Student Evaluation of Teaching (SET): Clues on how to interpret written feedback

Guillaume Schiltz[a]

[a]Department of Physics, ETH Zurich (Swiss Federal Institute of Technology), Switzerland

Abstract

In this paper we present the results of a study covering 218 written comments submitted in the formal university SET questionnaire of two undergraduate physics lectures for engineering students. Concerning the SET-metrics, one of the lectures was rated as critical, while the other lecture had good results. The analysis is based on the praise and criticism framework elaborated by Hyland/Hyland (2001) for written feedback. Our findings, which also relate written feedback to quantitative variables and contrast the results between critical and good evaluations, provide a deeper insight for both, teachers and educational developers, on how to interpret written comments in a quality management process.

Keywords: *student evaluations of teaching; students' comments; content analysis; quality management; student voice; formative feedback.*

1. Introduction

The feedback from student evaluation of teaching (SET) is a major instrument to measure the degree of faculty achievements. SET is carried out with rating forms, online or paper, that students have to fill in at the end of the term. Typical items of the questionnaire include Likert-scale questions on the teacher performance, the provided material and the class organization (Marsh, 2007). Based on a defined metrics, those items are statistically compiled to measure the teacher's teaching effectiveness. The results of SET are often used to decide on tenure or promotion of the teacher (Kember et al., 2002). Another feature of SET, however, consists of providing feedback to the teacher (Yao&Grady, 2005). To support this formative purpose of the SET, the questionnaire often includes free-text questions, where students can comment on their personal experience in more detail.

Whereas the analysis of SET data mainly focuses on quantitative ratings, little is known about the impact of written feedback from free-text questions. Among the few studies dealing with written comments, Alhija&Fresko (2009) and Brockx et al. (2012) offer some valuable insight into quantitative aspects of free-text comments. Moreover, open-ended comments have been subject to linguistic analysis (Stewart, 2015) and were used for exploratory considerations (Hodges&Stanton, 2006; Stupans et al., 2016).

In this study we combine quantitative results together with lexical evidence in order to provide some interpretative hints, on how to link written comments to the overall questionnaire results.

2. Data and Coding

The data cover 218 written comments submitted in our formal university SET questionnaires. They result from two independent undergraduate physics lectures (table 1). According to the SET-metrics (based on Likert-scale questions) defined by the university, lecture A is regarded as a good lecture, whereas lecture B was identified as critical. Written comments are open-ended answers induced at the end of the questionnaire by the item *"Imagine that you are the lecturer teaching this course unit. What would you improve? What would you keep unchanged?"*. Their sole purpose is to provide feedback to the teacher and at ETH Zurich they are not relevant for the SET-metrics.

We based our coding scheme (figure 1) on the "Praise and Criticism" feedback points introduced by Hyland&Hyland (2001). Each feedback point is related to one of the 7 predefined content categories and identified either as critique, praise or suggestion. The categories have been selected according to the main themes of the preceding Likert-scale questions of the questionnaire. With an average of 55 words the comments turn out to be

rather extensive and almost all comments include a set of different feedback points (table 2).

	lecture A	lecture B	total
evaluation records	273	326	599
records with comments	88 (32%)	130 (40%)	218 (36%)
average comment word count	40	65	55
feedback points	158	265	423
overt criticism	23	72	95
overt praise	70	72	142
suggestions	65	121	186

Table 1. Student evaluations and written comments included in the study. Lectures A and B are independent physics introductory courses for undergraduate engineers.

