

International Conference on Education and Educational Psychology (ICEEPSY 2012)

The student as a commentator: students' comments in student evaluations of teaching

Brockx, B^{a*}., Van Roy, K^b., Mortelmans, D^a.

^aUniversity of Antwerp, Sint-Jacobsstraat 2, 2000 Antwerp, Belgium

^bLessius University College, Zandpoortvest 13, 2800 Mechelen, Belgium

Abstract

Little is known about students' comments in student evaluations of teaching. In the current study on around 70% of the SET-surveys received, a comment was written by a student. Most students' comments deal with, a good combination between theory and practice, the build-up of the course, and whether it is interesting and relevant. They emphasize aspects already asked for in the fixed-ended questions, but also provide information that was not part of the SET-survey. The results of a logistic regression analysis showed a consistency between students' comments and SET-scores. Surveys that were received with low-SET scores have a larger chance to receive negative comments about various aspects of the course. In general students seems to take their task as a commentator, seriously. As to the content of the comments received, the context plays an important role. This indicates that more research in different contexts is necessary and that one should take into account this context when conducting SET.

© 2012 The Authors. Published by Elsevier Ltd.

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

Keywords: Student evaluations of teaching, Students' comments, content analysis., qualitative research

1. Introduction

Student evaluations of teaching (SET) is the most widely spread method to conduct teacher evaluation at institutions for higher education (Arnold, 2009). Different survey designs exist to conduct SET, but most often a SET survey instrument consists of an amount of fixed-ended questions on which the student rates the teacher. The scores on these fixed-ended questions are then used to generate a statistical report. Regularly these fixed-ended questions are supplemented with some open-ended questions. There are several reasons to include open-ended questions in a SET-questionnaire. Open-ended questions allow a greater freedom of expression and are less influenced by the researcher as students don't have to fit their answer in standard answer categories (Foddy, 1994). Although open-ended questions allow for a greater range of expression and valuable information might get lost in answers to fixed-ended questions, little research has focused upon the answers given to these questions (Alhija & Fresko, 2009). This might have to do with the fact that interpreting written student comments is difficult as they appear idiosyncratic and

* Corresponding author. Tel.: +32 326555339

E-mail address: bert.brockx@ua.ac.be

anecdotal, seeming a series of random unconnected utterances on instructors, their teaching and the educational environment (Lewis, 2001, pp. 25-26; Hodges & Stanton, 2007). They need to be analyzed to reveal the aforementioned information and insights. Analyzing the answers requires an extensive time-consuming coding process (Ahrens & Pigeot, 2005). To add to the objectivity of the research inter-coder reliability is an important factor in this coding process (Johnson & Christensen, 2012).

In SET students' comments reflect the complexity of teaching environments and the way each student reacts to specific environments or teaching strategies (Lewis 2001, p. 26). In that perspective - as different authors (Alhija & Fresko, 2009; Hodges & Stanton, 2007; Smith & Welicker-Pollak, 2008) pointed out - comments can yield more and useful information on important educational issues and provide better insights in order to improve them. For instance, student comments are more specific than a statistical report and can contain actual ideas to improve teaching (Hammond, Taylor, & McMenamin, 2003). An analysis of students' comments can provide useful insight into aspects of teaching and courses that students find important, and verify whether these aspect are different from the ones that are generally measured in SET-instruments (Symmons, 2006). Furthermore, students' comments in SET are important for teachers. Some studies indicate that instructors prefer written comments (as opposed to feedback based on statistical data) (Centra, 1993; Lewis, 2001; Svinicki, 2001). The answers to these questions have a greater potential to influence them, than statistical reports (Alhija & Fresko, 2009).

Only a small body of knowledge exists about students' comments in SET, most studies used a quantitative approach and focused on the frequency of the comments that were received. These studies showed great variation in the percentage of students that took the effort to write down a comment. In some studies the percentage of students that entered or wrote down a comment was low (10-12%) (Theall & Franklin, 1991), in other studies the proportion of students giving a comment was between 40 and 50% (Alhija & Fresko, 2009; Hardy, 2003) while in others it was even between 60–70% (Oliver, Tucker, & Pegden, 2007). The frequency of the comments received is also related to the direction of the comments. In general students are more eager to provide positive comments than negative ones (Alhija & Fresko, 2009). Although the amount of received comments may be less for negative comments, the negative comments are more specific and offer more detail (Alhija & Fresko, 2009).

