



When to Ask Student Evaluation Questionnaire: before the Exam, after the Exam or at both Occasions?

Vasja Vehovar and Miha Matjašič,
*Centre for Social Informatics, Faculty of Social Science,
University of Ljubljana*



Student evaluation of teaching (SET)

SET is subject of most intriguing social science research, with numerous methodological controversies. For example, it is still not clear if SET truly measures the effectiveness of the teacher or it measures friendliness and likeliness.

We address here in this study a specific aspect related to the timing of SET, i.e. **whether to do it before or after exam.**

The dilemma is of high administration importance and it also raises serious methodological questions.

Surprisingly, only a very limited and non-conclusive research can be found on this issue, from Frey (1976) to Guangming Li et al. (2018) and Zipser&Minciell (2018).



SET Timing

Three strategies for timing of the SET survey:

1. Before exam
2. After exam
3. Before and after exam



Before Exam

Advantages:

- No influence of the exam/grades on SET;
- Similar timing for all students (e.g. last week of course);
- Effective administration tools exist to force responding: e.g. application to exam is not allowed until the student evaluates the course, or, evaluation is done in the classroom at the end of the course via mobile phones.

Disadvantage:

- Weak evaluation of **efforts** (hours) spent given the number of ECTS, **competences**, and particularly quality and fairness of the **exam**.



After Exam

Advantages:

- Competent evaluation of exam, effort and competences.

Disadvantages:

- Impact of grade lead students to unreliable response, while teachers are stimulated for high grading.
- Much less convenient and effective administration procedures are available to force students into SET, so danger of seriously low response rate exist.
- Timing may have a long tail, with differential memory effects, increased variety of uncontrolled circumstances, pressure before enrolment into next study year (which is a typical measure of forcing in the after-exam strategy).



Before Exam AND After Exam

Advantages:

- Enjoys benefits of both, before and after, strategies;
- Avoids disadvantages of both strategies.

Disadvantage:

- Increased burden for students: the number of items (questions) remains the same, but the number of survey sessions actually doubles (two instead of one);



SET at University of Ljubljana (UL)

- After decades of discussions and reluctances, the online SET was introduced in 2015/2016 for all 26 faculties.
- Questions are asked **before** (7 items for the course, 6 on each teacher of the course, plus related open questions) and 5 items **after** exam (3 items on exam, plus evaluation of efforts and competences).
- Contrary to expectations, system works relatively well, despite its complexity and large diversity of the faculties, which are relatively independent, with non-centralised administration.



SET administration at UL

- A questionnaire **before** exam is offered to students at 70% of course (Dec 15 for 1st semester and May 10 for 2nd semester). The before questionnaire disappears after student applies to exam, but students cannot apply to ANY exam before handling (i.e. responding or refusing) evaluations for ALL due courses of that semester.
- A questionnaire **after** opens when results are recorded. Student typically respond (or refuse) within few days.
- Once activated, questionnaires remain open (with permanent passive alert in administrative profile) until answered or refused. Without resolving all evaluations the enrolment into next study year in Fall is not possible.

