Discrepancies in purposes of student course evaluations: what does it mean to be "satisfied"?

Iris Borch[1] · Ragnhild Sandvoll[1] · Torsten Risør[2,3]

Abstract

Student evaluation of teaching is a multipurpose tool that aims to improve and assure educational quality. Improved teaching and student learning are central to educational enhancement. However, use of evaluation data for these purposes is less robust than expected. This paper explores how students and teachers perceive how different student evaluation methods at a Norwegian university invite students to provide feedback about aspects relevant to their learning processes. We discuss whether there are characteristics of the methods themselves that might affect the use of student evaluation. For the purpose of this study, interviews with teachers and students were conducted, and educational documents were analysed. Results indicated that evaluation questions in surveys emerged as mostly teaching-oriented, non-specific and satisfaction-based. This type of question did not request feedback from students about aspects that they considered relevant to their learning processes. Teachers noted limitations with surveys and said such questions were unsuitable for educational enhancement. In contrast, dialogue-based evaluation methods engaged students in discussions about their learning processes and increased students' and teachers' awareness about how aspects of courses improved and hindered students' learning processes. Students regarded these dialogues as valuable for their learning processes and development of communication skills. The students expected all evaluations to be learning oriented and were surprised by the teaching focus in surveys. This discrepancy caused a gap between students' expectations and the evaluation practice. Dialogue-based evaluation methods stand out as a promising alternative or supplement to a written student evaluation approach when focusing on students' learning processes.

Keywords Evaluation methods · Evaluation use · Quality enhancement · Higher education · Student evaluation of teaching · Utilisation-focused evaluation

✉ Iris Borch
iris.h.borch@uit.no

Extended author information available on the last page of the article

1 Introduction

In the last two decades, the use of evaluation has been proliferated in European higher education concurrent with an increase in educational evaluations and auditing by quality assurance agencies (European University Association 2007; Hansen 2009; Stensaker and Leiber 2015). The overall goal for evaluators and the *raison d'être* of educational evaluation are to improve teaching and student learning (Ryan and Cousins 2009, pp. IX–X).

In evaluation research, the use of the evaluation data is one of the most investigated topics (Christie 2007; Johnson et al. 2009). It is evident that, despite the high number of collected evaluations, use of evaluation data remains low (Patton 2008). Inspired by existing research and with an intention to increase use of evaluation for the intended purpose, Michael Quinn Patton developed the utilisation-focused evaluation (UFE) approach (Patton 1997, 2008). Essential to UFE is the premise "that evaluations should be judged by their utility and actual use" and that "the focus in UFE is on intended use by intended users" (Patton 2008, p. 37). Utility to UFE is strongly related to intended use and shall therefore be related to the purposes of evaluation.

Inspired by utilisation-focused evaluation, this paper investigates the intended use of student evaluation for the overall purposes of evaluation: improved teaching and student learning. We explore this intended use in relation to evaluation methods. Student evaluation of teaching (SET) is one of several components in educational evaluation and does not readily lead to improvements in teaching and student learning (Beran and Violato 2005; Kember et al. 2002; Stein et al. 2013).

The majority of student evaluations of teaching are retrospective quantitative course evaluation surveys (Erikson et al. 2016; Richardson 2005). Qualitative evaluation is seen as a viable alternative to quantitative evaluation methods (Darwin 2017; Grebennikov and Shah 2013), but has been subject to less empirical research (Steyn et al. 2019). Additionally, few empirical studies of SET have focused on aspects relevant for student learning (Bovill 2011; Edström 2008), even though improved student learning is promoted as the main purpose of educational evaluation (Ryan and Cousins 2009, pp. IX–X).

Querying data from The Arctic University of Norway (UiT), wherein both quantitative and qualitative evaluation methods appear, we will explore the experiences and perceptions of student evaluation among students and academics with the following research question:

> How do different evaluation methods, such as survey and dialogue-based evaluation, invite students to provide feedback about aspects relevant to their learning processes?