Still, little is known about the content. When consulting the literature on students comments in SET, the study by Alhija & Fresko (2009) is the only recent study in an international refereed journal that focused on the content of the students' comments that was found. The results of this study showed that comments could be classified into three main domains, eight primary content areas, and forty five secondary content areas: the course (content, assignments, and general), the instructor (teaching style, personal traits, and general) and the context of instruction (scheduling issues and student composition). Six of the 45 secondary content categories used in the study were mentioned by more than 10% of the students: general evaluation of the instructor, general evaluation of the course, interest generated by the course, interest generated by the instructor, contribution to learning, and clarity of instruction.

As mentioned above an analysis of students' comments can provide useful insight into aspects of teaching and courses that are important to students. One can even view them as a window into the process of student learning and intellectual development, and as a result from their own personal epistemological beliefs (Hodges & Stanton, 2007). As "Schools shape and change beliefs both as purveyors of knowledge and as epistemological training grounds for developing students" (Schraw, 2001, p. 460) it could be expected that the content of students written comments also depends on the type and nature of their study field. Especially with regard to the study field 'Education and training', the hypothesis could be postulated that the comments of those students should differ from the comments of students of other study fields since research shows that when programs are focused on encouraging students to reflect on epistemological beliefs, they become more metacognitively aware and their epistemological beliefs change (Brownlee, Purdie, & Boulton-Lewis, 2001). It is possible to refer to the model of personal epistemology proposed by Bendixen and Rule (2004) wherein metacognition oversees the model, including the mechanism of change. In fact, those others propose that the more metacognitively aware a student is, the more lasting belief change may be (Bendixen & Rule, 2004, p. 74). The hypothesis mentioned above is worth to be looked at as a considerable part of the courses of the students of our sample belongs to a program in the study field 'Education and training' and deals whit reflection on epistemological beliefs.

Several authors found correlations between SET-scores and other measures of teaching quality (e.g. alumni rating, evaluations by trained observers) (Richardson, 2005; Roche & Marsh, 2000). The results of these studies provide evidence for the convergent validity of SET. Since the comments in students' evaluations of teaching cannot be seen as a different measure of teaching quality as students' comments and SET-scores are collected by the same instrument, on the same moment and are provided by the same student, convergent validity is not under consideration. Still, a consistency between student ratings and students' comments can be expected, which can provide evidence for the fact that students take SET seriously and provide consistent feedback. The following hypothesis can be postulated: Comments differ for students who's evaluations result in low, average or high SET-scores and there is a consistency between these comments and the SET-score..

The research about students' comments in SET is still limited and results (for example about the proportion of comments received) show great variation. Moreover, the narrow body of knowledge about student written comments is limited to studies conducted at universities in the United States, Israel and Australia (Alhija & Fresko, 2009). The aim of the present study is to contribute to this relatively small body of knowledge surrounding written student comments. The context and the SET-instrument being used in the current study is different from other studies. The results of the current study can be compared with previous studies and add to the reliability of previous and current findings.

Based on the literature described above and the hypotheses formulated, the current study will provide an answer to five research questions:

1. How many comments are received?
2. What are the areas and dimensions commented upon; is there a difference between positive and negative comments?
3. Do students' comments deal with other aspects, than the dimensions that are measured by fixed-ended questions of the survey?
4. Do comments from students of the study field of 'education and training' differ from comments in the other study fields?
5. Is there a certain consistency between SET-scores and the comments that are written by students?

2. Method & Sample

2.1. Sample and procedure

The sample used in the present study holds 3742 SET- surveys collected at the Lessius Mechelen University College in Flanders (Belgium) during the fall and spring semester of the 2009-2010 academic year. The (undergraduate) courses concerned include both introductory and advanced courses hailed from different departments in 4 study fields, i.e. Business & Communication, Architecture, Education and Training and Health which offer a wide range of bachelor degree programmes in areas such as nursing, teacher education, interior design, business management and commercial sciences, information management, tourism, communication, journalism, etc.

For every phase of the educational programs between 1 and 3 courses were randomly selected each semester, which is the policy of the college with respect to course evaluation. Being informed about the survey several weeks before, the students were invited by e-mail immediately after their examinations to complete the questionnaire using a link to an online survey. (For the university college addressed in this study, examinations are an important part of the curriculum, and thus should be part of the evaluation as well.) Participation in the survey was not compulsory and was controlled using 'once-only' tokens for each student enrolled. In accordance with college policy the student's anonymity was maintained.

For further analysis the evaluations were merged with an administrative dataset, which included several student and course characteristics. The descriptive statistics are given in table 1.

