Lessons Learned 3, 2 (2023)
Submitted: 27.09.2023
Accepted: 18.12.2023
DOI: https://doi.org/10.25369/ll.v3i2.81
ISSN: 2749-1293 (Print); 2749-1307 (Online)

How good is my teaching (really)? Using student feedback constructively

B. Schlegel[1*], B. Watzka[2]

[1] *Chair of Magnetofluid Dynamics, Measurement and Automation Technology, Institute of Mechatronic Mechanical Engineering, Faculty of Mechanical Engineering, TU Dresden*
[2] *Jun. Prof.'in for Didactics of Physics, Institute of Physics, Faculty of Natural Sciences, Otto von Guericke University Magdeburg*

Abstract

Student feedback enables a differentiated view of the (perceived) usefulness of teaching and learning offers, methods and tools with regard to students' learning progress. In addition to didactic principles and teaching methodology, student feedback can be used to tailor teaching to students' needs. Evaluation for the further development of teaching and evaluation research are available as tools alongside other types of student feedback. Both types of evaluation are not used to analyze and control the quality of teaching, but can be used as instruments by teachers who want to further develop their teaching. Before a corresponding survey can be planned and carried out, the following questions must be discussed in detail: What specific question is to be answered with the help of the data to be collected? Which influencing factors play a role in the context of the question? What results are conceivable and what consequences would these results have for the design of teaching? Are these consequences relevant? The answers to these questions can be used to determine which form of evaluation is chosen and how it can be specifically designed.

Studentisches Feedback ermöglicht einen differenzierten Blick auf die (wahrgenommene) Nützlichkeit von Lehr-Lern-Angeboten, Methoden und Tools bezüglich des Lernfortschrittes bei den Studierenden. Neben fachdidaktischen Prinzipien und Lehr-Methodik lässt sich auf Basis studentischen Feedbacks Lehre bedarfsgerecht gestalten. Evaluation zur Weiterentwicklung der Lehre und Evaluationsforschung stehen neben anderen Arten studentischer Rückmeldung als Werkzeuge zur Verfügung. Beide Varianten von Evaluation dienen nicht der Analyse und Kontrolle der Qualität von Lehre, sondern können als Instrumente von Lehrenden eingesetzt werden, die ihre Lehre weiterentwickeln wollen. Bevor eine entsprechende Erhebung geplant und durchgeführt werden kann, müssen folgende Fragen ausführlich diskutiert werden: Welche konkrete Fragestellung soll mit Hilfe der zu erhebenden Daten beantwortet werden? Welche Einflussfaktoren spielen im Kontext der Fragestellung eine Rolle? Welche Ergebnisse sind denkbar und welche Konsequenzen hätten diese Ergebnisse für die Gestaltung von Lehre? Sind diese Konsequenzen relevant? Aus den Antworten auf diese Fragen lässt sich ableiten, welche Form der Evaluation gewählt wird und wie diese konkret gestaltet werden kann.

*Corresponding author: beatrice.schlegel@tu-dresden.de
This article was originally submitted in German.

1. Introduction

The innovations in teaching - driven or intensified by the changed conditions during the pandemic - have been reflected upon, discussed, adapted or even rejected in many ways in recent semesters. There is now a consensus that digital elements enrich our teaching, and that face-to-face courses (abbreviated to LV in the following) are part of good study conditions. Which elements are used in which form and are really suitable for advancing the students' learning process cannot be answered in general terms and is always subject-specific or even module-specific. In order to gain a differentiated insight into this, various types of student feedback can be obtained (classic feedback, tests, evaluation, etc.) In combination with the expertise of the teaching staff, they can lead to a decisive improvement in teaching.

This article is about evaluation and evaluation research as a means of improving teaching. The field of research is outlined, starting with pure evaluation through to empirical research, which requires more expertise in comparison. In section two, the functions of different evaluation approaches are explained by way of example, and then in the two subsequent sections, evaluation for the further development of teaching and evaluation research are discussed in more detail. Specific reference is made here to the workshop of the same name as part of the Lessons Learned Conference 2023. Our vision is for lecturers from different subject areas to jointly develop evaluation modules and use them in their own teaching in order to improve teaching in the long term and, if necessary, to further develop evaluation processes.

