Reasons for Responding in Student Evaluation of Teaching

Meinald Thielsch, Birte Brinkmöller and Boris Forthmann
Department of Psychology, University of Münster, Münster, Germany

Correspondence should be addressed to
Meinald T. Thielsch
Department of Psychology
University of Münster
Fliednerstr. 21
48149 Münster
thielsch@uni-muenster.de

Abstract
Nowadays, a systematic student evaluation of teaching (SET) is common at many universities. However, not all students evaluate all relevant courses and, as a consequence, feedback from a substantial number of students is missing. The current study examines response behavior with respect to demographic, motivational, and situational variables. Based on an online survey, data of $N = 640$ participants were analyzed. Using regression analysis based on multiple imputation, no significant effects for the reported participation in SET were found for demographic variables or opportunity costs. However, experienced salience of SET, aspects of social exchange, survey mode, and the general willingness to participate in surveys were all identified as significant predictors of SET participation. The results of our study can be helpful for evaluation managers to reduce non-response, especially stressing the importance of feedback and communication in SET.

Key words: SET, salience, social exchange theory, survey mode, survey fatigue

Acknowledgement
We thank Philipp Deing for his active help during the planning and investigation phase and Tina Dusend for supporting this study.

This is a pre-copyedited, author-produced PDF of an article accepted following peer review. Please cite as:

Thielsch, M. T., Brinkmöller, B. & Forthmann, B. (2018). Reasons for responding in student evaluation of teaching. *Studies in Educational Evaluation, 56*, 189-196. https://doi.org/10.1016/j.stueduc.2017.11.008

For many years student evaluation of teaching (SET) are a common part of higher education (e.g., Clayson, 2008; Spooren, Brockx & Mortelman, 2013). Issues in research that relate to SET dealt with the question how they should be conducted, which problems may appear using commonly applied measurements, which bias variables do exist and how valid SET are in general (e.g., Clayson, 2008; Marsh, 1984; Morrison, 2013; Narayanan, Sawaya & Johnson, 2014; Spooren et al., 2013). Such scientific work addressing SET facilitated the improvement of the evaluation processes in universities, emphasizing the importance of high quality evaluation instruments and the feedback of results (e.g., Nowakowski & Hannover, 2015; Penny & Coe, 2004; Schmidt & Loßnitzer, 2010). Nowadays, measures do exist even for special requirements, for example the evaluation of written exams (see Froncek, Hirschfeld & Thielsch, 2014).

Yet, the participation rate in SET is a general problem. Fluctuating rates between only 30% and 70% are often reported (Morrison, 2013; Spooren et al., 2013). In addition, various authors found lower response rates for online-based methods of data collection as compared to paper-based methods (e.g., Bacon, Johnson & Stewart, 2016; Ling, Philipps & Weihrich, 2012; Morrison, 2011; Stowell, Addison & Smith, 2012). Consequently, a substantial part of the targeted students may not give an opinion leading to a risk of systematical sampling errors. This issue is commonly referred to as non-response-bias (Adams & Umbach, 2011; Bacon et al., 2016; Sax, Gilmartin & Bryant, 2003). The aim of the present work is to examine possible reasons and influence factors for response behavior in SET.

Response Behavior in SETs

The fact that not all participants are able or willing to participate in a survey is a typical effect in empirical social research (Bosnjak, 2001; Groves, 2006). Generally, when the focus is on non-response, persons behavior can be roughly separated in two categories (see Bosnjak, 2001; Sax et al., 2003): a) non-response to single questions of a questionnaire (called “item non-response”) and b) persons that do not take part in the survey at all (called “unit non-response” or “total non-response”). The present study deals with students’ response behavior with respect to their participation in all SETs for which their participation was actually requested. In this context, it is important to understand possible reasons for non-response and how its detrimental consequences, such as bias, unfolds (Groves, 2006; Sax et al., 2003).

High rates of participation are not automatically the key to a successful avoidance of bias effects and under certain circumstances low rates of response can still lead to reliable data (see Groves, 2006; Sax et al., 2003). However, recent research in the field of SET fuels the idea of possible systematical bias-effects: Some authors report a higher rate of participation of women and students with good grades (Estelami, 2015; Reisenwitz, 2016). Bacon and colleagues (2016) assume on the basis of an analysis of around 8800 courses that if response rate is low students who were more committed and interested in the class were giving their opinions. This causes more accented data patterns: in SET with lower participation lecturers with positive evaluation receive even better scores and lecturers with bad evaluations are rated worse. When the response rate increases also less committed students participate in the survey and both groups of lecturers will be rated rather in the direction of the scale mid-point. However, the commitment of the students was not directly measured in the study of Bacon and colleagues (2016). Hence, it is questionable in which way further aspects next to demographical characteristics play a role in SET participation.

This will be discussed in the following section on a theoretical level and connected to the question of the current work.

Predictors of response behavior in SET

Several subjective-motivational and situational factors may influence response behavior in SET. In order to extend the work of Adams (2010), five of them are in the focus of the current study: (1) perception of SET as social exchange, (2) SET salience, (3) opportunity costs, (4) survey mode, and (5) general survey participation.

The social exchange theory (Emerson, 1976) posits that social interactions rely on a certain reciprocity or a cost-benefit relation (Dillmann, 1978). In addition, costs and benefits are taken into account before social interactions. Thus, if a person expects to receive a benefit investment in costs will be more likely. Groves, Cialdini and Couper (1992) assume an influence of the social exchange theory in response behavior. In SET, the necessary personal effort and dislike to evaluate courses can be considered as the costs. However, the students may experience a benefit from having learned something during the term in their courses – and feel obliged to honor this by SET participation.