Table 1 Descriptive statistics of the student and course characteristics

		N	%			N	%
Course year	First year course	840	45,8	Study field	Education & training	570	31.1
	Second year course	517	28,2		Other	1265	68.9%
	Third year course	478	26				
Course grade	Failed (< 50%)	361	19.7	Gender	Female	1314	71.6
	Average	1188	64.7		Male	521	28.4
	High (> 75%)	286	15.6				

2.2. (Research) Instrument

The course evaluations were collected using an adjusted version of the SET-37 questionnaire, which was developed at the University of Antwerp (Belgium), the validity and reliability of this instrument has been proved (D. Mortelmans & P. Spooen, 2009; Spooen, Mortelmans, & Denekens, 2007). The adjustment relates to the professional oriented characteristics of study programs and associated learning environments. The empirical test results - obtained under the same requirements of validity and reliability as the original version of the instrument - provided evidence of the (adjusted) instrument's reliability, validity and applicability in the new context (D Mortelmans & P Spooen, 2009; D. Mortelmans & P. Spooen, 2009).

The evaluation questionnaire used comprised two parts. In the first part students were presented with 40 Likert items representing 13 quasi-balanced scales and measured along a six-point scale ranging from 1 (completely disagree) to 6 (completely agree). The second part of the evaluation form contained the two following open-ended questions: "Which aspects did you find positive about this course and should be kept?" and "Which aspects did you find negative about this course and should be improved?"

Confirmatory factor analyses have repeatedly shown that an underlying factor arises in the SET-37 instrument (Brockx, Spooen, & Mortelmans, 2011; Spooen, 2010). This higher-order factor influences the dimensions that measure how the teacher builds, organizes, and executes the course it explains a large part of the variation in seven² underlying dimensions (22 items). The SET-score that'll be used in the present study is the sum-score on this higher-order factor. This sum-score was transformed into a categorical variable, with categories low (< 25th percentile), average and high (> 75th percentile) SET-scores.

2.3. Analytical strategy

To provide an answer to the five research questions, quantitative as well as qualitative research techniques are applied.

2.3.1. Qualitative analysis

As student written comments are printed material and may also be considered as a form of interviewing (Lewis 2001, p26); an inductive approach based on Grounded Theory as elaborated by Strauss and Corbin (1998) is applied. Scrutinizing the data, a codebook was constructed in an iterative process by coding and recoding a part of the comments several times. Two researchers conducted each of these research activities simultaneously and independently. The outcome of each step was discussed and the process resulted in a coding scheme consisting of 25 subcategories (secondary content area) grouped in 7 categories (primary content area) which are related to 4

² These seven underlying dimensions are as follows: clarity of objectives, value of subject matter, build-up of subject matter, presentation skills, contribution (course materials) to understanding of the subject, course difficulty, and help provided by the teacher during the learning process.

dimensions (overall evaluation, instructor, course and organizational context).

Using the above codebook and a new sample (494 positive comments and 470 negative), which was rated by each rater independently, the inter-rater reliability was examined. The Kappa value was calculated and showed a good inter-rater reliability: Kappa=.77 for positive comments, .81 for negative comments. These results were considered satisfactory (Landis & Koch, 1977). Finally, the coding of the remaining comments (1655 positive and 1573 negative) was completed. During the coding process one comment could receive multiple codes, when the comment dealt with multiple aspects of the course.

2.3.2. *Quantitative analysis*

The prediction of receiving certain comments or not involves a dichotomous dependent variable. The proper analytical tool to use, is then logistic regression (Carbera, 1994). Due to the hierarchical structure of the data the most appropriate technique to analyze the data, would be logistic multilevel analysis (Hox, 2010). Because the category that was coded most (a positive combination of theory and practice) was only mentioned 291 times or in 15.8% of the comments (n=1835), estimating multilevel models lead to serious iteration problems. Therefore a regular logistic regression analysis was used. For all analysis assumptions to perform logistic regression analysis were checked. First the problem of zero cells was examined (Menard, 2009). Zero cells cause instability and lead to possible unreliable results (Larose, 2006). In a next step the multicollinearity problem was examined by checking the VIF statistics and tolerance (DeMaris, 2004). In the last step the residuals were examined, the outliers (Pearson standardized residuals), leverage points (leverage values) and the influential cases (Santner & Duffy, 1989).

3. Results

3.1. *How many comments are received?*

On 2629 (70%) of the SET-surveys that were received, a comment was written. The difference between negative and positive comments is small on 2355 (62.85%) surveys a positive comment was written and on 60% (2245) a negative.

