Optimise your course evaluation system

Ad Scheepers identifies good practices in the use of Student Evaluations of Teaching (SET)

Student evaluations of teaching (SET) are widely used in higher education to assess course and teaching quality. SETs serve as an information source for teachers to be able to improve their teaching, provide information for students to select courses and for pedagogical research.

In the majority of cases, SETs also serve as an information source for management to assess the quality and effectiveness of teaching and teachers, decide salary and promotions or awards and help institutions to conduct programme reviews. SETs have an important place in the educational quality control cycle.

The use of SETs often leads to controversy and debate among faculty and management. The meaningfulness, methods used, validity, reliability, bias and (mis-)interpretation of student ratings are and have been extensively discussed and researched, as far back as the 1920s but especially in the last few decades.

There are lots of questions. Are SETs used in the right way? Are SET methods applied in the right way? Are SETs results used in a correct way? Are SETs reliable enough to make assessments about teaching quality? Are SETs the only way to evaluate teaching, teachers and course assessment? Can students judge quality in teaching?

In an attempt to create some clarity and structure and find useful recommendations and examples of good practices, we conducted both a field study and a literature review. The field study extracted information (interviews, websites, and questionnaire) from 50 institutions (mostly business schools) in Europa, Americas, Oceania and Asia. The literature review considered over 50 studies and meta-analyses over the last two decades.

"

The use of SETs often leads to controversy and debate among faculty and management

Evaluation practice

The field study showed that all teaching programmes in the sample use SET ratings in one way or another. Most schools use SET data for human resources (HR) purposes. The field study indicates that SETs are primarily used to gather information on teacher performance, course elements and assessment. SETs contain in almost all cases a combination of teaching aspects: instructor effectiveness, course quality, course content (including skills development related to the course content), organisational aspects and assessment quality. Student involvement is sometimes included.

Next to these aspects, overall opinions are very often included. Single-score ratings (overall rating of course and teacher) do occur in more than a quarter of the cases.

SETs are typically administered following direct teaching in the classroom, after the last lecture or on completion of the course but before exam results are in. The field research did not reveal specific SETs for distance learning, e-learning, or technology-enhanced learning. The use of online evaluation is widespread although in some instances paper-and-pencil administration still occur.

Schools and institutions seem satisfied with SET systems when they see the system is taken seriously and when it is comprehensive; but more often they are unsatisfied when the SET system is perceived as incomplete, has low student participation and is not closing the loop.

The most important stakeholders in the course and teaching quality assessment are teaching staff, management, support staff and students. However, from the field study as well as from some studies, it appears that roles and tasks are often diffuse and not made explicit. Moreover, there is a wide variety of departments and sections involved in some stages in the SET process. Often, tasks and roles are not marked off and actors in one stage of the process are not aware what are the roles and tasks of actors in other stages.

Figure 1 SET process and process steps as a PDCA quality cycle.

Fragmentation and tension

SET research results do not create a consistent picture and are fragmented. Characteristic in the research literature are the varying perspectives, aspects and subjects. The field of SET research is not only fragmented but also contains contradictory results (for example, on the relation between student performance and SET ratings).

The fragmentation and contradictions in the results make it hard to find clear answers in the SET research literature. Some quotes may illustrate this (emphases added):

• "Research on SET has thus far *failed to provide clear answers* to several critical questions concerning the validity of SET." (Spooren, Brockx & Mortelmans, 2013, p. 598.)

• "However, despite researchers producing dozens of studies validating several aspects of SETs ... the overall result has been inconclusive on many important issues …. "Part of the difficulty in referencing SET literature is the fact that one can *often find at least one study to support almost any conclusion.*" (Griffin, Hilton III, Plummer & Barret., 2014, p. 339.)

A central theme is the tension between the use of SETs to improve teaching effectiveness on the one hand and to qualify teaching faculty on the other. Reliability, possible misinterpretation, misuse and bias are important issues in this dilemma. From the literature review, several alternative methods are mentioned (such as observation tools).