2. Functions of evaluation

Course evaluation, such as that carried out at TU Dresden by the Center for Quality Analysis (ZQA), is primarily used to monitor the quality of teaching at a university or higher education institution in all subject areas. The evaluation used by the ZQA is based on the Heidelberg Inventory of Course Quality [1]. This type of evaluation makes it possible to record the current status and compare it with a quality standard that reflects the expectations of teaching. The evaluation form [2] contains statements about the teacher, which are rated by the students on a five-point scale from "strongly agree" to "strongly disagree". Examples of statements are that the teacher "...conveys the course content clearly" or "...is available for consultation if required." In addition, assessments of the course such as "The lecture has expanded my knowledge" as well as assessments of the requirements of the course, workload, student commitment and others are asked. The results of this evaluation give the teacher an impression of the students' views and initial indications of where improvements can be made. However, they cannot be interpreted as to which improvements should be made and in what way. This is not the aim of this type of evaluation.

The type of evaluation presented in this article aims to find indications for the concrete improvement of teaching. Conversely, they allow few conclusions to be drawn about the quality of teaching.

An initial comparative example should illustrate this difference. In the ZQA course evaluation, students are asked under the heading "Use of digital teaching formats" how they rate the digital formats used in the course (page 3 [2]). The formats to be evaluated are livestream, video recordings, PowerPoint presentations, discussion forums and other formats. In the evaluation of the further development of teaching, students are asked about their preferred variant for the course with the option of selecting one of the following answers:

- Lecture only in presence,
- videos only,
- Live broadcast via YouTube only,
- Hybrid variant I: Lecture in presence, online participation,
- Hybrid variant II: Lecture in presence, videos,
- Hybrid variant III: Lecture in presence, online participation, videos,
- Other (with the option to enter something).

Statements on the quality of digital formats can be derived from the results of the ZQA survey. The results of the evaluation on the further development of teaching provide an indication of which format is more suitable for students.

A major limitation of evaluations is the ability to establish causal relationships. Although evaluations can identify correlative relationships, they do not allow reliable statements to be made about cause and effect. For example, an evaluation could show that students who prefer to follow lectures in video format rate their knowledge acquisition as higher. This suggests a connection between the lecture format "video" and the subjective perception of knowledge acquisition, but the actual cause of this phenomenon remains unclear. Specific research designs with control groups are required to determine reliable cause-and-effect relationships.

It should be noted at this point that the development of teaching can never take place purely on the basis of evaluation. The expertise of the teacher is of central importance. On the one hand, it is based on their in-depth understanding of the subject and their research experience, and on the other hand on their didactic and pedagogical know-how. Findings from (subject) didactics and pedagogy provide important pointers for the reorientation and development of teaching. For example, according to Deci and Ryan's self-determination theory, motivation depends on the extent to which the three basic psychological needs of experiencing competence, autonomy and social integration are met [3]. Consequently, in order to motivate students, teaching-learning situations must be created that meet these needs. However, students have different types of needs. This can depend on the subject and also vary according to the content to be taught. In order to find out to what extent the design of the teaching-learning situations actually meets the students' needs, an evaluation should be carried out.

3. Course evaluation for the further development of teaching

Two examples are used below to illustrate the possibilities and limitations of evaluation for the further development of teaching.[1]

Possibilities and limits

Example 1 - the intermediate query

An intermediate question in the course can help to repeat and consolidate knowledge. Possible time periods for interim questions are the beginning, middle or end of the course. Which variant is preferred by the students can be determined by a simple query as part of an evaluation, as in the evaluation of the Measurement and Automation Technology module (abbreviated to MAT in the following). The result for the winter semester 2022 can be seen in Figure 1.

