This item is the archived peer-reviewed author-version of:

I liked your course because I believe in (the power of) student evaluations of teaching (SET) : students perceptions of a teaching evaluation process and their relationships with SET scores

Reference:

Spooren Pieter, Christiaens Wim.- I liked your course because I believe in (the power of) student evaluations of teaching (SET) : students perceptions of a teaching evaluation process and their relationships with SET scores
Studies in educational evaluation - ISSN 0191-491X - 54(2017), p. 43-49
Full text (Publisher's DOI): <https://doi.org/10.1016/J.STUEDUC.2016.12.003>
To cite this reference: <https://hdl.handle.net/10067/1429750151162165141>

I liked your course because I believe in (the power of) student evaluations of teaching (SET). Students' perceptions of a teaching evaluation process and its relationships with SET scores.

Abstract. Previous research has shown that many student, teacher, and course characteristics are related to SET scores. However, the relationship between students' perceptions of (the value of) of SET procedures and their SET scores has not yet been studied. At the University of Antwerp (Belgium), 974 students from 6 different academic disciplines completed both a SET form and an adapted version of the Students' Perceptions of a Teaching Evaluation Process Questionnaire (Surrat & Desselle, 2007). Structural equation modelling was used to examine the relationships between students' perceptions of a SET procedure and their SET scores. The analysis revealed a statistically significant positive relationship (.20) between a latent construct measuring the perceived value of SET and SET scores, as derived from a global factor measuring general teaching competency. Students who valued SET procedures tended to provide higher SET scores. It is important to recognize that based on their conceptions of their role as evaluators of university professors, students might differ in their perceptions of teaching competence.

Keywords: Student evaluation of teaching; validity; students' perceptions; teacher evaluation; higher education; structural equation modelling

Introduction¹

Student evaluations of teaching (SET) are the most common tool for assessing teaching in contemporary higher education (Knapper, 2001). Its widespread use could be attributed to the apparent ease of collecting data and presenting and interpreting results (Penny, 2003). Many institutions overly rely on SET data as the dominant and sometimes sole indicator of teaching quality (Onwuegbuzie et al., 2007). There are three main purposes of SET: (a) to improve teaching quality (i.e., feedback for the improvement of teaching), (b) to inform tenure/promotion decisions, and (c) to demonstrate an institution's accountability (by means of, among other things, adequate procedures for ensuring teaching quality) (Kember, Leung & Kwan, 2002).

Since the introduction of the first teacher-rating scale in 1915 (Marsh, 1987), SET instruments and procedures have provoked heated discussions among various stakeholders (i.e., policy makers, teachers, and students). In general, their concerns include (a) the differences in the ways in which students and teachers perceive effective teaching, (b) the relationships between SET scores and factors that are unrelated to “good teaching” (Centra, 2003; Marsh, 2007), (c) SET procedures and practices (e.g., the contents of SET reports, the depersonalization of the individual relationship between teachers and their students due to the standardized questionnaires and respondents' anonymity, the competency of SET administrators, low response rates), and (d) the psychometric properties of SET instruments. It is therefore not surprising that along with the introduction of SET procedures, a new research theme has emerged in the field of educational psychology and educational sciences. Moreover, Centra (2003) suggests

¹ The title of this paper was inspired by Remedios & Lieberman's (2008) paper, entitled “*I liked your course because you taught me well: The influence of grades, workload, expectations and goals on students' evaluations of teaching*”.

that “no method of evaluating college teaching has been researched more than student evaluations” (p. 495).

As a result, many comprehensive reviews are available (see, e.g., Marsh, 2007; Onwuegbuzie et al., 2009; Spooren, Brockx & Mortelmans, 2013). In a nutshell, these studies indicate that strong evidence has been provided for several areas of its validity. However, unsettled concerns on this matter include, among others, the absence of a general theory of college teaching upon which SET instruments can be grounded and the psychometric properties of many in-house-developed questionnaires that are used for both the improvement of teaching and personnel decisions.

As mentioned above, another major issue in the research field concerns the relationships between effective teaching, as measured by means of SET data, and factors that are unrelated to good teaching (including course characteristics, student characteristics, and teacher characteristics) (Centra & Gaubatz, 2000). Specifically, well-known potentially biasing factors include class size, the teacher’s gender, and course grades (although grade bias may also support the validity hypothesis concerning the positive relationship between grades and SET, suggesting that students who learned a lot during a course – and thus probably received higher grades – reward their teacher for his/her educational activities) (Marsh, 1987). This issue involves the discriminant validity and divergent validity of SET, which has received considerable attention in the research literature (Spooren, Brockx & Mortelmans, 2013). Most leading SET researchers are convinced of the validity of SET data because they have found potentially biasing factors to be of little or no influence (Beran & Violato, 2005; Marsh & Roche, 2000; Centra, 2003; Remedios & Lieberman, 2008; Spooren, 2010); nonetheless, such “bias studies” continue to play a central role in the literature.