Salience is the context of teaching evaluation is the quality of an upcoming or running SET to stand out and to draw attention. This could be done by making SET-related information available and vivid in the minds of the students. Salience is an important characteristic of SET that can motivate students to participate (Adams, 2010). Implementations of SET may vary in how they prompt students to participate (see Berk, 2012) and, moreover, there are differences regarding the specific characteristics of an evaluation that motivate students (Groves & Peytcheva, 2008). This is why we investigate several factors of salience which possibly influence participation: the evaluation behavior of the fellow students and also awareness of the consequences of the SET. Furthermore, salience can be supported by lecturers constantly reminding the students of the evaluation and, thus, expressing appraisal towards SET. Finally, general importance of SET at the respective department was considered.

Furthermore, students experience opportunity costs – those are reflected by little time resources or in attractive alternative activities (Adams, 2010). For some students, the present education system in western countries comes with a high level of perceived stress caused (amongst others) by challenges in this new stage of life, the mere number of exams or a subjective experienced pressure to perform (e.g., Bechler & Thielsch, 2012; Brougham et al., 2009; Rayle & Chung, 2007). We assume that responding decreases when time resources are little and experienced opportunity costs are high. In the present study, we measured experienced time pressure beyond regular studies as indicators for opportunity costs. These were side jobs of the questioned students, or the distance between their home and their university which implicates travel time (Couper, Kapteyn, Schonlau, & Winter, 2007).

A prominent issue in SET research is the question to what extent an online based survey can be compared to a paper based SET and which of these modes yields better results (for a review see Morrison, 2013). Some studies found evidence for comparable results and return rates (e.g., Perret, 2013; Venette, Sellnow & McIntyre, 2010), other studies observed reduced return rates for online based SET (e.g., Fike, Doyle & Conelly, 2010; Ling et al., 2012; Stowell et al., 2012). Consequently, we will consider survey mode as a possible predictor of response behavior in the present study.

Finally, survey fatigue appears to be relevant in this context due to the high number of surveys among students which is also caused by a raising amount of web based studies (Adams, 2010; Groves et al., 2004; Porter, Whitcomb & Weitzer, 2004). These studies are not only related to teaching at universities. For example, students receive invitations for studies of market-, social and opinion research or commercial providers ask for ratings concerning (online) shopping, journeys or hotel visits. The constant request to take part in such surveys can lead to an excessive exposure and, as a consequence, students refrain from participation (Porter et al, 2004). According to Adams (2010) this effect is more likely when the number of the current courses which are to be evaluated is high or the students are in higher semesters and have already taken part in a lot of evaluations in the past.

Aim of the present study

The present study examines which of these five subjective-motivational and situational factors described above – (1) social exchange, (2) salience of SET, (3) students' opportunity costs (4) survey mode, and (5) students' general survey fatigue – next to demographical variables such as age and gender have an influence on the participation in SET (see Table 1). We expected that several factors influence the participation in SET and tested this assumption by means of hierarchical multiple regression.

Method

Sample

A total of 867 persons started the survey, N = 218 (25.14%) dropped out without providing information for any of the relevant study variables; N = 9 (1.04%) denied permission to use the data. Thus, the final sample – including missing values (see treatment below) – consisted of 640 students, 439 females (68.59%) and 197 males (30.78%). The participating persons were aged between 16 and 52 (M = 24.19, SD = 4.42). Altogether, German-speaking students from 118 different universities participated in the study. One-third (29.38%) studied at the University of Münster. Two further universities which were often named are the University of Koblenz-Landau (4.53%) and the University of Leipzig (3.44%). The most often named major was psychology (32.50%), followed by medicine (7.66%) and teaching (6.41%).

Importantly, the sample included the full range of evaluation behavior (see Table 2). There were participants who – according to their self-report – never took part in an SET (4.38% of the sample) and also those who reported that they never had missed an SET (24.84%). On average participants reported they had participated in 68.34% of their SETs.

Material

The items of the questionnaire were conceptually based on each of the mentioned theoretical aspects. Additionally, fifteen German universities – those with the highest scores in terms of the criteria 'involvement of students in evaluation of teaching' (for example, general distribution of SET results, informing students with about SET results) rated by the CHE university ranking 2012/2013 (www.che-ranking.de) were chosen – and all departments of the University of Münster were contacted via e-mail in order to obtain information regarding their currently used methods of evaluation. The provided information captured a variety of

different methods which served as a basis for item inclusion. This procedure justified applicability of the questionnaire for a wide range of participants. The questionnaire consisted of items regarding the reasons for response behavior, concerning the behavior of evaluation, and attitudes towards evaluation (for specific items see Table 1).

Table 1. Factors influencing participation in SET: Items used for regression analysis in the present study.