3.2. *What do these comments deal with?*

In table 2 the codebook is presented. The results show that 17.5% of the positive comments deal with combining theory and practice, and 13.23% about the built up of the course. Other important categories in the positive comments are formative evaluation (11.3%), relevance and interest (11.06%), contribution to learning (10.94%) and a specific component of the course (10.46%).

The most commented upon categories in the negative comments are: relevance and interest of the course which is mentioned in 9.98% of the negative comments, followed by built-up of the course (9,47%), course material miscellany (8,77%), a specific course component (7,82%), and load and difficulty of the course (7,31%). Students positive comments focus to a lesser extent on the context, and the evaluation compared to the negative comments. For example 5.15% of the negative comments deal with the evaluation method while this is only 1.86% of the positive comments. This difference is more or less the same for all evaluation categories, and context categories. For example the quantity of contact moments is mentioned in 2.99% of the negative comments and in 1.66% of the negative comments. Except for the categories that are related to the context and the evaluation, all other aspects are focused upon stronger in the positive comments compared to the negative comments, for example 7.6% of the positive comments are a general evaluation of the teacher (e.g.. good teacher), this aspect is only mentioned in .71% of the negative comments.

Positive comments contain more information. The results show that on average 1.52 codes were used to code a positive comment, and 1.14 for a negative.

Table 2: An overview of the codebook, the amount of codes used, and the percentage of comments that deals with the code and the link to the SET-instrument

Dimension	(Content) category	(Content) subcategory	N pos	Pct. Pos.	Pct. Pos	Pct. neg	Link to the SET-37
Overall evaluation	General	<i>Overall evaluation</i>	100	6.01	10	.64	-
Instructor	General	<i>Overall evaluation</i>	127	7.63	13	.83	-
	Teaching	<i>Use of didactic material</i>	114	6.9	15	.95	-
		<i>Teaching methods</i>	181	10.88	84	5.34	Presentation skills
		<i>Combination theory and practice</i>	291	17.50	108	6.87	Practice-oriented
		<i>Formative evaluation, feedback and support during the learning process</i>	188	11.30	99	6.29	Help of the teacher during the learning process, formative (examinations)
Personal	<i>Time management</i>	56	3.37	84	5.34	-	
	<i>Attitude and conduct towards students</i>	160	9.62	31	1.97	-	
	<i>Professionalism</i>	157	9.44	104	6.61	Multiple dimensions	
Course	Content	<i>Pointing out a course component without further comment</i>	174	10.46	123	7.82	-
		<i>Contribution to learning</i>	182	10.94	46	2.92	-
		<i>Clarity of objectives</i>	75	4.51	41	2.61	Clarity of objectives
		<i>Load, difficulty (level) prior knowledge</i>	304	1.80	115	7.31	Course difficulty, linking-up with advance knowledge
		<i>Relevance, contribution to professional competences, interest, scope, importance</i>	184	11.06	157	9.98	Value of the subject-matter
	Evaluation	<i>build-up</i>	220	13.23	149	9.47	Build-up of the subject-matter
		<i>Evaluation method and authenticity of the evaluation</i>	31	1.86	81	5.15	Authenticity of the evaluation (method is not part of SET-37)
		<i>Content validity</i>	35	2.10	56	3.56	Content validity of the examination(s)
		<i>Difficulty (level)</i>	6	.36	55	3.50	-
		<i>Score</i>	1	.06	21	1.34	-
Course materials	<i>Overall evaluation</i>	47	1.80	28	1.78	-	
	<i>Clarity, build-up and study friendly</i>	82	11.06	74	4.70	(Course materials) contribution to understanding the subject-matter	
	<i>Miscellany</i>	41	13.23	138	8.77	(Course materials) contribution to understanding the subject-matter	
context	Course organization	<i>Quantity of contact moments</i>	27	1.6	474	3.99	-
		<i>Sequence in training program</i>	4	.24	30	1.91	-
		<i>Schedule</i>	1	.06	11	.70	-
TOTAL:			2520		1790		
<i>Average amount of codes per comment</i>			1.51		1.17		

3.3. Do students' comments deal with other aspects, than the dimensions that are measured by fixed-ended questions³?

Some categories could be linked to the dimensions being measured by the fixed-ended questions in the SET-questionnaire, but some differed. The following 13 categories could not be linked: general overall evaluation of the course and teacher, the use of didactic materials, time management, attitude and conduct towards students, a specific course component that students point out, the contribution to learning, difficulty of the evaluation, the evaluation score, the course material general and all context aspects. This mostly has to do with the fact that students' comments, are less or more specific than the fixed-ended questions of the SET-survey.