The assumption of bias might affect teachers' attitudes toward SET, which are important for both obtaining and using SET data. After all, their usefulness for the improvement of teaching depends upon the extent to which teachers respond to SET and use them (Ballantyne, Borthwick, & Packer, 2000). Moore and Kuol (2005) observed two risks related to these responses: a fixation on minor issues (e.g., reworking the layout of a PowerPoint presentation) and demotivation, dejection, and withdrawal from the commitment to teaching effectiveness. Through interviews with faculty members, Yao and Grady (2005) found that teachers cared about student feedback from students, but they experienced anxiety and tension concerning the summative purposes of SET. Beran and Rokosh (2009) reported, based on a survey of 262 university teachers, that 84% of respondents supported the use of SET in general and that 62% felt that department heads and deans used SET reports properly.

Burden (2008, 2010) interviewed 22 award-winning teachers in tertiary education and observed a common recognition of the importance of SET. Nevertheless, only four of the teachers interviewed reported that SET led to improvement in their classes. Most teachers interviewed saw the feedback as nothing but little hints and tips, as the evaluations did not reflect their perspectives of good teaching. Likewise, Nasser-Abu Alhija and Fresko (2002) conducted a survey with 101 instructors at a teacher's college, finding that instructors considered SET of little value for the improvement of their teaching and that they made little or no use of students' feedback.

A recent study on this topic confirmed these distinct teachers' attitudes toward the effectiveness of SET and the concern about the aim and use of SET data (especially when used for personnel decisions) (Chan, Muk & Zeng, 2014). The authors therefore suggested that all stakeholders (teachers, students and administrators) should be fully

aware of the purposes of SET, which include both the improvement and evaluation of teaching.

Therefore, students' attitudes toward the goals of SET are also important in the collection of SET data. If students see no connection between their efforts in completing SET questionnaires and the outcomes of these evaluations (e.g., teacher awards or improvements in teaching or course organization), such evaluations may become yet another routine task, thus leading to mindless evaluation behaviour (Dunegan & Hrivnak, 2003).

In a study across 20 institutions, Kite, Subedi and Bryant-Lees (2015) found that students in general and across institutions, course disciplines, class standings and genders were positive about the SET process. Spencer and Schmelkin (2002) concurred; based on a mail survey administered to a random sample of students, they found that students were generally willing to participate in SET procedures. Nevertheless, students had little confidence that their evaluations were actually taken into account by administrators or teachers. A survey of pharmacy students performed by Surrat and Desselle (2007) revealed similar results, suggesting that students perceived SET as appropriate and necessary, although they were occasionally frustrated to see that their feedback did not improve ensuing teaching efforts.

Students are also ambivalent about the relative utility of the SET process. Chen and Hoshower (2003) observed that according to students, the most attractive outcome of a teaching-evaluation system is that the feedback provided improves teaching efforts. Their study showed that the motivation of students to participate in evaluations depended significantly on their expectations concerning the outcome. On the other hand, the idea that SET is also used for administrative decision-making was not found

to be an important motivator for student participation in SET surveys. Iqbal, Lee, Pearson and Albon (2016) found that students were motivated to participate in SET procedures when they perceived that the results would be used by the teachers. Nonetheless, both the timing of SET and the number of forms to be completed affected students' participation. Concerning the latter, a non-response analysis by Spooren and Van Loon (2012) revealed that among other student characteristics, such as course grades and seniority, students who were overburdened with a large number of SET surveys were less likely to participate. The authors suggested that administrators should be aware of this and monitor a fair dispersion of SET surveys during the academic year to avoid over-surveying or to use sampling strategies, as suggested by Kreiter and Laksham (2005).

From this overview, it is clear that students value SET procedures and are generally willing to participate in SET surveys. However, they think that in many cases, their feedback is only limitedly taken into account by their teachers. Researchers have not addressed the question of whether students' attitudes toward a teaching evaluation process affect their evaluation behaviour when completing a SET form. One could understand a student's perceived value of SET procedures to improve a teacher's instructional skills as an indication of *student engagement* (Appleton, Christenson, & Furlong, 2008) and *school engagement* (Fredricks, Blumenfeld, & Paris, 2004); more specifically, it may indicate engagement's social/behavioural/participatory component (i.e., engagement in school life) and emotional component (i.e., engagement of the heart, which includes how a student feels about the ways and the working of the school) (Yazzie-Mints, 2007). From this point of view, it can be argued that strongly engaged students will be more positive about their institution's teaching-evaluation system. Moreover, an analogous argument can be put forth concerning students' appreciation of