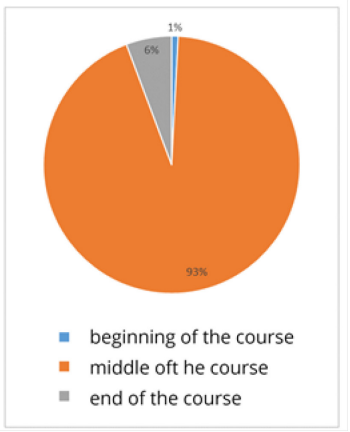

Figure 1: Evaluation example - Interim survey - Question: Which variant do you prefer? n=122

93% of the students surveyed stated that they preferred the mid-lecture question.

What can be deduced from the results obtained?

From this, a design tip can be derived, namely to include the intermediate questions in the middle of the next courses of the module. It is not possible to deduce why the students prefer this variant. Possibly because the process

[1] The student survey takes place online during the course.

of listening and taking notes is interrupted and this "break" has a positive effect on concentration. This is only an assumption and is not confirmed or refuted by the evaluation. The results also do not indicate at what point in the course intermediate questions should generally be included or what function they have in the learning process.

Example 2 - Understanding the students

The aim of teaching is (among other things) for students to understand technical contexts. It is the task of the teacher to support the process of understanding and to create teaching-learning opportunities in which insights are possible. In order to improve the teaching offer, it can be useful to find out in which teaching-learning situation students understand the most. In the evaluation of the MAT module, students were asked to what extent they agreed with the following five statements on a four-point scale from one "disagree" to four "strongly agree":

(1) I was able to follow the teacher's explanations without any problems.
(2) I was able to understand everything from what was written down.
(3) It was only when I worked through my transcript that I was able to understand the connections.
(4) I only really understood the connections through the exercises.
(5) It was only after attending the exercises that I really understood the connections.

The results of the survey are shown in Figure 2.

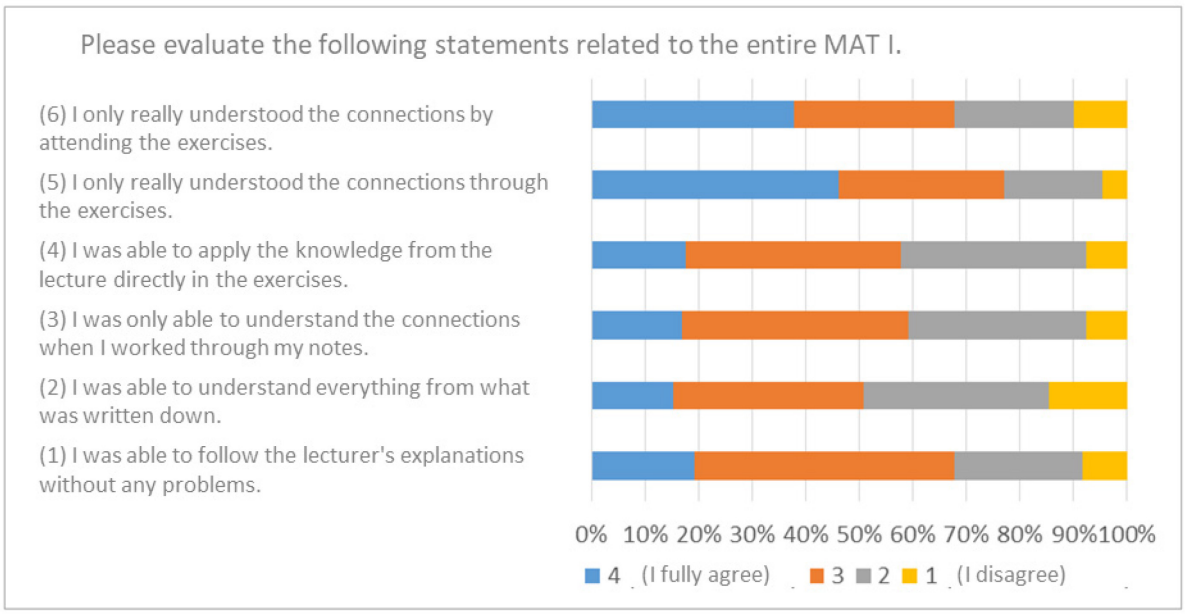

Figure 2: Extract from the results of the evaluation of the MAT module in the winter semester 2023