3.4. Differences in comments between the department of education and others.

To examine whether SET-surveys received in the study field of 'Education and training', have a larger chance to receive certain comments, and to examine a possible consistency between SET-scores and the comments students write down, several logistic regression analyses were calculated. In table 3 the results of these analyses are shown, controlling for gender, year of study and course grades. Only those analyses in which the field of study ('Education and training' or other) or the SET-score have a significant effect are shown. Only the data in which students provided either a negative or positive comment are used for these analyses (n= 1835)

The results in table 3 show that there are significant differences between the study field of 'education and training' and other study fields. SET-surveys that are received in the study field 'education and training', have a significant *larger chance to receive a negative comment* about the difficulty of the evaluation, the clarity of the objectives and the and the sequence of the course in the training program compared to the other fields of study. The odds ratio in table 3 shows that the chance is 5.05 times larger to receive a negative comment about the clarity of objects, compared to the other fields of study and 2.02 times larger to receive a negative comment about the evaluation difficulty compared to other fields of study. The chance to receive a negative comment about the sequence of the course in the training program is 2.66 times as large.

There is a *significant smaller chance* to receive negative comments about the relevance, contribution to professional competences and interestingness (odds-ratio = .63) of the course compared to other fields of study.

The results of the analysis of the positive comments show a significant *smaller chance* to receive *positive comments* about a general evaluation of the teacher (e.g good teacher) (odds ratio = .59), but a significant *larger chance* to receive comments about the combination of theory and practice (odds ratio=1.79), and teaching method (odds ratio 1.76) in the department of 'education and training' compared to other fields of study.

3.5. Consistency between SET-scores and comments?

The fifth research question deals with the possible difference in students' comments when a low, average or high SET-score has been received and whether there is a consistency between this SET-score and students' comments (Low SET-scores meaning a larger chance to receive comments on negative aspects, high SET-scores meaning a large chance for comments on positive aspects). The results in table 3 show a consistency between the SET-score and the comments received. When a SET-survey with a comment has been received, the chance that it deals with following negative categories: teacher professionalism (odds ratio = 1.66), clarity of the objectives (odds ratio = 2.19), the content validity of the evaluation (odds ratio = 1.83), difficulty & adaptation to foreknowledge (odds ratio

³ The following dimensions are measured in the fixed ended questions: clarity of course objectives, value of subject matter, build-up of subject matter, presentation skills, harmony organization course-learning, (course materials) contribution to understanding subject matter, course difficulty, help of the teacher during the learning process, authenticity of the examination, linking-up with advance knowledge, content validity of the examination, formative examination and practice oriented.

= 1.87), and the relevance, contribution to the professional competences and interestingness of the course (odds ratio = 1.60) is significantly larger when the answers to the fixed-ended questions resulted in a low SET-score, compared to a survey with average SET-scores.

The chance that a comment deals with the general *positive* evaluation of the course is *significantly smaller* (odds ratio = .31) when the SET-score from the survey was low compared to average.

Table 3: results of the logistic regression analyses (***) $p < .001$ (***) $p < .01$ (*) $p < .05$