(teaching in) a course. Strongly engaged students, who perceive themselves as being an important stakeholder in the academic community, will be more positive about all activities (including teaching practices) that take place in that academic community. In this study, we tested this hypothesis and explored whether students' views of SET practice are correlated with their actual SET scores. The research questions for this study were:

- (a) How do students value SET practice in this institution? (RQ1)
- (b) Are there any correlations between students' perceived value of SET practice and their SET scores (even when controlling for a number of background variables)? (RQ2)

Method

Participants. Study participants included 974 undergraduate students from 6 different academic disciplines (science; social and political science; law; literature and philosophy; economics; and pharmaceutical, biomedical and veterinary sciences) at the University of Antwerp (Belgium). Participants completed both a SET form for one course in which they were enrolled the previous semester and an adapted version of the Students' Perceptions of a Teaching Evaluation Process Questionnaire (Surrat & Desselle, 2007) in class. Fifteen courses were involved in the study. The number of students who evaluated a course ranged from 16 to 183 (mean=65 students; SD=45; median= 62 students). All students were given the same instructions by the authors before completing both surveys. They were also asked for their identification number, which enabled the authors to link the two surveys with a number of demographic variables, including their grades for the course they evaluated and their academic discipline, gender, and seniority (freshman vs. sophomore at the time they participated in the course). All students were guaranteed anonymity and were assured that both their

responses on the surveys and their demographic data would be used only for the present study; 927 students (95%) agreed with this procedure and recorded their identification number at the top of both questionnaires. Participants were distributed in terms of gender (66% female), field of study (24% exact and natural sciences, 76% social sciences and humanities), student's seniority (26% freshmen, 74% sophomores) and their course grades for the course they evaluated in the present study (on a scale that ranges from 0 to 20 with mean = 12.86 and SE = 2.64). The sample represented 7% of the entire student body at this institution and approaches its distribution in terms of students' background characteristics (57% female students, 33% exact and natural sciences students, 15% freshmen), although female students, social sciences and humanities students and freshmen are slightly over-represented in the sample.

Measures. All students completed the SET37_QS questionnaire (Spooren, Mortelmans & Christiaens, 2014), which consists of nine closed-ended items scored on a six-point scale that ranges from 'strongly disagree' to 'strongly agree'. (It also includes two open-ended questions at the end of the questionnaire, which were not used in the analysis.) (Appendix A). Confirmatory factor analysis revealed a global factor behind the nine items in this instrument, measuring global teaching competence (Spooren, Mortelmans & Christiaens, 2014). All respondents then self-reported their perceived value of SET practice by means of the Students' Perceptions of a Teaching Evaluation Process Questionnaire (Surrat & Desselle, 2007), which consists of 15 items scored on a five-point scale that ranges from 'strongly disagree' to 'strongly agree'. This instrument was based on the literature, anecdotal information that the authors acquired from faculty members and students, and expert feedback. In this instrument, nine items elicit

students' perceptions of SET practice, and six items focus on students' behaviour when involved in a SET procedure.

Analytic strategy. With respect to RQ1, descriptive statistics were tabulated on the scores we collected with Surratt and Desselle's (2007) instrument to gauge students' perceptions of SET practice. Exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were used to determine the existence of any latent constructs behind the 15 items in this instrument. We also conducted a number of t-tests to determine whether there were statistical differences in these factors between male and female students, between science students and other students and between students with different seniority (i.e., freshmen vs. sophomores), which served as the controlling variables. A subsequent analysis used structural equation models (SEM) to test the correlations between the two variables, i.e., students' perceived value of SET practice and their SET scores in a course they had taken the previous semester (RQ2).

Results

How do students value SET practice in this institution (RQ1)? The item means, standard deviations (SD) and frequencies are reported in Table 1 for each item of Surratt and Desselle's questionnaire. In general, the majority of students were positive about SET, indicating that they viewed SET systems as appropriate and desirable. Moreover, almost 90% of the respondents (strongly) agreed that SET is necessary to provide accountability for teaching quality (item 1). Students also believed that SET might influence subsequent teaching efforts (i.e., the *formative* purpose of SET practice). However, there was more doubt on the question regarding whether professors actually take into account students' suggestions in the 'open comments' section of a SET survey

(item 6). Furthermore, students varied in their answers on the question regarding whether SET results should be included in a university's decision on promotion or tenure of professors (i.e., the *summative* purpose of SET practice).

Concerning their participation in SET procedures, students reported that their decision to participate in a SET survey was hardly affected by specific characteristics of the professor, such as likeability (item 13) or teaching skills (item 14). Moreover, most students indicated that their SET feedback was not strongly influenced by a professor's personality (item 12), the amount of course work (item 11), or the nature of the course being evaluated (item 15).