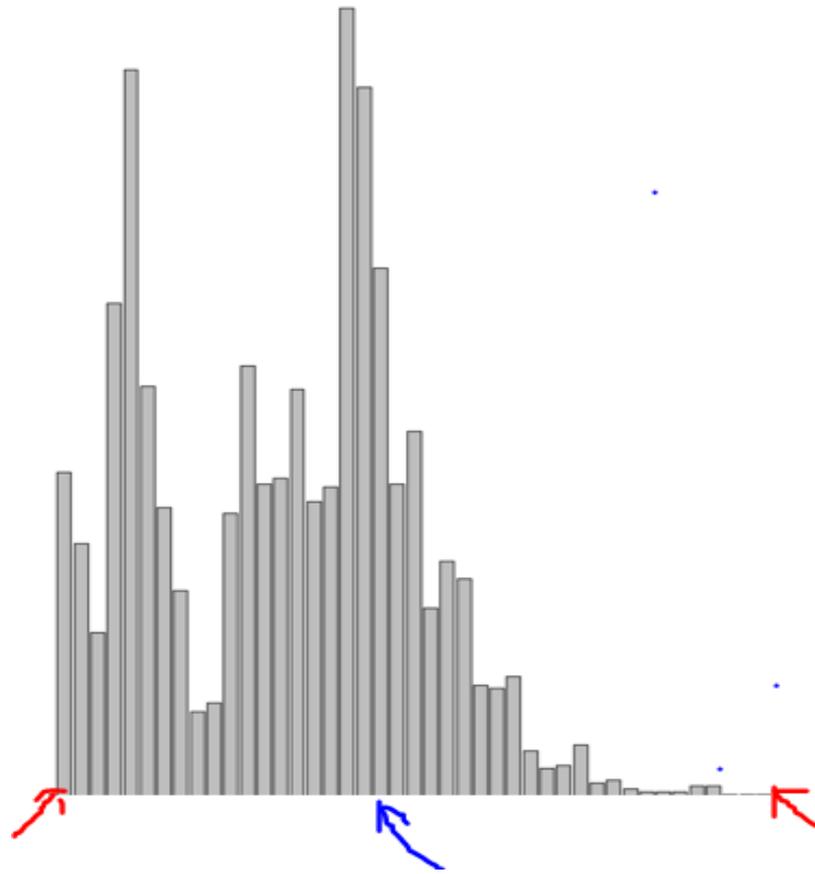


Experiment 2018/2019

- At the Faculty of Social Sciences, each semester students are evaluating around 6 courses, before and after exam.
- In experiment, the **SEVEN** items, related to course (satisfaction, consistency, autonomy, literature, informed, online content, grading) from **BEFORE** exam, were repeated also **AFTER** exam, in addition to **FIVE** after exam items (**content**, **clarity** and **grading** of exam, **competences** and *efforts*). The first **FOUR** items of these FIVE items **AFTER** exam can be treated as outcomes.
- The *effort* item from AFTER exam set was also repeated BEFORE, but was found unrelated with any other item.



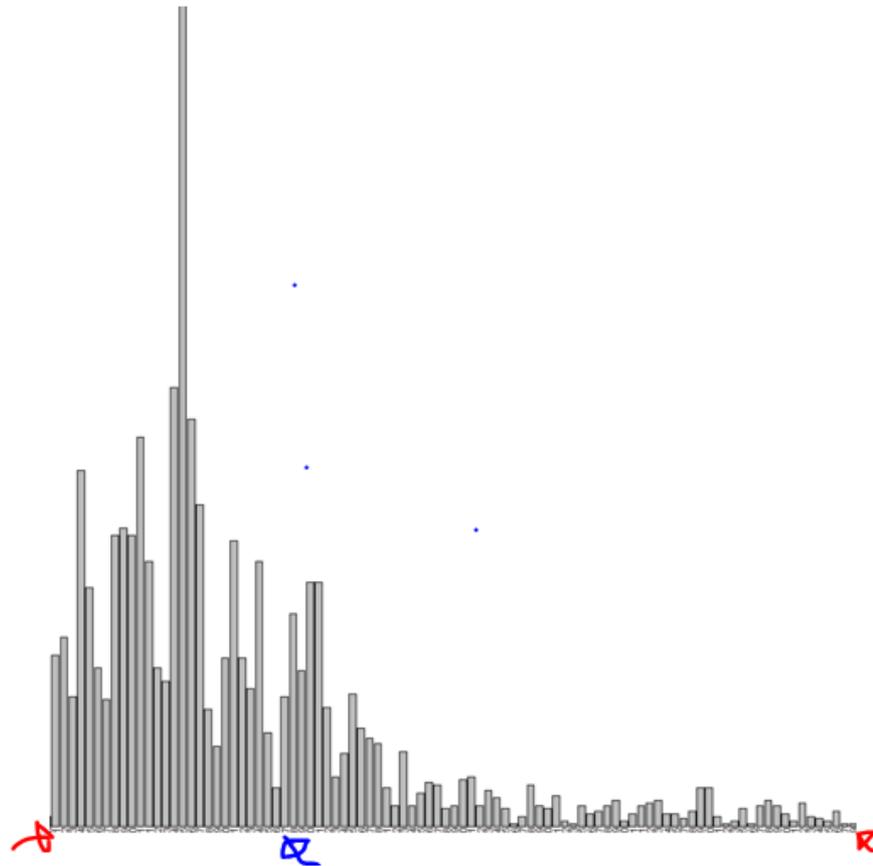
Distribution of the BEFORE surveys, 2st semester (red arrows: May 10th - Jun 25st)



Blue arrow: end of 2nd semester - May 29 (start Feb 17)