Role in regression analysis	Item-content	Answer mode
Dependent variable	If you remember previous evaluations of teaching, how often did you evaluate all relevant courses?	In sum, I evaluated __ % of my courses.
Block 1 – demography	Gender	female / male
Block 1 – demography	Age	*Open answer format*
Block 2 – social exchange	As a student, I can identify myself with my university.	7-point Likert scale (ranging from *totally disagree* to *totally agree*)
Block 2 – social exchange	I am satisfied with my study performance.	7-point Likert scale (ranging from *totally disagree* to *totally agree*)
Block 2 – social exchange	My subject was my first choice.	7-point Likert scale (ranging from *totally disagree* to *totally agree*)
Block 2 – social exchange	I am happy with my choice of subject.	7-point Likert scale (ranging from *totally disagree* to *totally agree*)
Block 3 – salience	Lecturers of my subject highly value the evaluation of teaching.	7-point Likert scale (ranging from *totally disagree* to *totally agree*)
Block 3 – salience	Generally, in my field of study teaching evaluations are highly valued.	7-point Likert scale (ranging from *totally disagree* to *totally agree*)
Block 3 – salience	My peers at the university evaluate all of their courses.	7-point Likert scale (ranging from *totally disagree* to *totally agree*)
Block 3 – salience	Did you know, that the evaluation of teaching can lead to consequences for teachers and future teaching?	yes / no
Block 4 – opportunity costs	Did you work alongside your studies?	yes / no
Block 4 – opportunity costs	Is your current residence situated more than 10km away from the university?	yes / no
Block 5 – survey mode	How is the evaluation of teaching of your subject conducted?	paper based / computer based / both
Block 6 – survey fatigue	Semester (total)	*Open answer format*
Block 6 – survey fatigue	What was your evaluation behavior at the beginning of your studies? At the start of my studies I ...	5-point Likert scale (ranging from *evaluated much less* to *evaluated much more frequent*)
Block 6 – survey fatigue	During a semester, I take part in many surveys.	7-point Likert scale (ranging from *totally disagree* to *totally agree*)

Procedure

The targeted age-group for this study was almost fully reachable via the Internet (95% to 100% Internet usage in the target group at the time of the study; see Eimeren & Frees, 2013). Thus, the questionnaire was administered within an online testing environment called EFS

Survey 10 (provided by the Questback GmbH). We used several ways for participant acquisition: (1) the online-panel PsyWeb (a collaborative online research panel of the Universities of Münster, Leipzig, Munich, and the Applied University of Osnabrück https://psyweb.uni-muenster.de/), (2) different student groups on the social network Facebook, (3) flyer in all faculties and a canteen of the University of [anonymized for peer review], (4) invitations and a reminder sent via e-mail to 48 student councils of several German universities and to 15 student councils of the University of [anonymized for peer review]. The field inquiry took place from 25.06.2013 to the 22.10.2013, so that one could participate both during the summer semester 2013 and the beginning of the winter semester 2013/2014.

At the beginning of the survey participants were informed about the anonymous treatment of their data, the voluntariness of participation, the estimated study time and the responsible researchers. In the second step socio-demographical data were gathered (age, gender, university etc.). Afterwards information about the participation of the respondents in SET and the general praxis of SET at their respective university were acquired. In the subsequent part the students were asked to report their personal opinion towards the university, the SET and their subject (Table 1, see Appendix for original wording in German). Moreover, potential reasons for non-response were directly asked (in the present study not analyzed in detail). At the end, the participates had the chance to request a report of the study results, give their opinion about the survey and release or exclude their data from the analysis. Finally, participants were thanked and the responsible researchers were named again. The study time was about 19 minutes.

Results

The pattern of missing values was analyzed in detail prior to analysis. Two problems here need to be taken into account: (a) study variables are both, continuous and categorical and, thus, common methods for testing MCAR vs. MAR (Jamshidian & Jalal, 2010; Little, 1988) were not applicable for the full set of variables, and (b) items were grouped and presented in a fixed order and, as a consequence, the missing data mechanism was monotone (see Table 2). Thus, following suggestions by Ridout and Diggle (1991), logistic regression was used to test the pattern of missing data block-wise. That is, a variable for the second of two consecutive presentation blocks was created indicating if participants dropped out of the study or not. This dichotomous variable was then predicted by all variables (a stepwise logistic regression was run) that had been answered up to the first block of the consecutive pair. This procedure revealed that missingness in four out of seven presentation blocks was significantly predicted by previously answered study variables. Thus, the MCAR assumption was violated and MAR was a more reasonable assumption here. Consequently, all data analyses were based on multiple imputation by means of the R package mice (van Buuren & Groothuis-Oudshoorn, 2011).

Descriptive statistics and bivariate correlations for the variables of the five motivational and situational factors and the reported evaluation behavior can be found in Table 2. Importantly, all bounded study variables covered the full range of possible values.

Table 2. Descriptive statistics and correlation matrix of reported evaluation behavior and several independent variables (standard errors of correlations are in parantheses).