Table 1

Students' responses to survey items regarding students' perceptions of their teaching evaluations (Surratt & Desselle, 2007) (N = 974)*

	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree	Overall Response, Mean (SD)
Students' perception of SET						
1. Some form of student evaluation of teaching is necessary to provide accountability for teaching quality.	8 (0.8)	19 (2.0)	88 (9.1)	505 (52.0)	351 (36.1)	4.21 (0.75)
2. SET feedback influences the quality of the professor's future teaching efforts.	11 (1.1)	42 (4.4)	238 (24.8)	499 (51.9)	171 (17.8)	3.81 (0.82)
3. SET results should not be a significant factor in the university's decision to promote/tenure a professor.	77 (8.0)	281 (29.2)	239 (24.8)	268 (27.8)	98 (10.2)	3.03 (1.14)
4. All faculty should be SET-evaluated for all courses in which their contribution was substantial.	7 (0.7)	53 (5.5)	176 (18.3)	548 (56.8)	180 (18.7)	3.87 (0.80)
5. Senior faculty need not be SET-evaluated as frequently as junior faculty with less teaching experience.	396 (40.9)	389 (40.1)	89 (9.2)	78 (8.0)	17 (1.8)	1.90 (1.00)

6. Most professors read and consider the comments section of the SET results.	24 (2.5)	141 (14.9)	531 (56.1)	229 (24.2)	22 (2.3)	3.09 (0.76)
7. From my understanding of when a professor receives SET feedback, it is possible that the professor may retaliate on the final exam after receiving poor SET scores.	203 (21.2)	308 (32.1)	239 (24.9)	163 (17.0)	46 (4.8)	2.52 (1.14)
8. It is appropriate for students to judge teaching effectiveness.	17 (1.8)	13 (1.3)	62 (6.4)	555 (57.3)	322 (33.2)	4.19 (0.76)
9. Professors who receive the best SET scores are not necessarily the most effective teachers.	63 (6.6)	193 (20.1)	305 (31.7)	334 (34.8)	66 (6.9)	3.15 (1.03)
	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree	Overall Response, Mean (SD)
Influences on students' evaluations						
10. The professor's personality influences my SET scoring.	42 (4.3)	110 (11.3)	202 (20.7)	494 (50.9)	122 (12.6)	3.56 (0.99)
11. I tend to give better SET scores in courses that require less work.	200 (20.6)	488 (50.4)	151 (15.6)	109 (11.2)	21 (2.2)	2.24 (1.00)
12. I complete the SET evaluation every time I am given the opportunity.	46 (4.8)	112 (11.6)	252 (26.1)	381 (39.4)	175 (18.1)	3.55 (1.06)
13. I am more likely to complete the SET evaluation when I really like or really dislike the professor (as opposed to neutral feelings).	213 (22.0)	340 (35.2)	164 (17.0)	157 (16.2)	93 (9.6)	2.56 (1.26)
14. I am more likely to complete the SET evaluation when I feel the professor taught especially well or especially poorly.	203 (21.0)	333 (34.4)	165 (17.1)	163 (16.9)	103 (10.7)	2.62 (1.28)
15. The nature of the course (subject matter) influences my SET scoring.	52 (5.4)	256 (26.5)	227 (23.5)	368 (38.1)	64 (6.6)	3.14 (1.05)

Note. The results show absolute frequencies and valid percentages between brackets.

*Since not all students responded to all statements, the number of responses for each item varied from 947 to 971.

A priori t-tests revealed a number of statistically significant differences between groups of students based on their gender, seniority, and academic discipline. Concerning *student's gender*, male students ($M_{\text{item5}}=2.05$, $SD_{\text{item5}}=1.07$; $M_{\text{item9}}=3.33$, $SD_{\text{item9}}=1.06$)

scored higher on item 5 ($t(879)=2.87, p=.004$) and item 9 ($t(869)=3.15, p=.002$) than female students ($M_{\text{item}5}=1.83, SD_{\text{item}5}=.93; M_{\text{item}9}=3.09, SD_{\text{item}9}=.99$), indicating that they tended to agree more with the ideas that senior faculty should not be evaluated as much as junior faculty and that the best professors are not always given the highest SET scores by their students. With respect to the latter, we also found a statistically significant difference concerning *academic discipline*: Students in the field of natural and life sciences ($M_{\text{item}9}=2.94, SD_{\text{item}9}=1.11$) scored lower on this item ($t(959)=3.40, p=.001$) than students from other academic disciplines ($M_{\text{item}9}=3.22, SD_{\text{item}9}=1.00$). Students' *seniority* also revealed statistical differences: Sophomores ($M_{\text{item}8}=4.23, SD_{\text{item}8}=.73; M_{\text{item}12}=3.60, SD_{\text{item}12}=1.04; M_{\text{item}14}=2.54, SD_{\text{item}14}=1.30$) scored higher on item 8 ($t(967)=2.79, p=.005$) and item 12 ($t(964)=2.68, p=.008$) and lower on item 14 ($t(965)=3.17, p=.002$) compared to freshmen ($M_{\text{item}8}=4.07, SD_{\text{item}8}=.83; M_{\text{item}12}=3.39, SD_{\text{item}12}=1.11; M_{\text{item}14}=2.83, SD_{\text{item}14}=1.21$). This suggests that mature students are more confident of their role as evaluators of teaching practice and are more likely to participate in SET procedures.