The first three statements on the teacher's explanations and the notes were rated as "strongly agree" by less than 20%. In comparison, twice as many fully agreed with the statements on the exercises and exercises (statements (4) and (5)). If the results of ratings four (strongly agree) and three are combined, the differences are partially balanced out. For example, the two statements on the explanations (1) and on attending the exercise (5) both achieved almost 70% agreement. The statement "I was able to understand everything from what was written down" received the lowest level of agreement (16% fully agree, four and three together: 51%). The statement "I only really understood the relationships through the exercises" received the highest level of agreement (46% fully agree, four and three in total: 77%).

Students who disagreed or strongly disagreed

(one or two) with the statement "I was able to follow the teacher's explanations without any problems" were also asked explicitly why they were not able to follow the explanations very well or at all. The following answers were given, among others:

- Can't keep up with the writing,
- Because I can only write or follow the lecture,
- too fast,
- Too little time was scheduled in the lecture for the amount of material.

What can be deduced from the results obtained?

The students surveyed rated the exercises as more effective for understanding contexts than the lecture. What was written down supported the understanding process significantly less than the other teaching-learning offers. It could therefore be helpful to include more exercises and to integrate practice examples into the course in order to illustrate and apply what has been written down.

Apparently, the amount of material is too large for some students to be able to write and think at the same time in the course. It could be helpful to offer some of the content to be taught in a script so that not all content has to be copied out. It should also be considered whether the amount of material can be reduced.

Limitations: Causal relationships between the design of teaching-learning situations and the support of student learning processes are not examined in this evaluation. Insights gained from the in-depth inquiry of sub-groups are only valid for the student group. They cannot be generalized. Causal deductions and the generalization of findings are only possible in evaluation research, which is the subject of Chapter 4.

Development of an evaluation module

The last section showed examples of the potential that evaluation offers for the development of teaching and its limitations. Evaluation results provide points of reference for decisions and show where changes can be made. More global questions lead to statements that relate to more than one teaching-learning situation. More specific questions provide information on individual aspects. In order to ask precisely tailored questions that lead to usable results, a corresponding development process is required, which is characterized by discussion of the following questions:

- Which teaching/learning situation, method or content is involved? Global or specific?
- What does the teacher want to know from the students?
- What evaluation results are conceivable?
- What are the consequences of each conceivable outcome?
- Are these consequences relevant / interesting and can they be implemented?

Once these questions have been clarified, the following steps lead to an evaluation module:

Step 1: Search for corresponding modules in existing inventories,
Step 2: Adaptation of modules or new construction,
Step 3: Check whether questions can actually be answered with the module,
Step 4: Test and adjust if necessary.

As part of a workshop at the Lessons Learned Conference 2023, an evaluation module was developed with a group of four lecturers to be used directly in their teaching in future. The aim is to evaluate, interpret and discuss the results of the evaluation together. This type of collaboration enables evaluation and teaching development to take place more constructively, as different groups of students and different perspectives of the teachers can be compared. The questions discussed in the team and the evaluation module developed are presented below. **Teachers are invited to incorporate the module into their own evaluation.**

The team has decided to focus on a method for consolidation and with a high level of student activity that can take up to ten minutes in a course, such as an interim question. The question to be answered with the help of the evaluation is: How is the method accepted by the

students? Does it support the process of understanding? It is therefore a concrete question. Possible results are (a) the method is well accepted, or (b) the method is not well accepted, or (c) the method is partially well accepted. The following consequences were derived from the individual results: for the result (a) the method is used more frequently, for (b) the reason for this is investigated and adjusted if necessary, for (c) the method is used in suitable courses but not more frequently. These consequences were assessed by the teachers as relevant and feasible.