		Final dataset					Imputed datasets																		
		M	*SD*	Min	Max	%-missing	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Evaluation behaviour	1	68.34	32.83	0	100	9.69%	68.26	32.84	--																
Gender	2	0.31	0.46	0	1	0.63%	0.31	0.46	-.06 (.04)	--															
Age	3	24.19	4.42	16	52	0.16%	24.19	4.42	-.04 (.04)	.07 (.04)	--														
Identification with university	4	4.88	1.58	1	7	22.81%	4.87	1.58	.22*** (.05)	-.06 (.05)	-.03 (.05)	--													
Satisfaction with performance	5	5.21	1.48	1	7	22.81%	5.20	1.48	.14** (.04)	-.15** (.05)	-.05 (.05)	.32*** (.04)	--												
Subject first choice	6	5.82	1.71	1	7	22.81%	5.81	1.71	.20*** (.05)	-.01 (.05)	-.03 (.05)	.30*** (.04)	.24*** (.04)	--											
Satisfaction with subject	7	5.78	1.44	1	7	22.81%	5.78	1.44	.15*** (.04)	-.04 (.05)	.02 (.05)	.51*** (.03)	.35*** (.04)	.39*** (.04)	--										
Lecturers value SET	8	4.49	1.52	1	7	17.50%	4.47	1.53	.32*** (.04)	-.01 (.05)	-.06 (.05)	.30*** (.04)	.10* (.04)	.17*** (.04)	.25*** (.04)	--									
Importance of SET	9	4.71	1.63	1	7	17.50%	4.68	1.64	.37*** (.04)	-.05 (.04)	-.09 (.05)	.26*** (.04)	.12** (.04)	.18*** (.04)	.22*** (.04)	.70*** (.02)	--								
Participation of peers	10	3.86	1.61	1	7	17.50%	3.83	1.62	.46*** (.03)	-.08 (.04)	-.08 (.05)	.18*** (.04)	.07 (.05)	.13** (.05)	.13** (.04)	.29*** (.04)	.41*** (.04)	--							
Consequences known[a]	11	0.29	0.45	0	1	17.66%	0.30	0.46	-.19*** (.04)	-.02 (.04)	.02 (.05)	-.09* (.04)	-.16*** (.04)	-.15** (.04)	-.12** (.04)	-.30*** (.04)	-.26*** (.04)	-.06 (.04)	--						
Work: yes vs. no[a]	12	0.43	0.50	0	1	6.25%	0.42	0.49	.05 (.04)	-.03 (.04)	-.19*** (.04)	-.01 (.04)	.05 (.04)	.03 (.04)	.00 (.05)	.04 (.04)	.01 (.04)	.01 (.04)	-.04 (.04)	--					
Distance to university[a]	13	0.65	0.48	0	1	5.78%	0.65	0.48	.10* (.04)	-.07 (.04)	-.14** (.04)	.07 (.04)	.05 (.05)	.04 (.04)	.03 (.04)	.04 (.04)	.08 (.04)	.09* (.04)	-.07 (.04)	.05 (.04)	--				
Paper-pencil vs. computer-based	14	0.27	0.44	0	1	12.81%	0.27	0.44	.05 (.04)	.01 (.04)	-.02 (.05)	.03 (.04)	.01 (.04)	.05 (.04)	.00 (.05)	.14*** (.04)	.24*** (.04)	.04 (.04)	-.17*** (.04)	.03 (.04)	.04 (.04)	--			
Paper-pencil vs. both	15	0.28	0.45	0	1	12.81%	0.29	0.45	-.09* (.04)	.07 (.04)	.09 (.05)	-.00 (.04)	-.07 (.05)	.05 (.05)	.02 (.04)	.02 (.04)	.00 (.04)	-.17*** (.04)	.00 (.05)	-.04 (.04)	-.03 (.04)	-.38*** (.04)	--		
Number of semesters	16	7.31	4.00	1	37	5.78%	7.33	4.08	-.02 (.04)	.00 (.04)	.46*** (.03)	.01 (.05)	-.01 (.05)	-.04 (.05)	-.02 (.05)	-.07 (.05)	-.05 (.05)	-.02 (.05)	-.09* (.04)	-.13** (.04)	.01 (.04)	-.02 (.04)	-.00 (.04)	--	
Previous SET behavior	17	3.17	0.85	1	5	18.28%	3.18	0.86	-.13** (.05)	-.03 (.04)	.04 (.05)	-.10* (.05)	.04 (.05)	-.05 (.05)	-.04 (.05)	-.11* (.04)	-.10* (.04)	-.13** (.04)	.00 (.04)	-.04 (.04)	-.05 (.04)	-.03 (.04)	.07 (.04)	-.06 (.05)	--
Survey participation	18	4.78	1.59	1	7	17.50%	4.74	1.61	.47*** (.03)	-.13** (.04)	-.05 (.05)	.25*** (.04)	.12** (.04)	.13** (.04)	.17*** (.04)	.29*** (.04)	.31*** (.04)	.32*** (.04)	-.14** (.04)	.06 (.04)	.14** (.04)	.06 (.04)	.06 (.04)	-.05 (.05)	-.19*** (.04)

Notes. Final data: $N = 640$. Multiple imputation results are based on $m = 100$ imputed datasets. Means, standard deviation, correlations, and inference for correlations were pooled by means of functions provided by the R package miceadds (Robitzsch, Grund, & Henke, 2017). Standard errors of correlations (depicted in parentheses) are in the same measurement unit as correlations. [a]Dummy-coded with 0 = yes and 1 = no. Gender: 0 = female and 1 = male. Paper-pencil vs. computer-based: paper-pencil 0 and computer-based = 1. Paper-pencil vs. both: paper-pencil = 0 and both = 1.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p$ .001.

Moderate to large correlations of the dependent variable evaluation behavior with frequency of participation of one's peers and self-estimated frequency of survey participation in general were found. Furthermore, we noticed moderate correlations of reported evaluation behavior and importance of SET as it was communicated by the teacher, and perceived importance of SET in general. Small correlations were found between evaluation behavior and all variables relating to the social exchange theory, the extent to which consequences of SET were known, and previous evaluation behavior. High correlations between independent variables were found for identification with one's home university, satisfaction with the chosen subject, general status of the SET, and the importance of SET as indicated by the teacher.

