyer, C. J., Baum, P., Hanna, R. W., & Chapman, K. S. (2004). Gathering faculty teaching evaluations by in-class and online surveys: their effects on response rates and evaluations. *Assessment & Evaluation in Higher Education, 29*, 611-623. doi:10.1080/02602930410001689171

Emery, C. R., Kramer, T. R., & Tian, R. (2003). Return to academic standards: a critique of students' evaluations of teaching effectiveness. *Quality Assurance in Education, 11*, 37-47. doi:10.1108/09684880310462074

Feldman, K.A. (1997). Identifying exemplary teachers and teaching. Evidence from student ratings. In R. Perry & J. Smart (Eds.), *Effective teaching in higher education. Research and Practice* (pp. 368-395). New York: Agathon.

Galbraith, C., Merrill, G., & Kline, D. (2011, in press). Are student evaluations of teaching effectiveness valid for measuring student outcomes in business related classes? A neural network and Bayesian analyses. *Research in Higher Education*. doi:10.1007/s11162-011-9229-0

Gamliel, E. & Davidovitz, L. (2005). Online versus traditional teaching evaluations: mode can matter. *Assessment & Evaluation in Higher Education, 30*, 581-592. doi:10.1080/02602930500260647

Johnson, R. (2000). The authority of the student evaluation questionnaire. *Teaching in Higher Education, 5,* 419-434. doi:10.1080/713699176

Kember, D., Leung, D., & Kwan, K. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment and Evaluation in Higher Education, 27,* 411-425. doi:10.1080/0260293022000009294

Kreiter, C. D. & Laksham, V. (2005). Investigating the use of sampling for maximising the efficiency of student-generated faculty teaching evaluations. *Medical Education, 39*, 171-175. doi:10.1111/j.1365-2929.2004.02066.x

Leung, D. Y. P. & Kember, D. (2005). Comparability of data gathered from evaluation questionnaires on paper and through the internet. *Research in Higher Education, 46*, 571-591. doi:10.1007/s11162-005-3365-3

Liu, Y. (2006). A Comparison Study of Online versus Traditional Student Evaluation of Instruction. International *Journal of Instructional Technology and Distance Learning, 4,* 15-29. Retrieved from http://www.itdl.org/

Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319–383). New York: Springer.

Mortelmans, D. & Spooren, P. (2009). A revalidation of the SET37-questionnaire for student evaluations of teaching. *Educational Studies, 35*, 547-552. doi:10.1080/03055690902880299

Penny, A. R. (2003). Changing the agenda for research into student's views about university teaching: four shortcomings of SRT research. *Teaching in Higher Education, 8*, 399-411. doi: 10.1080/13562510309396

SAS Institute Inc. (2008). *SAS/STAT® 9.2 User's Guide*. Cary, NC: SAS Institute Inc.

Searle, S. R., Casella, G. & McCulloch, C.E. (1992). *Variance Components*. New York: John Wiley and Sons Inc.

Seldin, P. (1993). The use and abuse of student ratings of professors. *Chronicle of Higher Education, 39*, A40.

Snijders, T. & Bosker, R. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage Publishers.

Spooren, P. (2010). On the credibility of the judge. A cross-classified multilevel analysis on student evaluations of teaching. *Studies in Educational Evaluation, 36*, 121-131. doi:10.1016/j.stueduc.2011.02.001

Vennette, S., Sellnow, D., & McIntyre, K. (2010). Charting new territory : assessing the online frontier of student ratings of instruction. *Assessment & Evaluation in Higher Education, 35*, 101-115. doi: 10.1080/02602930802618336