Figure 3 shows the evaluation module in the Limesurvey survey tool, which is provided by the Bildungsportal Sachsen for TU Dresden [4].

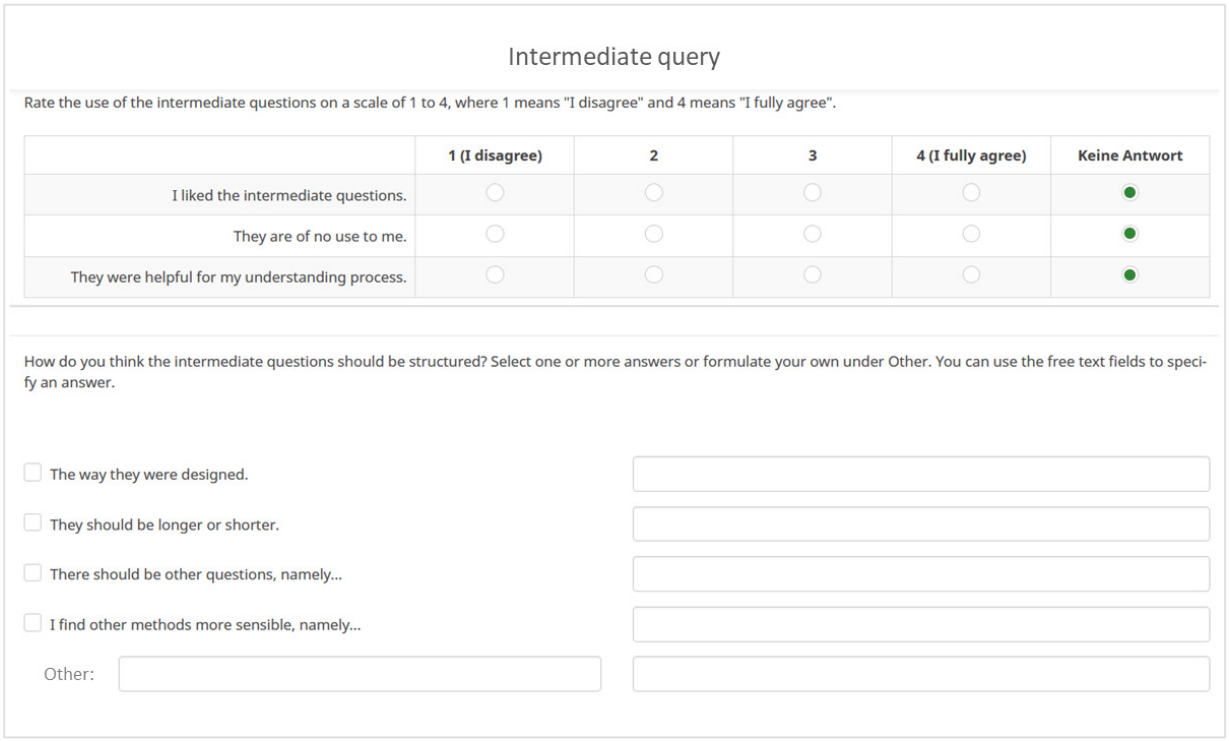

Figure 3: Evaluation module intermediate query

The module can be part of a course evaluation as part of quality control or also part of a more comprehensive evaluation for the further development of teaching. It is conceivable to use the module as a short evaluation in combination with general information such as subject of specialization, semester, possibly also gender and origin, if corresponding comparisons between student groups are interesting and helpful.

4. Evaluation research

Evaluation research uses both quantitative and qualitative research methods. It can pursue formative objectives to improve ongoing processes or summative objectives to make final assessments. However, evaluation research goes beyond simple feedback. It searches in controlled environments for the causes and mechanisms behind the observed results.