In this exploratory study, we investigate how different evaluation practices focus on aspects relevant to student's learning processes. We do not attempt to measure student learning in itself, but to scrutinise students' and academics' perceptions of how well course evaluation methods measure aspects they regard as relevant to student learning processes. We will discuss whether there are characteristics of the methods themselves that might affect the intended use. In this paper, academics refer to teaching academics, also named as teachers.

Furthermore, does the term 'student evaluation' refer to evaluations developed and initiated locally, comprised of students' feedback about academic courses in health profession education programmes. The definition of evaluation in the internal quality assurance system at UiT states that "Evaluation is part of the students' learning process and the academic environments' self-evaluation" (Universitetet i Tromsø 2012).

1.1 Student evaluation in higher education

It has been argued that evaluation in and of higher education is a balancing act between control, public accountability and quality improvement (Dahler-Larsen 2009; Danø and Stensaker 2007; Raban 2007; Williams 2016). In practice, the main function of student evaluation has shifted in the last decades from teaching development to quality assurance, which is important for administrative monitoring (Chen and Hoshower 2003; Darwin 2016; Douglas and Douglas 2006; Spooren et al. 2013). However, improved teaching and student learning are still advocated as objectives in policy documents in Norway where this study is conducted (Meld. St. 7 2007–2008; Meld. St. 16 2016–2017). Moreover, students (Chen and Hoshower 2003; Spencer and Schmelkin 2002) and teachers (Nasser and Fresko 2002) identify improved teaching and student learning as main purposes of student evaluation. Despite an overall aim to improve teaching and the generally positive attitudes of academics towards evaluations (Beran and Rokosh 2009; Hendry et al. 2007; Nasser and Fresko 2002; Stein et al. 2013), studies conclude that the actual use of evaluation data for these purposes is low (Beran et al. 2005; Kember et al. 2002; Stein et al. 2013).

Research has identified several explanations for why academics do not use survey responses: superficial surveys (Edström 2008), low desires to develop teaching (Edström 2008; Hendry et al. 2007), little support with respect to how to follow up the data (Marsh and Roche 1993; Neumann 2000; Piccinin et al. 1999), absence of explicit incentives to make use of these data (Kember et al. 2002; Richardson 2005), time pressure at work (Cousins 2003), scepticism as to the relevance of students' feedback in teaching improvement (Arthur 2009; Ballantyne et al. 2000) and a belief that these surveys are mainly collected as part of audit and control (Harvey 2002; Newton 2000). Well-known biases in student evaluation might also play a role in academics' scepticism towards use of student evaluation data for improvement of teaching (Stein et al. 2012). Research on bias in student evaluation has shown that several aspects of courses have a negative impact on student ratings, many of which academics cannot control or change: quantitative courses get more negative ratings than humanistic courses (Uttl and Smibert 2017), bigger group sizes affect student ratings negatively compared with smaller group sizes (Liaw and Goh 2003), graduate courses and elective courses are rated more favourably than obligatory courses (Patrick 2011) and female teachers receive lower ratings than male colleagues (Boring 2017; Fan et al. 2019).

Student evaluation data is more likely to be used in contexts where academics aim for constant improvement of teaching and courses (Golding and Adam 2016) and receive consultation and support on how to use evaluation data for course development (Penny and Coe 2004; Piccinin et al. 1999; Roche and Marsh 2002). Few evaluations collect in-depth information about student learning processes, such as which aspects of courses students consider as important for their learning (Bovill 2011). A recent meta-

analysis by Uttl et al. (2017) concludes that high student ratings of teaching effectiveness and student learning are not related. From our point of view, it is necessary to go beyond the lack of correlation between highly rated professors and student learning, and seek knowledge about the complexity of SET and its intended uses. It is noteworthy that this meta-analysis included only conventional surveys, which is the most dominant evaluation method. The use of surveys for obtaining student feedback on teaching and academic courses is time-efficient and often focuses on students' satisfaction with a course and the teachers (Bedggood and Donovan 2012; Richardson 2005), or the teacher's performance (Ryan 2015).