	Difficulty and adaption to foreknowledge (neg)			Attitude/conduct towards students (neg)			Teacher professionalism (neg)			Evaluation content validity (neg)			Relevance, contribution to prof. competences, interest (neg)			Evaluation difficulty (neg)			Clarity of objectives (neg)										
	Est.	odds ratio	sig.	Est.	odds ratio	Sig.	Est.	odds ratio	sig.	Est.	odds ratio	sig.	Est.	odds ratio	sig.	Est.	odds ratio	sig.	Est.	odds ratio	sig.								
Intercept	3,135		***	-3,51		***	-2,78		***	-3,53		***	-2,59		***	-4,1		***	-5,11		***								
Field of study 'education'	-0,14	0,87		-0,63	0,54		0,23	1,26		-0,05	0,96		-0,46	0,63		0,70	2,02	*	1,62	5,05	***								
SET-score low	0,62	1,87	**	-0,57	0,56		0,51	1,66	*	0,60	1,83	*	0,47	1,60	*	0,35	1,42		0,78	2,19	*								
SET-score excellent	-0,05	0,96		-1,47	0,23	*	-0,33	0,72		-0,70	0,50		-0,01	0,99		-0,35	0,70		-1,01	0,36									
course in second year	-0,35	0,70		-0,66	0,52		-0,07	0,93		0,19	1,21		0,25	1,28		0,76	2,13	*	0,65	1,92									
course in third year	-0,20	0,82		-0,36	0,70		-0,15	0,86		0,06	1,06		0,20	1,23		0,03	1,03		1,40	4,06	***								
course grade failed	0,07	1,07		0,58	1,79		0,24	1,27		0,42	1,52		0,52	1,68	*	0,59	1,81	*	-0,91	0,40									
course grade high	-0,55	0,58		0,72	2,05		-0,56	0,57		-0,68	0,50		0,29	1,34		-0,82	0,44		-0,84	0,43									
Gender (male)	-0,35	0,71		-0,15	0,86		-0,58	0,56	*	-0,13	0,88		-0,31	0,73		-0,33	0,72		-0,22	0,80									
Nagelkerke R ²	,03			0,05			0,04			0,04			,03			0,06			0,17										
Sequence in training program (neg)																													
Teacher professionalism (pos)						Teaching method (pos)						General teacher evaluation (pos)						Combining theory and practice (pos)						General course evaluation (pos)					
Est.	odds ratio	sig.	Est.	odds ratio	sig.	Est.	odds ratio	sig.	Est.	odds ratio	sig.	Est.	odds ratio	sig.	Est.	odds ratio	sig.	Est.	odds ratio	sig.									
Intercept	-4,82		***	-2,137		***	-2,38		***	-2,58		***	-1,16		***	-2,85		***	-2,85		***								
Field of study 'education'	0,98	2,66	*	-0,21	0,81		0,56	1,76	***	-0,53	0,59	*	0,58	1,79	***	-0,34	0,72												
SET-score low	-0,13	0,87		-0,15	0,86		0,17	1,19		-0,49	0,61		-0,22	0,80		-1,16	0,31	**											
SET-score excellent	-0,82	0,44		0,38	1,46	*	0,27	1,31		0,16	1,17		-0,20	0,82		0,34	1,41												
course in second year	0,85	2,34		0,10	1,11		-0,08	0,92		-0,01	0,99		-0,16	0,86		0,12	1,12												
course in third year	0,63	1,88		-0,14	0,87		0,03	1,03		0,14	1,14		0,23	1,26	***	0,01	1,01												
course grade failed	0,26	1,30		-0,50	0,61	*	-0,54	0,58	*	-0,33	0,72		-0,77	0,46		-0,15	0,86												
course grade high	-1,50	0,22		-0,22	0,80		-0,03	0,97		-0,05	0,95		-0,41	0,67		-0,01	0,99												
Gender (male)	0,43	1,54		-0,38	0,68		-0,20	0,82		0,47	1,60	*	0,01	1,01		0,40	1,49												
Nagelkerke R ²	,06			0,02			,05			0,03			0,04			,05			,05										

The difference between high and average SET-scores, show that there is a significant *larger chance* that a *positive comment* about teacher professionalism (odds ratio = 1.46) is written in a survey with a high SET-score compared to an average SET-score. For the *negative comments* the results show that SET-surveys with high SET-scores have a significantly smaller chance to receive a negative comment about the conduct and attitude towards students (odds ratio = .23).

4. Conclusion and discussion.

Previous research about students' comments in SET is limited. With this study a contribution was made on five different aspects, to this yet unexplored terrain of research.

First, the results of the current study, showed a rather high number of written comments in SET surveys. In other studies the amount of comments received ranged between 10 -70%, (Alhija & Fresko, 2009; Theall & Franklin, 1991). In the current study in 70% of the SET-surveys received, a comment was written. Although, the difference is small this study confirms the findings of the study by Alhija & Fresko (2009) that more positive comments (62.85%) are received than negative comments (60%).

Secondly, the current study showed that most of the students' positive comments deal with the combination between theory and practice (17.5%). The most commented upon negative category is relevance and interest of the course (11.3%). When the direction of the comments is not taken into account, three most mentioned categories were combining theory and practice, the built-up of the course, and relevance and interest. These are aspects that students probably find important. It indicates that students do not evaluate light-hearted and that their main concern is not shallow.