Content Categories	
C	Content: i.e. syllabus, level of complexity, curriculum
L	Lecturing: i.e. teacher performance, slides, experiments
M	Material: i.e. script, notes, master solutions
A	Activities: i.e. clicker, exercises
S	Support: individual support, online and in class
O	Organization (of the course, of exercise classes)
I	Alignment: exercices – lecture – exam

Feedback Points		
teacher	**t+**	overt praise, e.g. *"The teacher was available for …"* → Cat. **S**
	t–	overt criticism, e.g. *"The teacher copies his script …"* → Cat. **L**
activity	**a+**	overt praise, e.g. *"Exercises helped to understand …"* → Cat. **I**
	a–	overt criticism, e.g. *"The script was erroneous …"* → Cat. **M**
suggestion	**s+**	supporting, e.g. *"Keep the clicker questions …"* → Cat. **A**
	s–	Improving, e.g. *"Less quantum mechanics …"* → Cat. **C**

Mitigation	
pairing	explicit combination of praise and criticism, e.g. *"… however …"*
hedges	verbally softening feedback, i.e. *"sometimes …"*
details	providing examples and/or explanations, e.g. *"… because …"*

Figure 1. Coding scheme. A single comment mostly includes several feedback points.

In addition we distinguished praise and criticism to whether it addresses the teacher as a person or the activity of teaching. In order to specify the degree of politeness, we also codified mitigation strategies. Pairing occurs when praise and criticism are used in combination. Hedges refer to the lexical mitigation of any feedback and details were recorded when concrete examples or further details are mentioned to underpin the feedback.

Most of the comments are written in German, but English is used as well. Furthermore, the comments range from single keywords to complex sentences, while using emoticons and special characters. Relying on automated or semi-automated analysis tools, as used in other studies (Stupans et al., 2016; Zaitseva et al., 2013), turned out to be inapplicable. All comments were hand-coded and double-checked for reliability.

comment	t+	t-	a+	a-	s+	s-	mitig.
Many errors in the script! Clicker questions are helpful and make the somehow dry lecture less monotonous.			A	M L			pairing hedges

Table 2. Example coding of a comment (3 feedback points).

3. Results and discussion

3.1 Critical lectures tend to entail more written feedback.

A total of 423 feedback points could be identified, 158 for lecture A and 265 for lecture B. This results in an average of 1.8 feedback points per comment for lecture A and an average of 2.0 feedback points for lecture B. The overall feedback for the critical lecture B, thus, was significantly more extensive.

3.2 Overt criticism is less frequent in good lectures.

In lecture A, overt praise (t+, a+) (n=70) occurred three times as often as overt criticism (t-, a-) (n=23). In contrast, for lecture B the occurrence of overt praise and overt criticism was identical (n=72).

3.3 Written Feedback addressing the teacher as a person is primarily positive.

Overt praise referencing the teacher (t+) could be identified in both lectures (A: n=31, B: n=26). Overt criticism (t-) only occurred 4 times in lecture B. Those instances, however, were heavily mitigated. Otherwise they would have classified for offensive comments. Stewart (2015) showed evidence that the praising tendency is generally directed to the teacher's person (t+) and criticism to the product of the teacher's actions (a-). E.g. "The teacher was highly motivated" vs. "The lecture was boring". We could only support these findings for the category L.

3.4 For critical lectures, students tend to provide more negative than positive feedback.

In lecture A the occurrence of positive feedback points (t+, a+, s+) (n=82) and of negative feedback points (t-, a-, s-) (n=76) was nearly identical. In lecture B, however, we identified almost twice as much negative feedback points (n=175) than positive ones (n=90).

3.5 Negative feedback is forthright and mostly lacks lexical mitigation.

Lexical mitigation (pairing, hedges) was used extremely sparse. Argumentative mitigation (details) occurred in both lectures with a clear peak in lecture B (table 3). Some of this additional information was helpful for further improvements of the lecture.

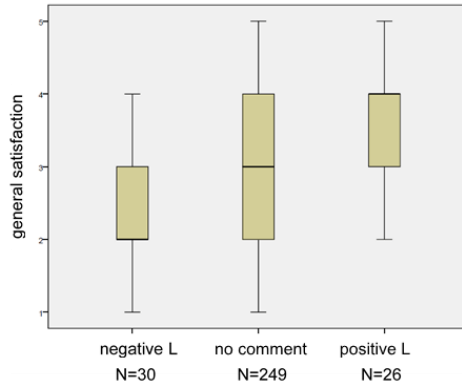|	lecture A	lecture B
pairing	17 (19%)	21 (16%)
hedges	7 (8%)	12 (9%)
details	42 (48%)	84 (65%)

Table 3. Distribution of mitigation strategies

3.6 Written feedback can be biased, i.e. positive and negative feedback does not echo the general results of the evaluation.