Subsequently, we compared hierarchical multiple regression models in which the independent variables were entered according to theoretical blocks (see Table 3). As a measure of effect size for each block, we considered increase in R^2 with .02 (small effect), .13 (moderate effect), and .26 (large effect). These cut-offs correspond with common cut-offs for Cohen's f^2 effect size which can be easily transformed to the R^2 metric (Faul, Erdfelder, Buchner, & Lang, 2009). In the first block demographical variables were included in the model. This block did not explain any variance of the stated evaluation behavior in a meaningful way. In the next step variables pertaining to the social exchange theory were included which led to a significant incremental variance explained of 7.1% (small effect). In particular, identification with the own university and first choice of subject predicted the reported evaluation behavior in this block. In the third step, all salience variables were included in the model which led to a significant incremental R^2 of 21.1% (moderate effect). In this block general importance of SET, participation of peers, and awareness of consequences were predictive for evaluation behavior. The inclusion of the opportunity cost variables and also of survey mode did not further improve prediction of evaluation behavior. In the last step, general survey participation raised explained variance by 8.3% (small effect) to an amount of 37.7%. In addition, it should be noted that when all variables entered the model, also the contrast of paper-pencil assessment vs. both (paper-pencil and computer based) was significant (indicating that using both survey modes was detrimental for evaluation behavior). This latter observation was robust even when survey mode was entered last into the model. However, it should be noted that when survey mode was entered last, the increase in R^2 was a little higher (entered last: .008; entered second last: .003), but it did not surpass the conventional threshold for a small effect size. Moreover, confidence intervals for the estimated coefficient of paper-pencil vs. both did substantially overlap (entered last: 95%-CI: [-12.27, -1.09]; entered second last: 95%-CI: [-9.63, 2.18]). Thus, this difference should not be overstated.

Thus, the independent variables of the blocks relating to the social exchange theory, salience, survey mode, and survey fatigue were statistically relevant for the variation in reported evaluation behavior.

Table 3. Hierarchical regression analysis results.

	Demography		Social exchange theory		Salience		Opportunity costs		Survey mode		Survey fatigue	
	b	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Intercept	69.59***	1.64	69.20***	1.61	70.68	1.72	67.88	2.67	70.09	3.05	73.62	6.13
Gender	-4.30	2.94	-3.06	2.90	-1.46	2.61	-1.28	2.61	-1.07	2.61	1.09	2.46
Age	-1.25	1.46	-0.86	1.42	0.46	1.30	0.83	1.32	0.96	1.33	0.95	1.38
Identification with university			5.28**	1.84	2.33	1.63	2.32	1.64	2.32	1.64	0.90	1.53
Satisfaction with performance			1.55	1.58	1.50	1.41	1.45	1.42	1.27	1.43	1.13	1.35
Subject first choice			4.75**	1.66	3.01*	1.49	2.97*	1.49	3.15*	1.50	3.13*	1.40
Satisfaction with subject			-0.17	1.81	-1.35	1.56	-1.33	1.56	-1.39	1.56	-1.49	1.47
Lecturers value SET					2.01	1.89	1.99	1.88	1.90	1.88	0.94	1.75
Importance of SET					4.29*	1.87	4.31*	1.87	4.97*	1.93	4.19*	1.79
Participation of peers					11.75***	1.37	11.68***	1.37	11.22***	1.41	8.68***	1.36
Consequences known[a]					-6.64*	2.98	-6.41*	2.98	-6.85*	3.00	-5.87*	2.81
Work: yes vs. no[a]							2.42	2.41	2.48	2.41	1.51	2.30
Distance to university[a]							2.54	2.54	2.57	2.54	0.50	2.42
Paper-pencil vs. computer-based									-4.13	3.10	-5.61	2.95
Paper-pencil vs. both									-3.73	3.01	-6.68*	2.84
Number of semesters											-0.01	0.33
Previous evaluation behavior											-0.44	1.44
Survey participation											10.48***	1.34
ΔR^2	–		.071		.211		.003		.003		.083	
ΔF (df_1,df_2)	–		8.54*** (4,5134.85)		35.61*** (4,7348.46)		1.03 (2,13414.50)		1.18 (2,9577.95)		21.30*** (3,6485.07)	
R^2	.006		.077		.288		.291		.294		.377	
F (df_1,df_2)	1.46 (2,8172.35)		6.43*** (6,10078.57)		19.08*** (10,17315.89)		16.45*** (12,24352.37)		14.46*** (14, 31434.79)		17.34*** (17,41265.47)	

Notes. Final dataset: $N = 640$. Multiple imputation results are based on $m = 100$ imputed datasets. Regression coeffients, R^2, and F values were pooled by means of functions provided by the R package mice (van Buuren & Groothuis-Oudshoorn, 2011). For model comparison F values the Wald-type method according to Meng and Rubin (1992) was used. [a]Dummy-coded with 0 = yes und 1 = no. Gender: 0 = female und 1 = male. Paper-pencil vs. computer-based: paper-pencil 0 and computer-based = 1. Paper-pencil vs. both: paper-pencil = 0 and both = 1. The dependent variable was not scaled in order to leave regression coefficients directly interpretable in terms of percentages. All (pseudo-)continuous variables were *z*-scaled and, thus, the intercept of the models is an estimate for the dependent variable when inserting the reference category for categorical predictors and the mean for continuous predictors. Regression coefficients for categorical predictors represent mean differences on the percent scale. Regression coefficients for continuous predictors represent the expected change of the dependent variable for a 1*SD* change of the predictor. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Discussion

In this study, we examined possible predictors of participation in SET. Participation of students is crucial for the validity and generalizability of SET results. Consequently, knowledge about variables and subjective motivational reasons facilitating participation are of highest practical relevance. In the present study, we systematically analyzed blocks of variables from various theoretical views by means of hierarchical multiple regression. The variables of the salience block had the largest effect on the reported evaluation behavior. In addition, we found effects for the variables concerning survey fatigue and social exchange theory.