Summative evaluation in particular focuses on two main objectives. On the one hand, it is possible to examine how a treatment condition (independent variable: e.g. the teaching format) influences an observation/measurement (dependent variable: e.g. learning performance). On the other hand, summative evaluations are about carrying out impact analyses or making predictions. Here, for example, questions could arise such as: "How strongly does an increase in weekly quizzes by one task influence learning performance?" or "What can be predicted about learning performance based on the number of quizzes?

The royal road of research

In experiments, specific conditions are controlled and variables are manipulated in order to investigate causal relationships. To ensure that other unexpected variables are not responsible for observed effects, a pre-post control group design with randomized assignment of participants is often used. In a pre-post control group design, the dependent variable is measured before the intervention (pre-test) and after the intervention (post-test), with an additional comparison with a control group that does not receive an intervention. The participants are randomly assigned to the group with or without the intervention. This design prevents not only the independent variable 'teaching format', but also other factors such as prior knowledge, from influencing the dependent variable 'learning performance'.

The ideal research process begins with questions and hypotheses derived from theory. These then determine the research design. Let us assume that we want to investigate the influence of two independent variables - 'teaching format' (with the levels 'flipped classroom' and 'traditional') and 'prior knowledge' (with the levels 'high' and 'low') - on a dependent variable such as 'learning performance'. In this case, a 2x2 design would be suitable, which can evaluate potential interactions between the variables in particular.

		teaching format	
		traditional	Flipped classroom
prior knowledge	low	ab Var. Lernleistung*	ab Var. Lernleistung*
	high	ab Var. Lernleistung*	ab Var. Lernleistung*

** Standardized questionnaire*

Figure 4: Classic 2x2 factorial design

Once the research design has been defined, the next step is to operationalize and select suitable, ideally standardized, measurement instruments such as questionnaires. Recommended sources for standardized measurement instruments are platforms such as www.testarchiv.eu in general and www.physport.org specifically for physical content.

The detailed test plan is then drawn up. This specifies when exactly which steps (e.g. pre-test) are to be carried out using which methods (e.g. questionnaire) for which groups of test subjects (e.g. a group experiencing the classic teaching format with little prior knowledge) and over what period of time (e.g. 45 minutes for a pre-test). Simpler experimental designs, as are often used in evaluations, are limited to the teaching material (treatment) and a subsequent test (post-test). An example of an application would be the investigation of differences in learning performance following the use of a particular teaching format, as shown in Figure 5.

Figure 5: Simple test plan with post-test

A major problem with assessing learning performance using such simple experimental designs is that learning performance could be influenced by other factors, such as prior knowledge. Therefore, it will not be possible to conclude from the results that a specific teaching format leads directly to a specific learning performance.

To address this problem, many research projects rely on pre-night test designs, as shown in Figure 6. However, even these designs do not necessarily allow for a clear causal relationship between the independent variable or treatment (here the teaching format) and the changes observed between the measurements (e.g. learning gains). There could still be an uncontrolled third variable, such as cognitive ability, that explains the difference between the two measurement points.

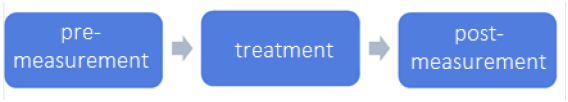

Figure 6: Simple test plan with pre- and post-test.

In order to exclude uncontrolled third variables, pre-post designs with control groups in conjunction with randomization are helpful, as shown in Figure 7. In this design, the control group does not receive any treatment, but is tested at both measurement times with regard to the dependent variable. However, even this design is not without limitations, especially when it comes to drawing causal conclusions. One potential stumbling block could be that the pre-measurement influences the subsequent treatment. This could happen, for example, if the pre-measurement makes it clear to participants which aspects of the treatment are considered particularly relevant.

Figure 7: Experimental design with pre-test, post-test and control group.

The Solomon 4 group enplan [5] offers a solution. It not only takes into account the main factors, but also controls the possible effects of the measurements themselves, as shown in Figure 8.