Are comments biased? We conducted a one-way ANOVA to assess feedback for category L on the values of the SET-variable "general satisfaction" (table 4). Evaluation records were divided into one of the three subpopulations: "positive L", "negative L", "no comment". Only for the critical lecture B, the SET-value differed statistically significant for the three subpopulations (figure 2). The SET-value decreased from "positive L" (3.6) to "no comment" (3.0) and further to "negative L" (2.3). For lecture A, however, the SET mean values of all subpopulations remained nearly unchanged.

lecture	SET	t+	t–	a+	a–	s+	s–	$\Sigma_{(+)}$	$\Sigma_{(-)}$	$\Sigma_{(tot)}$
A	3,7	30	0	14	5	7	15	51	20	142
B	2,9	17	4	18	10	9	34	44	48	184

SET is the mean value of the general satisfaction expressed by a 5-scale Likert question (1="very unsatisfactory", 5="very satisfactory"). It is the key variable in the SET-metrics for quality management.

Table 4. Distribution of feedback points related to the category L (lecturing) in both lectures.

21 out of 326 records were omitted because they had pos. L
together with neg. L, or they were missing a SET value.

Figure 2. Distribution of the SET-value for lecture B according to the three subpopulations (SPSS Boxplots).

4. Conclusion

Often lecturers feel confused and disappointed when reading students' comments, especially negative ones (Hodges&Stanton, 2007). With our study we offer a framework to interpret comments in the broader context of the evaluation results. Comparing comments from critical and good evaluations turned out be extremely helpful. The fact that even good evaluations show a considerable number of critical feedback points was surprising. Identifying a possible bias in critical evaluations was another revealing finding. Even though the same problem is addressed in several independent comments, this does not a priori point to a major deficiency. Further data will be needed to support our results and we are planning to pursue the study with additional evaluation data sets.

Acknowledgments

References

Alhija, F. N.-A., & B. Fresko (2009). Student evaluation of instruction: What can be learned from students' written comments? *Studies in Educational Evaluation*, 35, 37–44.

Brockx, B., Van Roy, K., & D. Mortelmans (2012). The student as a commentator: Students' comments in student evaluations of teaching. *Procedia-Social and Behavioral Sciences*, 69, 1122–1133.

Hodges, L. C., & K. Stanton (2007). Translating comments on student evaluations into the language of learning. *Innovative Higher Education*, 31, 279–286.

Hyland, F., & K. Hyland (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing*, 10.3, 185–212.

Kember, D., Leung, D. Y. P. & K. P. Kwan (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment & Evaluation in Higher Education*, 27, (5), 411-425.

Marsh, H. W. (2007). Students' evaluations of university teaching: A multidimensional perspective. In R. P. Perry & J C. Smart (Ed.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective* (pp.319-384). New York: Springer.

Stewart, M. (2015). The language of praise and criticism in a student evaluation survey. *Studies in Educational Evaluation*, 45, 1–9.

Stupans, I., McGuren, T., & A. M. Babey (2016). Student Evaluation of Teaching: A Study Exploring Student Rating Instrument Free-form Text Comments. *Innovative Higher Education*, 41, 33–42.

Yao, Y., & M. Grady (2005). How do faculty make formative use of student evaluation feedback?: A multiple case study. *Journal of Personnel Evaluation in Education*, 18, 107–126.

Zaitseva, E., Milsom, C. V., & M. Stewart (2013). Connecting the dots: Using concept maps for interpreting student satisfaction. *Quality in Higher Education*, 19(2), 225–247.