If the categories most commented upon are compared with the categories in the study by Alhija & Fresko (2009) one important difference is found. In the current study many comments dealt with a good combination between the theory and practice. That so many comments dealt with this topic has probably to do with the fact that the surveys were collected at a university college which is more professional and practical oriented than universities. Moreover this university college also promotes itself as a university college that focuses on the link between theory and practice. When surveys would have been collected at a university, students might find another aspect more important. Universities have a stronger focus on academic skills, and therefore the combination of theory and practice might be less important for these students. This finding also has implications for the use of a generic SET questionnaire. When a generic questionnaire is used, institution specific aspects are not measured. This would mean that in this case the combination between practice and theory would not have been measured, which is an important aspect for both students, and the institution itself.

Some general differences exist between negative and positive comments. Negative comments have a stronger focus on the evaluation and context categories compared to positive comments, while the positive comments focus more on all other aspects (general aspects, the course, and the teacher). Positive comments are more extensive than negative comments, but negative comments are almost always more specific. More codes were needed to code positive comments, than negative comments, but in the negative comments the general categories were almost never coded (e.g. general evaluation course, general evaluation teacher). Student comments have been accused of not being serious (Wachtel, 1998), but findings in the current study show that students seem to take SET seriously. If students provide an answer to the question "*Which aspects did you find negative about this course and should be improved?*" they almost always give a valuable specific improvement aspect (e.g. the pace of the course could be faster for me).

Thirdly, the study showed that some content in the comments of students corresponds to the dimensions being measured with the-SET instrument. However, also some categories could not be directly linked with the dimensions in the SET-instruments. Students use comments to emphasize what they already indicated in the fixed-ended questions but also to provide extra information. In the current study, the categories that were commented upon most by students are covered by the SET-instrument. Linking categories between students' comments and dimensions measured by the SET-instrument showed that comments of students are sometimes more specific, than the dimensions measured in the SET-instrument and sometimes less specific. Students sometimes use a comment to provide a general evaluation of the teacher (e.g. good teacher), while the goal of the fixed-ended questions is to collect a broader range of information. In other comments more specific aspects are mentioned (e.g. the examination was too difficult compared to the things we dealt with in class), while in the fixed-ended questions an evaluation of

the difficulty of the course in general is measured. A mix of open-ended and fixed-ended questions means the best of both worlds.

Fourth, the results showed differences in students' comments between the study field 'education and training' and other study fields. The chance is significantly larger that a comment deals with following negative aspects: the difficulty of the evaluation, the clarity of the objectives, and the sequence of the course in the training program and following positive aspects: combining theory and practice, and the teaching method when the survey is received in the study field 'education and training' compared to other study fields. Students in a certain department might find other things more or less important than students from other study fields. This indicates that the context is an important aspect to consider.

Fifth, the hypothesis that there is a consistency between the SET-score and the comments students write down is confirmed. When a survey with a low SET-score has been received the chances that the comment deals with various negative aspects (e.g. teacher professionalism, content validity of the evaluation, etc.) are larger compared to SET-surveys with an average SET-score. The chance that a survey with a high SET-score is received with a positive comment about teacher professionalism is also larger compared to surveys with average SET-scores.

The first steps were taken in the research towards students' comments in SET, but future research is necessary. For the present study data were collected at a Flemish university college, but replicating this study in other contexts is necessary as indicated students' comments are probably influenced by the context. In other contexts students might find other aspects more or less important. A comparison of students' comments in a university and a university college can yield interesting results.

In sum it is possible to conclude that a student as a commentator seems to take his task seriously and gives specific feedback and consistent feedback about the course and on how to improve it. As to the content of the comments the context plays an important role, and one should take this into account when conducting SET.