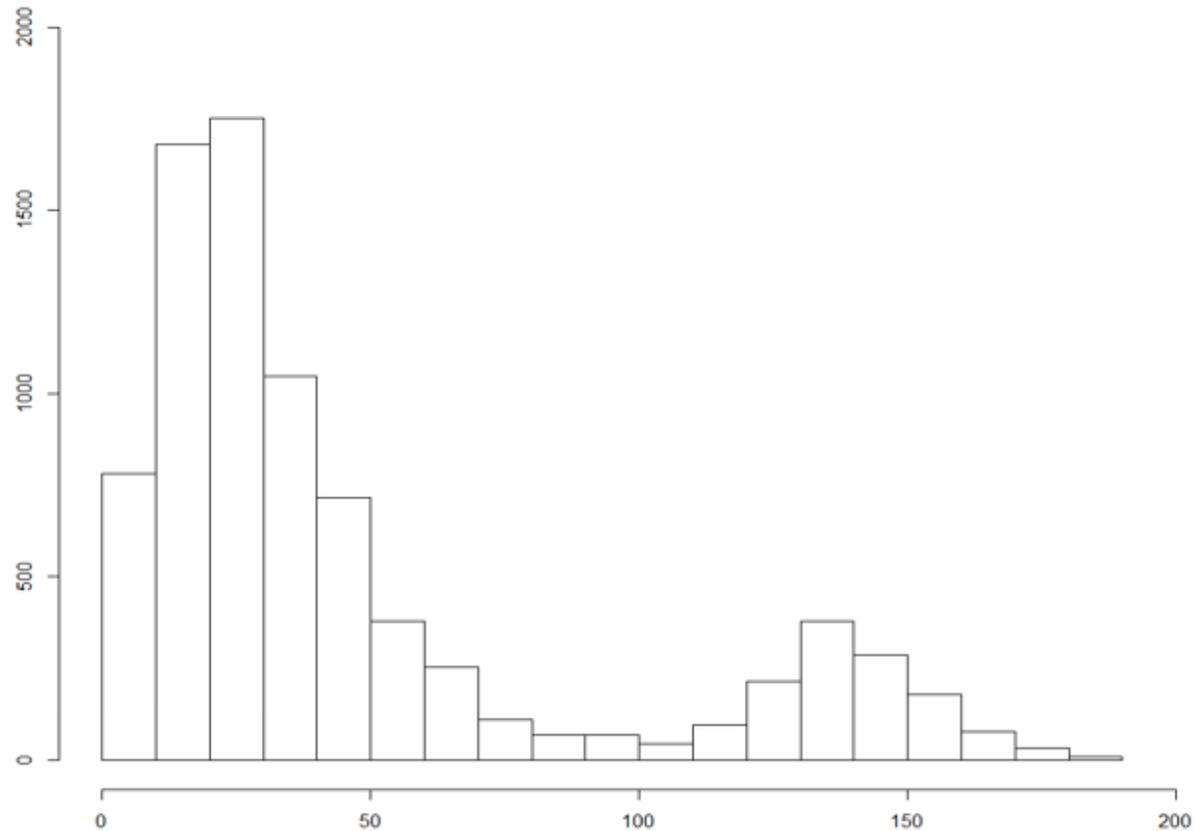
Distribution of the AFTER surveys, 1st semester (red arrows: Jan 20 and Apr 30)



Blue arrow: start of 2nd semester - Feb 17 (ending May 29)



Difference BEFORE – AFTER (in days) for the corresponding course



Two peaks: 30 days and 130 days.



Correlations of items in SEVEN

- Cronbach alpha for SEVEN items in **BEFORE** and also in **AFTER** was almost the same, slightly above 0.90.
- Between-item paired correlations of SEVEN items in BEFORE are 0.5-0.7 and they are similar also in AFTER.
- **BEFORE-AFTER** correlations (i.e. test-retest reliability) for each of the SEVEN items are around **0.5**, with „general satisfaction“ being the highest (0.58).
- However, the paired item correlations of **SEVEN** items with **FOUR** outcome variables (measured only in AFTER) are around **0.4** for SEVEN items are asked **BEFORE**, but **0.6** for the same SEVEN items when asked AFTER.
- No correlation of any of the item was found with time of the survey or with the BEFORE-AFTER time difference.