In a second step, we tested the assumed factor structure of this instrument by means of a confirmatory factor analysis (CFA) that included two factors (i.e., students' perceptions of SET practice (item 1-9) and students' behaviour when involved in a SET procedure (item 10-15)). As the data have a clear hierarchical structure (i.e., students are nested in courses, suggesting that individual observations are not completely independent) (Nasser-Abu Alhija & Hagtvet, 2006; Spooren, 2010; Wagner et al., 2013), this should be considered when analysing the data (Kreft & De Leeuw, 1998; Hox, 2010). A full multilevel SEM procedure, as proposed by Muthén (1994), was not possible due to the small number of units at the second level ($n = 15$). However, we took into account this complex sample structure by running a CFA model with the 'type

= complex feature' in MPlus software (Muthén & Muthén, 1998–2010), which specified the course level in the analysis and allowed the application of corrected standard errors for the study variables. The results yielded a questionable fit to the data ($\chi^2=472.39$; $df=89$, $p<.0001$; $RMSEA=.07$; $SRMR=.08$, $CFI=.76$), and item loadings on the targeted factors were very low ($<.40$) for nine items. The conceptual grounds of the instrument were thus not confirmed in the data. Therefore, we decided to explore the data by means of both exploratory and confirmatory factor analysis and searched for any factors (of both conceptual and empirical relevance) measuring students' perceived value of SET that could be included in our further analysis. Our sample was split randomly into two datasets, which enabled us to run an number of EFA-models to extract one or more factors, that then could be tested and confirmed via a CFA-model.

In the first sample, all 15 items were submitted to an exploratory factor analysis (EFA) using principal axis factoring and a varimax rotation (SPSS version 22.0), with the additional requirement that factor loadings were higher than .40 (Field, 2009). The selection of factors was based on the Kaiser criterion, the scree plot, and a minimum of three variables loading on one factor (Velicer & Fava, 1998). This left us with one factor consisting of four items that measure students' perceived value of SET. The total explained variance of this factor equalled 36.39%, and its internal consistency was acceptable (Cronbach's coefficient: .68). The remaining items in this one-factor model are presented in Table 2. All of these items obviously measure student's perceptions on the use and the relevance of SET-procedures. Students who score high on this factor, are bystanders of the use of SET for both monitoring teaching quality (items 1,item 4), and improving teaching (item 2). They also believe that students are able to provide appropriate evaluations of their teachers (item 8).

Table 2

Factor loadings for exploratory factor analysis with varimax rotation of students' perceived value of SET (Sample 1, N=467).

Item N°	Item	Factor loading
1	Some form of student evaluation of teaching is necessary to provide accountability for teaching quality.	.78
2	SET feedback influences the quality of the professor's future teaching efforts.	.56
4	All faculty should be SET-evaluated for all courses in which their contribution was substantial.	.52
8	It is appropriate for students to judge teaching effectiveness.	.52

Confirmatory factor analysis (CFA) in the second sample, which took into account the nested structure of the data, confirmed this factor structure and provided an adequate fit to the data ($\chi^2=3.81$; $df=2$, $p=.14$; $RMSEA=.04$; $SRMR=.02$, $CFI=.99$) (Hu & Bentler, 1995; MacCallum, Browne, & Sugawara, 1996). The standardized factor loadings of the items, corrected standard errors, and intra-class correlations are shown in Table 3. All loadings are statistically significant and substantial. Intraclass correlations for all items are lower than .05, suggesting little variance at the class level (Snijders & Bosker, 1999). With respect to this factor, ignoring the nested structure of the data (because of the low number of units at the class level) will thus not cause much bias in the model fits or parameter estimates in the subsequent analyses (Pornprasertmanit, Lee, & Preacher, 2014).