5. References

- Ahrens, W., & Pigeot, I. (2005). *Handbook Of Epidemiology*: Springer.
- Alhija, F. N.-A., & Fresko, B. (2009). Student Evaluation of Instruction: What Can Be Learned from Students' Written Comments? *Studies in Educational Evaluation*, 35(1), 37-44.
- Arnold, I. J. M. (2009). Do examinations influence student evaluations? *International Journal of Educational Research*, 48(4), 215-224.
- Bendixen, L. D., & Rule, D. C. (Writer) (2004). An Integrative Approach to Personal Epistemology: A Guiding Model [Article], *Educational Psychologist*: Taylor & Francis Ltd.
- Brockx, B., Spooren, P., & Mortelmans, D. (2011). Taking the grading leniency story to the edge. The influence of student, teacher, and course characteristics on student evaluations of teaching in higher education. *Educational Assessment, Evaluation and Accountability*, 23(4), 289-306.
- Brownlee, J., Purdie, N., & Boulton-Lewis, G. (2001). Changing Epistemological Beliefs in Pre-service Teacher Education Students. *Teaching in Higher Education*, 6(2), 247-268.
- Carbera, A. F. (1994). Logistic regression analysis in Higher Education: An applied perspective. In J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research* (Vol. 10, pp. 225 - 256). New York: Agathon Press.
- Centra, J. A. (1993). *Reflective faculty evaluation : enhancing teaching and determining faculty effectiveness* (1st ed.). San Francisco: Jossey-Bass.
- DeMaris, A. (2004). *Regression With Social Data: Modeling Continuous and Limited Response Variables*: Wiley-Interscience.
- Foddy, W. H. (1994). *Constructing Questions for Interviews and Questionnaires: Theory and Practice in Social Research*: Cambridge University Press.
- Hammond, I., Taylor, J., & McMenamin, P. (2003). Value of a structured participant evaluation questionnaire in the development of a surgical education program. *Australian and New Zealand Journal of Obstetrics and Gynaecology*, 43(2), 115-118.
- Hardy, N. (2003). Online Ratings: Fact and Fiction. *New Directions for Teaching and Learning*.

- Hodges, L. C., & Stanton, K. (2007). Translating Comments on Student Evaluations into the Language of Learning. *Innovative Higher Education*, 31(5), 279-286.
- Hox, J. (2010). *Multilevel Analysis: Techniques and Applications*. New York: Routledge.
- Johnson, B., & Christensen, L. B. (2012). *Educational Research: Quantitative, Qualitative, and Mixed Approaches*: Sage Publications.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174.
- Larose, D. T. (2006). *Data Mining Methods And Models*: Wiley-Interscience.
- Lewis, K. G. (2001). Making Sense of Student Written Comments. *New Directions for Teaching and Learning*.
- Menard, S. W. (2009). *Logistic Regression: From Introductory to Advanced Concepts and Applications*: Sage.
- Mortelmans, D., & Spooren, P. (2009). *Onderwijsevaluatie door studenten aan de KHM. Eindrapport van de valideringstudie naar de bruikbaarheid van het UA-evaluatie-instrument voor opleidingsonderdelen. Report of a validity study, to the use of the UA evaluationinstrument for courses.*
- Mortelmans, D., & Spooren, P. (2009). A revalidation of the SET37 questionnaire for student evaluations of teaching. *Educational Studies*, 35(5), 547 - 552.
- Oliver, B., Tucker, B., & Pegden, J. (2007). *An investigation into student comment behaviours: Who comments, what do they say, and do anonymous students behave badly?* Paper presented at the Australian Quality Forum. from <http://www.auqa.edu.au/auqf/pastfora/2007/program/papers/c11.pdf>
- Richardson, J. T. E. (2005). Instruments for Obtaining Student Feedback: A Review of the Literature. *Assessment and Evaluation in Higher Education*, 30(4), 387-415.
- Roche, L. A., & Marsh, H. W. (2000). Multiple dimensions of university teacher self-concept - Construct validation and the influence of students' evaluations of teaching. *Instructional Science*, 28(5-6), 439-468.
- Santner, T. J., & Duffy, D. E. (1989). *The Statistical Analysis of Discrete Data*: Springer-Verlag.
- Schraw, G. (2001). Current Themes and Future Directions in Epistemological Research: A Commentary. *Educational Psychology Review*, 13(4), 451-464.
- Smith, K., & Welicker-Pollak, M. (2008). What can they say about my teaching? Teacher educators' attitudes to standardised student evaluation of teaching. *European Journal of Teacher Education*, 31(2), 203-214.
- Spooren, P. (2010). On the credibility of the judge: A cross-classified multilevel analysis on students' evaluation of teaching. *Studies in Educational Evaluation*, 36, 121-131.
- Spooren, P., Mortelmans, D., & Denekens, J. (2007). Student Evaluation of Teaching Quality in Higher Education: Development of an Instrument Based on 10 Likert-Scales. *Assessment & Evaluation in Higher Education*, 32(6), 667-679.
- Strauss, A. L., & Corbin, J. M. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*: Sage Publications.
- Svinicki, M. D. (2001). Encouraging Your Students to Give Feedback. *New Directions for Teaching and Learning*, 2001(87), 17-24.
- Symmons (2006). *Listening to the student voice at the University of Sydney: Closing the loop in the quality enhancement and improvement cycle.* Paper presented at the Australian Association for Institutional Research (AAIR) Forum.
- Theall, M., & Franklin, J. (1991). Using student ratings for teaching improvement. 1991, 83-96.