Impact of the exam

- The difference in the means of SEVEN items (BEFORE-AFTER) is positive (i.e. student are happier after exam) and statistically significant for all seven items, but it is very small, within the interval 0.01-0.10 on 1-5 scale.
- The average difference (increase) across all SEVEN items from BEFORE to AFTER is $d=0.06$, i.e. from **4.17** to **4.23**.
- Regression of this difference (d) on FOUR outcome variables from AFTER gives $R^2=0.12$. The item „satisfaction with grading“ from the FOUR outcome variables has the largest impact on d . It creates the increase (regression coefficient) of $\text{Beta}=0.16$ ($t=10.0$).



Data quality

No difference in data quality is detected:

- **Share of straight-liners** (units with same answers to all SEVEN items) is the same (35%) BEFORE and AFTER.
- **Refusals** (unit nonresponse): BEFORE 31%, AFTER 32%.
- **Item non-response rate** is same and very small (1%).

However, for **27% of course evaluations done BEFORE, the students were not exposed to AFTER survey**. Some reasons:

- students abandon study before being forced to answer,
- faculty administrative deviations allow students are not forced to resolve (answer or refuse) after-exam survey,
- violations of administrative rules exist by teachers.



Conclusions

- The level of satisfaction with the course is almost the same, whether items are asked BEFORE or AFTER exam; the average increase in the means is 4.17 \rightarrow 4.23.
- The FOUR outcome items explain some ($R^2=0.12$) of this BEFORE-AFTER difference ($d=4.23-4.17=0.06$).
- No difference in response quality for BEFORE or AFTER.
- Test-retest (BEFORE-AFTER) item reliability is only around 0.5 for each of the SEVEN items.
- The SEVEN course evaluation items have much higher correlation with the FOUR outcome variables when asked AFTER exam (0.6) compared to BEFORE exam (0.4),
- Overall participation in the AFTER exam survey is much lower compared to BEFORE exam survey.



Test-retest (BEFORE-AFTER exam) results for item SATISFACTION ($r=0.58$)

		Satisfaction AFTER exam					
Likert scale		1	2	3	4	5	Total
Satisfaction BEFORE exam	1	47	37	27	19	18	148
	2	37	102	101	86	36	362
	3	29	69	221	250	79	648
	4	15	71	206	884	529	1705
	5	11	28	72	404	1390	1905
	Total	139	307	627	1643	2052	4768

Almost a quarter (23%) of evaluations change from positive to non-positive, or, from negative to non-negative.