Thus, salience of SET appears as a central factor when trying to increase participation. The indicators used in this study suggest that a positive evaluation climate is crucial. In this line of thinking we demonstrated that general importance of SET and participation of peers strongly predicted the reported evaluation behavior. To increase salience of SET several measures are reasonable. For example, reminding students to participate in a way that concurrently provides survey access or incentivizing participation are likely steps of an appropriate action plan in this regard (see Berk, 2012).

Concerning survey fatigue, a reverse effect as compared to what was suggested by literature was found (e.g., Adams, 2010; Groves et al., 2004; Porter et al., 2004): students reporting to generally participate in surveys quite often also stated a higher participation in SET. A general positive attitude and openness towards surveys is reflected here. Yet, persons responsible of a university's SET have only limited impact on such variables.

Regarding the indicators concerning the social exchange theory the identification with the university and the subject (operationalized by first choice) were of particular concern. This indicates that, according to the theory, students who identify stronger also tend to feel obliged to take part in SET. This result fits with the assumption of Bacon and colleagues (2016) who assume a higher participation in evaluation of dedicated students.

Frequently discussed variables such as, for example, survey mode in terms of paper-based vs. computer-based (see e.g. Ling et al, 2012; Morrison, 2011; Stowell et al., 2012) or a possible impact of demographical variables such as gender (see Reisenwitz, 2016) did not predict reported participation in SET in this study. With regard to the latter, effects of demographical variables, further research is necessary. In addition, these aspects are not modifiable by persons responsible for SET and, thus, could only be considered as potential bias-variables. Concerning the survey mode, it is likely that an increasing usage and acceptance of computer-based online survey techniques (see Gosling & Mason, 2015) have prevented previously reported detrimental effects of online SET to occur. The decision which mode of survey to use should nevertheless be made carefully – yet, when directly asked, students prefer the online mode (Donovan, Mader & Shinsky, 2007). Particularly, our findings indicate that it might have adverse effects to use both, paper-based and computer-based questionnaires, in combination at the same time, as this is associated with a significant decrease in reported SET participation. However, the found effect is rather small.

Implications for research and praxis

The current findings point to future research that focuses particularly on subjective motivational reasons of the students. In line with this, general willingness to take part in surveys should be investigated more deeply. Indirect and direct influencing factors of this relationship could also be addressed in such studies.

In order to increase response rates in SET it would be worthwhile to design SET as salient as possible (see Berk, 2012) while using only one survey mode (paper-pencil or online). Therefore, a SET needs to attract the students' attention right from the start of an assessment period. The current response rate could be used here to encourage participation by means of the following mechanism: The perception that other students evaluate may drive a positive effect on participation. Furthermore, students and also teachers need to be involved in order to strengthen the identification with the university. This enhanced identification can lead to a positive attitude of the students and may increase participation.

Generally, it remains to consider that not only high response rates are crucial for the validity of SET (see Groves, 2006; Sax et al., 2003). Furthermore, potential bias-effects due to non-response in the respective context should be taken into account (see Bacon et al., 2016).

Limitations and future research

Some aspects of the current study may limit the generalizability of the presented results. Yet, some limitations directly lead to possible approaches for future research. First of all, self-ratings are prone to common memory-biases. Moreover, there may exist a sample bias in the present study: All participants are German speaking and, because of the chosen ways of recruitment, a substantial part of them is studying at the same university. Furthermore, with 33.88% psychology students are overrepresented in the sample of the present survey. Consequently, generalizability of the results to the entirety of students should be treated with caution. Additionally, 70% of the sample consists of female students which is mainly driven by the typical gender distribution of psychology students (approximately 80% female). This could as well restrict the generalizability of found results. At least, in regard of drop-out from the present study it can be concluded that male and female participants do not differ. Still, a replication of our findings in a different sample is desirable.

With respect to the SET survey mode the present study investigated paper-based, computer-based and mixed approaches. It will be promising in future research on SET participation to further differentiate between possible devices used in computer-based evaluations (e.g., desktop computer, laptop, mobile devices). In doing so, it might be valuable to distinguish if students are asked to evaluate during the course or via online surveys in their spare time.

Finally, it needs to be noted that students who refuse to take part in surveys in general will also not be well represented in the present study. A realistic basic rate of total nonresponse in SET is hard to estimate on individual level. The average stated participation rate in our data was 68.34%, which fits in the range of generally reported participation rates in SETs (Morrison, 2013; Spooren et al., 2013). Thus, while the full range of values was covered in the sample it should be taken into account that participants who take part more often in SETs might be somewhat overrepresented in the sample.

Conclusion

Persons responsible for a SET should try to avoid or at least diminish biases caused by non-response. The results of the present study indicate that from a subjective-motivational point of view especially SET salience is a crucial factor for student participation. In addition, further aspects originating from the social exchange theory (such as the identification with ones' subject or identification with ones' home university), but also general factors such as joy of participating in surveys need to be taken into account. In a positive evaluation climate and a productive university atmosphere these variables can be constructively influenced. In particular, the communication during SET can help to raise the salience and diminish non-response.

References

Adams, M. J. D. (2010). No evaluation left behind: Nonresponse in online course evaluations. (Dissertation, North Carolina State University). Retrieved March 30, 2017, from http://repository.lib.ncsu.edu/ir/handle/1840.16/6364

Adams, M. J. D., & Umbach, P. D. (2011). Nonresponse and Online Student Evaluations of Teaching: Understanding the Influence of Salience, Fatigue, and Academic Environments. Research in Higher Education, 53(5), 576–591. doi:10.1007/s11162-011-9240-5.