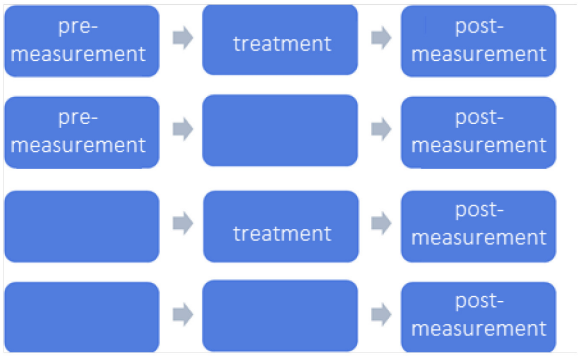

Figure 8: Solomon 4 group plan with double control group (treatment and measurement)

Problems in the implementation of research projects

For many university courses, students register themselves based on their interests, their requirements or their course of study. It would not be practical to randomly assign them to a specific course or teacher.

Randomizing students could be perceived as unfair, especially if one course or teaching method is seen as superior or more desirable. Students may feel disadvantaged if they are randomly assigned to a less favored course (especially if it is the control group).

The implementation of randomization in university courses would require considerable organizational effort. Systems would have to be set up to ensure that the allocation of students is correct and truly random.

Given all of these considerations, conducting randomization in university courses can be so costly and complex that it outweighs the potential benefits, especially when the main goal of the evaluation is to gather feedback to improve teaching and not necessarily to establish causal relationships.

A way out of the dilemma

A simple way to address the problems mentioned is to carry out impact analyses using regression or path analyses. Even if these approaches do not show causalities in the strict sense according to , they make it possible to explain differences in the dependent variables through the influences of independent, confounding and third variables. It is therefore advantageous to have a detailed understanding of potential confounding and third-party variables in addition to the dependent variable and to record these as well. For example, differences in cognitive nitive abilities [6], in men tal cognitive nitive stress [7] or in the men specific prior knowledge could influence the variance of the dependent variable. These factors could also contribute to the differences caused by the independent variable .

5. Conclusion and outlook

The discussion in the last two chapters shows how different the functions and questions of evaluation (research) can be. This results in different approaches to the development and design of the instruments. These differences determine the interpretation of the results and the consequences for teachers. Taking this into account and defining it clearly is the basis of a successful evaluation. In the workshop "How good is my teaching (really)?" as part of

the Lessons Learned Conference 2023, teachers discussed evaluation (research), its different functions and how to develop and design it. In a collaborative process, an evaluation module was created that can be used by teachers (description in section section 3). In addition, different research approaches were discussed with regard to their practical bility. The common conclusion was to test fewer difference hypotheses, as these are generally associated with complex and complex designs. Instead, the focus was shifted to impact analyses in order to better explain variations in a dependent variable .

The Lessons Learned Conference 2024 will then focus on the joint evaluation, assessment and discussion of the results. Our vision of joint evaluation development in the Lessons Learned Community will thus be further advanced.

Literature

[1] https://tu-dresden.de/zqa/qualitaetsanalyse/lehrveranstaltungsevaluation
[2] Sample questionnaire lecture of the Center for Quality Analysis of the TU Dresden. https://tu-dresden.de/zqa/ressourcen/dateien/qa/lve/LVE_Vorlesung_virtuell.pdf?lang=de
[3] Deci, Edward L.; Ryan, Richard M. The self-determination theory of motivation and its significance for pedagogy Zeitschrift für Pädagogik 39 (1993) 2, pp. 223-238
[4] https://bildungsportal.sachsen.de/umfragen/
[5] Döring, Nicola; Bortz, Jürgen. Research Methods and Evaluation in the Social and Human Sciences (2016). Berlin, Heidelberg: Springer.
[6] Heller, Kurt A.; Perleth, Christoph. Cognitive ability test for 4th through 12th grades, revision (2000). Göttingen: Hogrefe.
[7] Leppink, J., Paas, F., Van der Vleuten, C.P.M. et al. Development of an instrument for measuring different types of cognitive load. Behav Res 45, 1058-1072 (2013).