The field study shows that schools and institutions do use alternative sources: prior requirements; interview results from stakeholders; peer reviews; teacher self-evaluation. Respondents in the sample do not indicate using observational methods.
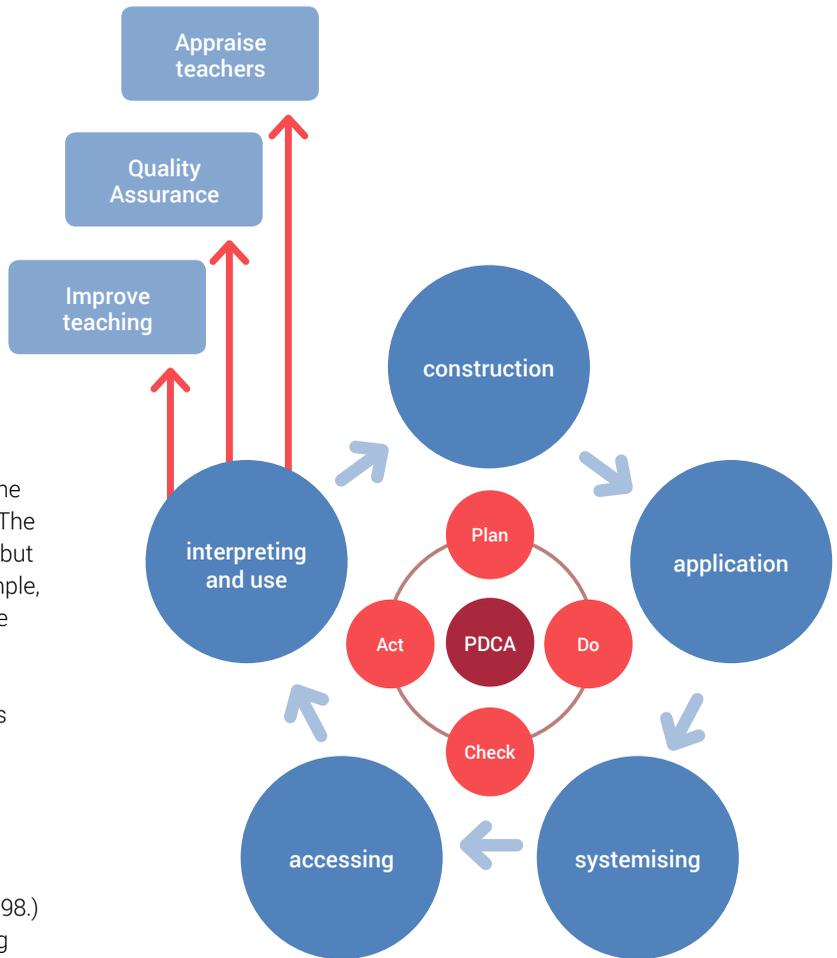
Quality control cycle

The literature review shows that SET-related analyses and conclusions are fragmented. Research mostly does not have a process perspective but focuses on isolated phenomena. They tend to be problem oriented instead of being solution oriented. In addition, in the majority of cases a clear overview and guarding of the whole process seems to be lacking. This may result in poor use of data, and misinterpretation of data (or worse, no relevant set of data, making the SET procedure symbolic and a waste of effort).

To be able to structure the results from the literature review and the field study, in our report we have looked at SET as a process with several stages. In this way the SET process forms a teaching quality control cycle (PDCA) (see Figure 1).

SETs

The literature review shows that SET-related analyses and conclusions are fragmented

In terms of framework, it appears that most research is on one of SET construction, application or interpretation and use. Surprisingly, systemising and accessing SET data have hardly been subject to academic research. However, these aspects of institutional organisation of SET are crucial in conducting the evaluation process efficiently and effectively.

Not properly organising and systemising SET data may negatively affect the other process stages. For example, if SET data cannot be accessed in a complete, transparent and systematic way it may prevent a proper check on reliability and validity of the data. This, in turn, may lead to missing the necessity to adjust evaluation questions and questionnaires. Equally important, it may lead to misinterpretations.

Risks of unreliability, bias, misinterpretations and inappropriate use can be especially high if the required expertise, accountability and supervision in the separate stages as well as for the process as a whole are not properly laid down.