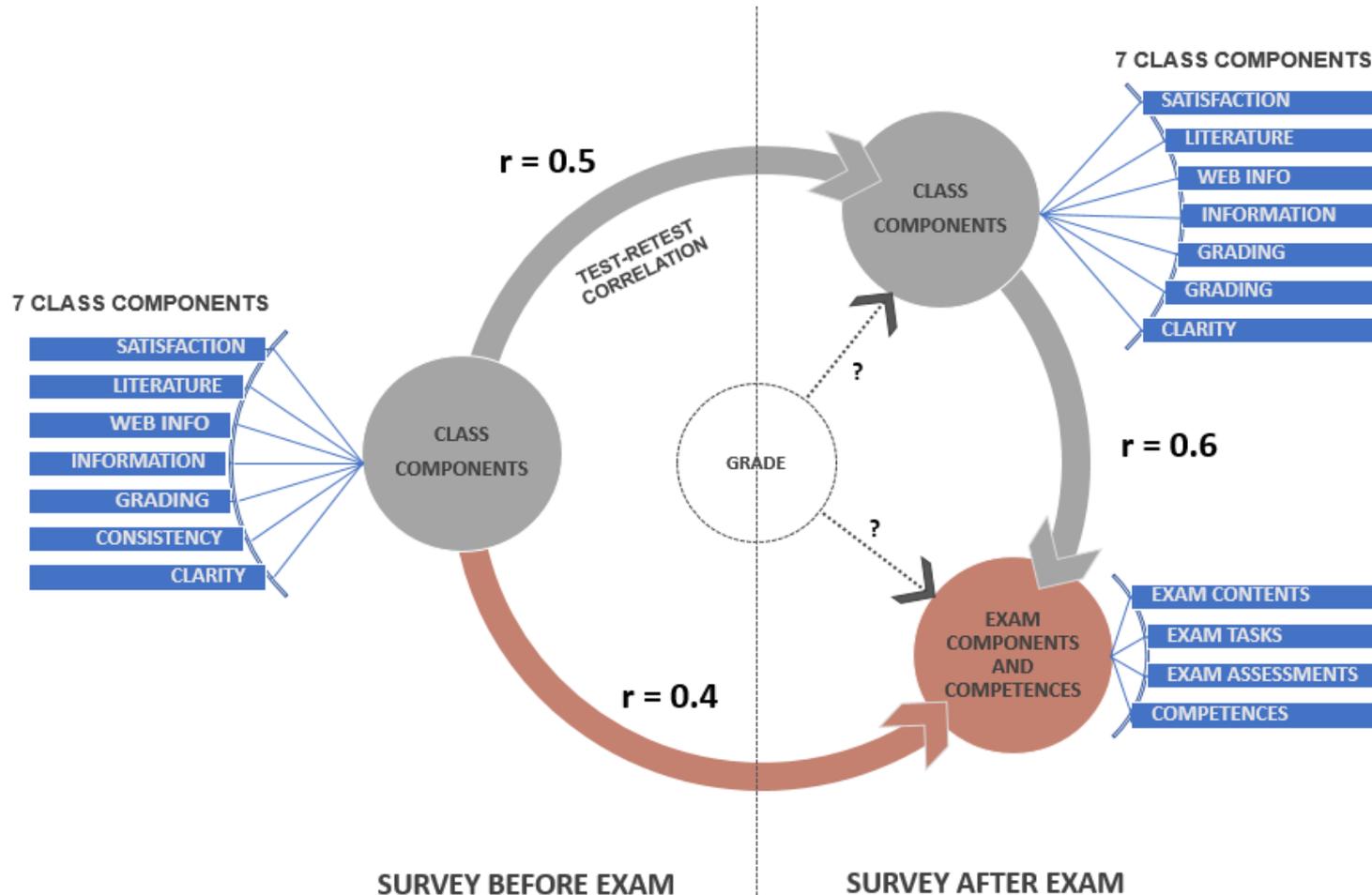


Discussion on reliability

- The item by item test-retest correlation for BEFORE items is relatively low, it is only around $r=0.5$ for each items.
- In total, 23% of individual evaluations (out of 4678) changed from positive to non-positive, or, from negative to non-negative in AFTER survey compared to BEFORE. This is considerable change. Only 56% evaluations remained the same in AFTER compared to BEFORE.
- Is this a normal „noise“ of reliability, or, there is a systematic effect due to exam and due to memory loss (arising in 30 or 130 days time lag between BEFORE exam and AFTER exam surveys), so that BEFORE evaluations are closer to true attitudes compared to AFTER ?



Correlations between blocks of components for course evaluations





Discussion on response rate

The main problem with SET at University of Ljubljana are administrative issues: how to provide student cooperation.

Namely, without some forcing strategy, the response rates are below 10%, while **extreme forcing** (e.g. evaluation as a precondition to enrolment into next year, with no option to refuse) would severely damage the response quality.

The warning example from a neighbouring University of Maribor (Slovenia), which uses only the AFTER-exam survey (with extreme forcing), shows that vast majority of students do straight-line in their evaluations.



Discussion on increase in satisfaction

- When SEVEN items on course evaluation are asked AFTER exam, the average across all item significantly increases compared to BEFORE, for 0.06 on scale 1-5.
- Given that vast majority of evaluations are in interval 4.0-4.5, difference 0.06 is not that small; as its relative value is above 5% of this interval.
- However, this is only the net increase for units that answered both surveys (BEFORE and AFTER), while gross increase is higher, because units that answered only BEFORE have higher means than those answered only AFTER. In addition, the difference among means at the course level are higher than individual based estimates.



Final discussion on three strategies

- **Before-only strategy:** Problems exist with evaluation of exam, competences, effort. How reliable it is to ask all this BEFORE exam? *This is an issue for future experiments.*
- **After-only:** Suffers from much lower response rates than before-only, due to lack of tools for forcing students. Plus, estimates for BEFORE items when asked AFTER exam are unreliable (due to impact of exam and time lag).
- **After AND before:** There is evidence (e.g. inquiry among students) that extra effort **is not burdensome** at all (particularly because survey can be refused with one click), which means this strategy has no disadvantages, only benefits. But there are also some anecdotal evidence that **it is burdensome**. *So this is an issue of further investigation.*