Table 3

Standardized factor loadings (Est.), corrected standard errors (SE), and intraclass correlations (ICC) of the 'perceived value of SET' factor (Sample 2, N=507)

Item N°	Est.	SE	ICC
1	.73	.06	.03
2	.58	.06	.02
4	.44	.05	.01
8	.54	.09	.01

Are there any correlations between students' perceived value of SET practice and their SET scores (RQ2)? To explore the relationships between students' perceived value of SET procedures (as measured by the four items from the previous analysis) and their SET scores (as measured by the nine items and the global factor in the SET37_QS instrument), we set up a number of structural equation models that included both the relationship between these variables and a number of exogenous factors at the student level, including course grade (for the course they evaluated; grades ranged between 0 and 20), academic discipline (1=natural and life sciences, 0=other disciplines), gender (1=female, 0=male), and seniority (1=freshman, 0=sophomore). All models were estimated with Lisrel8 software (Jöreskog & Sörbom, 1993). Since many variables are scored on an ordinal scale (five-point scale in the 'perceived value of SET' factor and six-point scale in SET37_QS), we used asymptotic covariance matrices based on polychoric correlations and weighted least square estimations (p. 45).

The results for the final model are presented in Figure 1. To scale the latent constructs, in each scale, the item with the highest loading was fixed to 1. All other loadings were set free. Only the statistically significant effects are shown. (Students' gender, for instance, showed no statistically significant effects.)

Note. GRADE = course grade; DISC = academic discipline; FRESH = seniority; PERC = perceived value of SET practice; P1-P2-P4-P8 = items in the ‘perceived value of SET’ factor; SET = global SET score derived from the items in the SET37_QS instrument; QS1-9 = items in the SET37_QS instrument.

Figure 1

Factors affecting the perceived value of SET and SET scores (N=927 students)

The model shows an acceptable fit to the data ($\chi^2=474.81$; $df=97$, $p<.0001$, $RMSEA=.06$, $SRMR = .09$, $GFI=.97$) and reveals a statistically significant positive relationship between student’s perceived value of SET practices and her SET scores (.20). This finding seems to confirm our hypothesis that the more a student values SET practice, the higher his/her SET scores will be. Furthermore, we observe a statistically significant effect of a student’s course grade on both his SET scores and his perceived value of SET practice. The former is in line with findings from previous SET research that also reported a small to moderate correlation between SET scores and student achievement, as measured by course grades (see, e.g., Feldman, 1989; Marsh & Roche, 2000; Spooren, 2010). Academic discipline also plays a role. We found statistically significant positive effects of this predictor on both variables, i.e., students from natural and life sciences valued SET practice more and issued higher SET scores than students

from other disciplines. Moreover, freshmen issued higher SET scores than sophomores. Nonetheless, all of the reported effects are rather small (ranging between .05 and .20).

Discussion and Conclusion

The aim of the present study was to explore the relationship between students' views of SET practice and their SET scores. The research aims for this study were (a) to deepen our insight into students' perceived value of SET practice at our institution and (b) to determine whether this perceived value of SET practice is related to students' actual SET scores.

Concerning the first question, we found that students are generally positive about SET, and they agree that SET can provide accountability for teaching quality. They also believe that SET might influence future teaching practice. However, they have their doubts concerning the actual use of students' comments by their professors. When asked about the use of SET reports for administrative decision-making (such as tenure or promotion), their answer remains unclear; students do not agree with each other with respect to statements on this matter. These findings are in line with previous studies on this topic. Moreover, we found that a number of student-related characteristics, such as gender, seniority, and academic discipline, influence their thoughts on the value of SET practice.

Concerning the second question, we found a statistically significant relationship between SET scores and students' perceived value of SET practice. Students who value SET procedures tend to provide higher SET scores. This might confirm our hypothesis that strongly engaged students are more positive about all activities in their institution, such as teaching and SET practice. Additionally, both SET scores and the perceived

value of SET practice are correlated with by students' grades, seniority and academic discipline. Nonetheless, all of the reported effects in the final model are rather small, ranging between .05 and .20. This is consistent with many previous studies on the relationships between student-related characteristics and SET scores. This supports the argument that SET scores depend primarily upon teaching behaviour (Barth, 2008; Greimel-Fuhrmann & Geyer, 2003). It is, however, important to recognize that based on their conceptions about their role as evaluators of university professors, students might differ in their perceptions of their professors' teaching competence.

The present study should be viewed as exploratory as it has a number of important limitations. First, as far as we know, this is the first study on the possible relationship between SET scores and students' perceived value of SET. Similar studies in other institutions are needed to confirm the results of this baseline study, which was performed in a particular institution using particular instruments. Second, the small number of courses ($N = 15$) used in the procedure did not allow us to take into account the hierarchical structure of the data in the most appropriate way. Subsequent procedures should use larger samples to confirm the results of the present study in a multilevel design. Third, we used only one SET form per student. Future analysis on this topic could combine several SET forms per student to confirm the existence of a 'trait', i.e., whether students who value SET practice indeed provide higher SET scores in *all* courses they evaluate. Fourth, it might be interesting to include a measure of 'student engagement' or 'student involvement' in future research studies to find out whether our assumption on the positive relationship between SET scores and students' perceived value of SET practice indeed is mediated by the extent to which a student feels that he or she is part of the academic community.