Guidelines

The field study shows a representative overview and a great variety of practices but it is hard to select best practices. However, despite the fragmentation and contradictions, we were able to formulate seven (general) guidelines, taking together all of the findings in the field study and the results and recommendations of the literature review. Following these guidelines might counter the most important problems that are found in the practice of evaluating and help optimising the SET process.

First, make sure your SET measurement are valid. Make sure to use measuring scales that suit the purpose of your SET. Depending on whether SET is used for assessing teacher quality, course improvement or quality assurance, choose the appropriate dimensions. Make use of validated instruments and dimensions with a theoretical basis. Standardise the SET instrument(s) across courses and time so as to optimise comparability and consistency. Regardless of the purpose of SET, be aware that course quality and teaching quality are multidimensional. Avoid using one-dimensional, overall indicators only.

Second, check the reliability of your SET assessment regularly and systematically. Make sure to optimise both the reliability of the instrument (questionnaire or scale reliability) as well as of the respondents. For an adequate interpretation of SET results, respondent reliability is essential because it indicates the level of agreement of respondents about course quality and teaching quality.

Third, to use SETs for course improvements as well as an HR instrument, consider the possibility of separating assessing teacher quality from course quality improvement opportunities. Use different sources for assessing teaching effectiveness.

It is clear from SET research and field experience that using only SET for judging teacher quality is limited.

Fourth, consider the optimisation of response rates. Low response rates do not necessarily have to be problematic. A low response rate may be reliable if you make sure your sample is sufficiently representative so that the respondent reliability is high. Several studies have indicated ways to improve this reliability and required response rates for given class sizes to assure reliability.

Alternatively, to increase response rates one might consider making SET mandatory as an integral part of a course or module. Guaranteeing the reliability of SETs, either by optimising representativeness of the sample or by making SETs mandatory, will increase the trust of faculty and they will rely on results more.

Fifth, improving the response quality can be achieved through either measures beforehand or measures afterwards.

Measures beforehand are: give academics strategies for interpreting reports; organise support and appropriate mentoring by peers and head of school; educate students to give professional and constructive feedback; forfeit anonymity in cases of abusive comments.

"

Following these guidelines might counter the most important problems that are found in the practice of evaluating and help optimising the SET process

It is clear from SET research and field experience that using only SET for judging teacher quality is limited

Measures that can be taken afterwards (after questionnaires are filled in) are: removing swear words (software) prior to the release of reports to the relevant co-ordinator; removing comments following the release of the report (only after scrutiny by a committee).

Sixth, address bias control. Make sure that the goals of SET are clearly stated, SET ratings are multidimensional, measuring scales are valid, consistency reliability is high, respondent reliability is high and the response sample is representative.

When interpretation is in accordance with the SET goals, characteristics of respondents are taken into consideration when interpreting results. Additional measures are: making sure interpreters of quantitative results know how to interpret the statistics; installing a uniform grading policy to prevent leniency; using alternative, multiple forms of performance (such as peer review, observation); being aware of the limitations in expertise of students to recognise good teaching; and making sure there is a (gender, ethnicity, age) balance in the teaching teams.

Finally, and very important, it is highly recommended to consider SET as a process with discernible and interconnected steps. The quality of SET will improve greatly if SET is seen as quality process with linked, coherent stages to create a closed quality cycle (PDCA).

When using computer software it is advisable to use a professional application that supports the integrated SET process. Be sure that the system encompasses and integrates all steps in the SET process. In that way, it will be a valuable addition to the SET process and the quality assurance system as a whole.

Considering SET as a quality cycle process also makes it necessary to overlook and guard the process as a whole, to make guidance and steering possible. It is advisable to create a position for a functionary department or section with an overview of the whole process, who guards the process and is responsible and accountable.

To be able to follow up on these guidelines, the report gives practical suggestions and recommendations in each of the stages of the process.

gf

About the Author

Ad W A Scheepers is Policy Advisor and Project Manager at the Rotterdam School of Management, Netherlands

For the full report referred to in this article see SET Project: Student Evaluations of Teaching, Measuring and Enhancing Course Quality and Teaching Quality (2019). https://equal.network/projects/