Bacon, D. R., Johnson, C. J., & Stewart, K. A. (2016). Nonresponse Bias in Student Evaluations of Teaching. Marketing Education Review, 26(2), 93–104. doi: 10.1080/10528008.2016.1166442.

Bechler, O. & Thielsch, M. T. (2012). Schwierigkeiten bei der Vorbereitung auf schriftliche Prüfungen [Difficulties in written exam preparation]. Zeitschrift für Hochschulentwicklung, 7 (3), 137-156.

Berk, R. (2012). Top 20 strategies to increase the online response rates of student rating scales. International Journal of Instructional Technology in Teaching and Learning, 8(2012), 98–107.

Bosnjak, M. (2001). Classifying response behaviors in web-based surveys. Journal of Computer-Mediated Communication, 6, 1–14.

Brougham, R. R., Zail, C. M., Mendoza, C. M., & Miller, J. R. (2009). Stress, sex differences, and coping strategies among college students. Current Psychology, 28(2), 85–97. doi:10.1007/s12144-009-9047-0.

Clayson, D. E. (2008). Student Evaluations of Teaching: Are They Related to What Students Learn?: A Meta-Analysis and Review of the Literature. Journal of Marketing Education, 31(1), 16–30. doi: 10.1177/0273475308324086.

Couper, M.P., Kapteyn, A., Schonlau, M., & Winter, J. (2007). Noncoverage and nonresponse in an Internet survey. Social Science Research, 36, 131-148.

Dillman, D. A. (1978). Mail and telephone surveys: The total design method. New York, NY: John Wiley & Sons

Donovan, J., Mader, C., & Shinsky, J. (2007). Online vs. traditional course evaluation formats: Student perceptions. Journal of Interactive Online Learning, 6(3), 158-180.

Emerson, R. M. (1976). Social Exchange Theory. Annual Review of Sociology, 2, 335-362.

Estelami, H. (2015). The Effects of Survey Timing on Student Evaluation of Teaching Measures Obtained Using Online Surveys. Journal of Marketing Education, 37(1), 54–64. doi: 10.1177/0273475314552324.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149-1160. doi: 10.3758/BRM.41.4.1149

Fike, D., Doyle, D., & Connelly, R. (2010). Online vs. paper evaluations of faculty: when less is just as good. Journal of Effective Teaching, 10(2), 42–54.

Froncek, B., Hirschfeld, G. & Thielsch, M. T. (2014). Characteristics of effective exams – Development and validation of an instrument for evaluating written exams. Studies in Educational Evaluation, 43, 79-87. doi: 10.1016/j.stueduc.2014.01.003.

Gosling, S. D., & Mason, W. (2015). Internet Research in Psychology. Annual Review of Psychology, 66(1), 877–902. doi:10.1146/annurev-psych-010814-015321

Groves, R. (2006). Nonresponse rates and nonresponse bias in household surveys: What Do We Know about the Linkage between Nonresponse Rates and Nonresponse Bias? Public Opinion Quarterly, 70(5), 646–675. doi: 10.1093/poq/nfl033.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. Public Opinion Quarterly, 56, 475-495.

Groves, R. M. & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias. A meta-analysis. Public Opinion Quarterly, 72(2), 167-189. doi: 10.1093/poq/nfnOl.

Jamshidian, M., & Jalal, S. (2010). Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. *Psychometrika*, *75*(4), 649-674. doi: 10.1007/S11336-010-9175-3

Ling, T., Phillips, J., & Weihrich, S. (2012). Online Evaluations vs. In-class Paper Teaching Evaluations : A Paired Comparison. Journal of the Academy of Business Education, 13, 150–162.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198-1202. doi: 10.1080/01621459.1988.10478722

Marsh, H. W. (1984). Students Evaluations of University Teaching - Dimensionality, Reliability, Validity, Potential Biases, and Utility. Journal of Educational Psychology, 76(5), 707–754. doi: 10.1007/1-4020-5742-3_9.

Meng, X. L., & Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, *79*(1), 103–111. doi: 10.1093/biomet/79.1.103

Morrison, R. (2011). A comparison of online versus traditional student end-of-course critiques in resident courses. Assessment & Evaluation in Higher Education, 36(6), 627–641. doi: 10.1080/02602931003632399.

Morrison, K. (2013). Online and paper evaluations of courses: a literature review and case study. Educational Research and Evaluation, 19(7), 585–604. doi: 10.1080/13803611.2013.834608.

Narayanan, A., Sawaya, W. J., & Johnson, M. D. (2014). Analysis of differences in nonteaching factors influencing student evaluation of teaching between engineering and business classrooms. Decision Sciences Journal of Innovative Education, 12(3), 233-265.

Nowakowski, A., & Hannover, B. (2015). Wie Lehrende die Ergebnisse studentischer Evaluationen wahrnehmen [How university lecturers perceive student evaluations of their teaching: An experimental manipulation of qualitative and quantitative feedback components]. Diagnostica, 61(3), 144–152. doi: 10.1026/0012-1924/a000144.

Penny, A. R., & Coe, R. (2004). Effectiveness of Consultation on Student Ratings Feedback: A Meta-Analysis. Review of Educational Research, 74(2), 215–253. doi:10.3102/00346543074002215.

Perrett, J. J. (2013). Exploring graduate and undergraduate course evaluations administered on paper and online: a case study. Assessment & Evaluation in Higher Education, 38(1), 85–93. doi: 10.1080/02602938.2011.604123.