References

- Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools*, 45, 369-386. doi: 10.1002/pits.20303
- Ballantyne, R., Borthwick, J., & Packer, J. (2000). Beyond student evaluation of teaching: Identifying and addressing academic staff development needs. *Assessment and Evaluation in Higher Education*, 25, 221-236. doi:10.1080/713611430
- Barth, M. M. (2008). Deciphering student evaluations of teaching: A factor analysis approach. *Journal of Education for Business*, 84, 40-46. doi:10.3200/JOEB.84.1.40-46
- Beran, T. N., & Rokosh, J. L. (2009). Instructor's perspectives on the utility of student ratings of instruction. *Instructional Science*, 37, 171-184. doi:10.1007/s11251-007-9045-2
- Beran, T. N., & Violato, C. (2005). Ratings of university teacher instruction : How much do student and course characteristics really matter ? *Assessment and Evaluation in Higher Education*, 30, 593-601. doi:10.1080/02602930500260688.
- 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
- Burden, P. (2010). Creating confusion or creative evaluation? The use of student evaluation of teaching surveys in Japanese tertiary education. *Educational Assessment, Evaluation and Accountability*, 22, 97-117. doi: 10.1007/s11092-010-9093-z
- Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, 44, 495-518. doi:10.1023/A:1025492407752.
- Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching ? *The Journal of Higher Education*, 71, 17-33. Retrieved from www.jstor.org/stable/2649280

- Chan, C.K.Y., Luk, L.Y.Y., & Zeng, M. (2014). Teacher's perceptions of student evaluations of teaching. *Educational Research and Evaluation*, 20, 275-289.
doi:10.1080/13803611.2014.932698
- Chen, Y., & Hoshower, L. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment and Evaluation in Higher Education*, 28, 71-88. doi: 10.1080/02602930301683
- Dunegan, K. J., & Hrivnak, M. W. (2003). Characteristics of mindless teaching evaluations and the moderating effects of image compatibility. *Journal of Management Education*, 27, 280-303. doi:10.1177/1052562903027003002
- Feldman, K. A. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education*, 30, 583-645. doi: 10.1007/BF00992392
- Field, A. P. (2009). *Discovering statistics using SPSS: and sex and drugs and rock 'n' roll* (Third ed.). London: Sage Publications.
- Floyd, F. L., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7, 286 – 299.
- Fredricks, J.A., Blumenfeld, P.C., & Paris, A.H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74, 59-109. doi: 10.3102/00346543074001059
- Greimel-Fuhrmann, B., & Geyer, A. (2003). Students' evaluation of teachers and instructional quality – Analysis of relevant factors based on empirical research. *Assessment & Evaluation in Higher Education*, 28, 229-238.
doi:10.1080/0260293032000059595

- Hox, J. J. (2010). *Multilevel analysis. Techniques and applications (2nd ed.)*. Mahwah, NJ: LEA.
- Hu, L. T., & Bentler, P.M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.) *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). Thousand Oaks, CA: Sage.
- Iqbal, I., Lee, J. D., Pearson, M. L., & Albon, S. P. (2016). Student and faculty perceptions of student evaluations of teaching in a Canadian pharmacy school. *Currents in Pharmacy Teaching and Learning*, 8, 191-199. doi:10.1016/j.cptl.2015.12.002
- Jöreskog, K., & Sörbom, D. (1996). *Lisrel 8: User's reference guide*. Chicago, IL: Scientific Software International.
- 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
- Kite, M. E., Subedi, P. C., & Bryant-Lees, K. B. (2015). Students' perceptions of the teaching evaluation process. *Teaching of Psychology*, 42, 307-314. doi:0098628315603062.
- Knapper, C. (2001). Broadening our approach to teaching evaluation. *New Directions for Teaching and Learning*, 88, 3-9. doi:10.1002/tl.32
- Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA/ Sage Publications.
- 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
- MacCallum, R. C., Browne, M.W., & Sugawara, H.M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130-149. doi:

- Marsh, H. W. (1987). Student's evaluations of university teaching: Research findings, methodological issues, and directions for further research. *International Journal of Educational Research*, 11, 253-388. doi:10.1016/0883-0355(87)90001-2
- 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.
- Moore, S., & Kuol, N. (2005b). A Punitive Bureaucratic Tool or a Valuable Resource? Using Student Evaluations to Enhance Your Teaching. In G. O'Neill, S. Moore, & B. McMullin (Eds.), *Emerging Issues in the Practice of University Learning and Teaching. Part 3: Developing and growing as a university teacher* (pp. 141-146). Dublin, University of Limerick: All Ireland Society for Higher Education (AISHE) Readings.
- Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluation of teaching: Popular myth, bias, validity or innocent bystanders? *Journal of Educational Psychology*, 92, 202-228. doi:10.1037/0022-0663.92.1.202
- Muthén, B. (1994). *Multilevel Covariance Structure Analysis. Sociological Methods & Research*, 22, 376-398. doi:10.1177/0049124194022003006
- Muthén, L.K., & Muthén, B. (1998-2010). *MPlus User's Guide (6th ed.)*. Los Angeles, CA: Muthén & Muthén.
- Nasser-Abu Alhija, F., & Hagtvet, K.A. (2006). Multilevel analysis of the effects of student and instructor/course characteristics. *Research in Higher Education*, 47, 559-590. doi: 10.1007/s11162-005-9007-y
- Nasser-Abu Alhija, F., & Fresko, B. (2009). Student evaluation of instruction: what can be learned from students' written comments? *Studies in Educational Evaluation*, 35, 37-44. doi:10.1016/j.stueduc.2009.01.002

- Onwuegbuzie, A. J., Witcher, A. E., Collins, K. M. T., Filer, J. D., Wiedmaier, C. D., & Moore, C. W. (2007). Students' perceptions of characteristics of effective college teachers: A validity study of a teaching evaluation form using a mixed-methods analysis. *American Educational Research Journal*, 44, 113-160. doi:10.3102/0002831206298169
- Onwuegbuzie, A. J., Daniel, L. G., & Collins, K. M. T. (2009). A meta-validation model for assessing the score-validity of student teaching evaluations. *Quality & Quantity*, 43, 197-209. doi:10.1007/s11135-007-9112-4
- Pornprasertmanit, S., Lee, J., & Preacher, K. J. (2014). Ignoring clustering in confirmatory factor analysis: some consequences for model fit and standardized parameter estimates. *Multivariate behavioral research*, 49, 518-543. doi: 10.1080/00273171.2014.933762
- Remedios, R., & Lieberman, D. A. (2008). I liked your course because you taught me well: The influence of grades, workload, expectations and goals on students' evaluations of teaching. *British Educational Research Journal*, 34(1), 91-115. doi: 10.1080/01411920701492043
- Snijders T. A. B. & Bosker R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modelling*. London: Sage Publications.
- Spencer, K. J., & Schmelkin, L. P. (2002). Student perspectives on teaching and its evaluation. *Assessment & Evaluation in Higher Education*, 27, 397-409. doi:10.1080/0260293022000009285
- 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
- Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching. A state of the art. *Review of Educational Research*, 83, 598-642. doi: 10.3102/0034654313496870

- Spooren, P., Mortelmans, D., & Christiaens, W. (2014). Assessing the validity and reliability of a quick scan for student's evaluation of teaching. Results from confirmatory factor analysis and G Theory. *Studies in Educational Evaluation*, 43, 88-94.
doi:/10.1016/j.stueduc.2014.03.001
- Spooren, P., & Van Loon, F. (2012). Who participates (not)? A non-response analysis on students' evaluations of teaching. *Procedia of Social and Behavioral Sciences*, 69, 990-996. doi: 10.1016/j.sbspro.2012.12.025
- Surratt, C. K., & Desselle, S. P. (2007). Pharmacy students' perceptions of a teaching evaluation process. *American Journal of Pharmaceutical Education*, 71(1), Article 06.
- Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, 3, 231-251. doi: 10.1037/1082-989X.3.2.231
- Yao, Y., & Grady, M. (2005). How do faculty make formative use of student evaluation feedback? A multiple case study. *Journal of Personnel Evaluation in Education*, 18, 107-126. doi:10.1007/s11092-006-9000-9
- Yazzie-Mintz, E. (2007). *Voices of students on engagement: A report on the 2006 High School Survey of Student Engagement*. Bloomington: Center for Evaluation & Education Policy, Indiana University. Retrieved January 26, 2016, from <http://files.eric.ed.gov/fulltext/ED495758.pdf>
- Wagner, W., Göllner, R., Helmke, A., Trautwein, U., & Lüdtke, O. (2013). Construct validity of student perceptions of instructional quality is high, but not perfect: Dimensionality and generalizability of domain-independent assessments. *Learning and Instruction*, 28 (December 2013), 1-11. <http://dx.doi.org/10.1016/j.learninstruc.2013.03.003>