Porter, S. R., Whitcomb, M. E. & Weitzer, W. H. (2004). Multiple Surveys of Students and Survey Fatigue. New Directions for Institutional Research, 121, 63-73. doi: 10.1002/ir.101.

Rayle, A., & Chung, K. (2007). Revisiting first-year college students' mattering: Social support, academic stress, and the mattering experience. Journal of College Student Retention, 9(1), 21–37. doi:10.2190/X126-5606-4G36-8132.

Reisenwitz, T. H. (2016). Student Evaluation of Teaching: An Investigation of Nonresponse Bias in an Online Context. Journal of Marketing Education, 38(1), 7–17. doi: 10.1177/0273475315596778.

Ridout, M. S., & Diggle, P. J. (1991). Testing for random dropouts in repeated measurement data. *Biometrics*, 1617-1621. doi: 10.2307/2532413

Robitzsch, A., Grund, S., & Henke, T. (2017). miceadds: Some additional multiple imputation functions, especially for mice. R package version 2.7–19. Retrieved from https://CRAN.R-project.org/package=miceadds

Sax, L. J., Gilmartin, S. K., & Bryant, A. N. (2003). Assessing Response Rate and Nonreponse bias in Web and Paper Surveys. Research in Higher Education, 44(4), 409–432. doi: 10.1023/A:1024232915870.

Schmidt, B., & Loßnitzer, T. (2010). Lehrveranstaltungsevaluation: State of the Art, ein Definitionsvorschlag und Entwicklungslinien [Course evaluations in German higher education institutions: State of the art, proposal for a definition and future trends]. Zeitschrift Für Evaluation, 9(1), 49–72.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the Validity of Student Evaluation of Teaching: The State of the Art. Review of Educational Research, 83(4), 598-642. doi: 10.3102/0034654313496870.

Stowell, J. R., Addison, W. E., & Smith, J. L. (2012). Comparison of online and classroom-based student evaluations of instruction. Assessment & Evaluation in Higher Education, 37(4), 465–473. doi: 10.1080/02602938.2010.545869.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. Retrieved from https://www.jstatsoft.org/article/view/v045i03

van Eimeren, B., & Frees, B. (2013). Rasanter Anstieg des Internetkonsums – Onliner fast drei Stunden täglich im Netz [Rapid rise of Internet use: Onliner are using the net nearly three hours per day]. Media Perspektiven, 2013(7–8), 358–372.

Venette, S., Sellnow, D., & McIntyre, K. (2010). Charting new territory: assessing the online frontier of student ratings of instruction. Assessment & Evaluation in Higher Education, 35(1), 97–111. doi: 10.1080/02602930802618336.

Appendix

Table A1. Original German item wordings in the present study, ordered corresponding to role in regression analysis.

Role in regression analysis	Item-content	Answer mode
Dependent variable	Wenn du an die vergangenen Lehrevaluationen zurückdenkst, wie häufig hast du alle in Frage kommenden Veranstaltungen evaluiert?	Ich habe insgesamt __ % meiner Veranstaltungen evaluiert.
Block 1 – demography	Geschlecht	weiblich / männlich
Block 1 – demography	Alter	*Open answer format*
Block 2 – social exchange	Ich kann mich als Student gut mit meiner Hochschule identifizieren.	7-point Likert scale (ranging from *stimme gar nicht zu* to *stimme vollständig zu*)
Block 2 – social exchange	Mit meinen Leistungen im Studium bin ich bislang zufrieden.	7-point Likert scale (ranging from *stimme gar nicht zu* to *stimme vollständig zu*)
Block 2 – social exchange	Mein Studiengang war meine erste Wahl.	7-point Likert scale (ranging from *stimme gar nicht zu* to *stimme vollständig zu*)
Block 2 – social exchange	Mit der Wahl meines Studiengangs bin ich zufrieden.	7-point Likert scale (ranging from *stimme gar nicht zu* to *stimme vollständig zu*)
Block 3 – salience	Von den Dozenten in meinem Fach wird viel Wert auf die Lehrevaluation gelegt.	7-point Likert scale (ranging from *stimme gar nicht zu* to *stimme vollständig zu*)
Block 3 – salience	In unserem Fach wird generell viel Wert auf die Lehrevaluation gelegt.	7-point Likert scale (ranging from *stimme gar nicht zu* to *stimme vollständig zu*)
Block 3 – salience	Meine Kommilitonen im Studium evaluieren alle anfallenden Veranstaltungen.	7-point Likert scale (ranging from *stimme gar nicht zu* to *stimme vollständig zu*)
Block 3 – salience	Wusstest du, dass die Lehrevaluation Konsequenzen für die Dozenten und die Lehre haben kann?	ja / nein
Block 4 – opportunity costs	Arbeitest du neben dem Studium?]	ja / nein
Block 4 – opportunity costs	Liegt dein derzeitiger Wohnort mehr als 10 km von der Uni entfernt?	ja / nein
Block 5 – survey mode	Wie wird die Lehrevaluation in deinem Fach erhoben?	mit Papierfragebögen / computergestützt / sowohl als auch
Block 6 – survey fatigue	Semester (gesamt)	*Open answer format*
Block 6 – survey fatigue	Wie war dein Evaluationsverhalten zu Beginn deines Studiums? Zu Beginn meines Studiums habe ich…	5-point Likert scale (ranging from *sehr viel weniger evaluiert* to *sehr viel häufiger evaluiert*)
Block 6 – survey fatigue	Im Laufe eines Semesters nehme ich an vielen Umfragen teil.	7-point Likert scale (ranging from *stimme gar nicht zu* to *stimme vollständig zu*)