Dialogue-based evaluation methods, however, have been suggested as a viable alternative to quantitative evaluation methods—an alternative with more potential to facilitate reflection and dialogue between students and educators about their learning. These dialogues can provide deeper and more context-specific feedback from the students and can be useful in course development (Cathcart et al. 2014; Darwin 2017; Freeman and Dobbins 2013; Steyn et al. 2019). There is significantly less research on qualitative evaluation methods compared with quantitative methods (Steyn et al. 2019).

In this study, we attempt to get insight about what SET is measuring using empirical data from a university where both dialogue-based and written evaluation methods take place. This may help us understand why student evaluation data is not used more actively to improve teaching and student learning. Furthermore, insight about what SET is measuring can play a role in the design of student evaluation. It may also lead to a better understanding of the low correlation between student learning and highly rated professors.

1.2 Analytical framework

In this study, we draw upon the central principle in UFE: that evaluation should be judged by its utility to its intended users (Patton 2008). Every evaluation has many intended uses and intended users. The utility depends on how the evaluation data is used to achieve the overall aims, which, as already stated, we regard as improved teaching and student learning. In the context of this study, central intended users of internal student evaluation data are academic leaders and teachers at the programme level. We also regard students as intended users because students are users of evaluation while they are studying, particularly when student evaluation is understood as part of students' learning processes. Moreover, evaluation data can play a role for future students when choosing which institution and programme they apply to. Intended users' perspectives on evaluation purposes and uses are essential to UFE. Drawing upon social constructivism, we consider student evaluation as a phenomenon constructed by actors in the organisation. As students and academics are central actors in evaluation, we regard their perspectives as important.

Involvement of intended users throughout the evaluation process is central to UFE. Such involvement is regarded as a way to establish a sense of ownership and understanding of evaluation, which in turn will increase use (Patton 2008, p. 38). Involvement of intended users can occur in the planning stage of an evaluation, by, for example, generating evaluation questions, together with the intended users, that they regard as relevant and meaningful (Patton 2008, pp. 49–52). Active involvement can

also be in analysis or implementation of findings. In essence, Patton (2008, p. 20) states that everything that happens in all the different stages of an evaluation process can impact use, from planning to implementation of findings. According to principles in UFE, evaluation should include stakeholders or users, and be planned and conducted in ways that acknowledge how both findings and processes are central to use. Moreover, findings and process should inform decisions that lead to improvement of the evaluated areas (Patton 2008). Learning based on educational evaluation is often described as solely the organisational or personal learning facilitated by the data described in evaluation reports (Niessen et al. 2009). In this study, however, learning in evaluation is regarded as a complex socially constructed phenomenon that occurs in different stages and at different levels in the evaluation process. Patton (1998) created the term *process use* to describe learning that happens at different levels *during* the evaluation process *before* an evaluation report is written. Process use refers to both individual changes in thinking and behaviour and as organisational changes in procedures and cultures "as a result of the learning that occurs during the evaluation process" (Patton 2008, p. 155).

In this study, learning at the individual level for students relates to both intended and unintended learning as consequences of evaluation processes. From a teacher perspective, 'process use' can be increased awareness and insights about student learning processes in a course. Since Patton launched the term process use, learning during the evaluation process has been acknowledged by many evaluators as important and is implemented in different approaches to evaluation (Donnelly and Searle 2016; Forss et al. 2002).

2 Design and methods

2.1 Institutional context

This study was conducted at The Faculty of Health Sciences, at UiT, a Norwegian university with 16,000 students, 3600 employees and eight faculties. This faculty is the largest in terms of number of students, with almost 5000 enrolled students, and it offers programmes and courses from undergraduate to graduate levels.

Norwegian legal act The act relating to universities and university colleges requires each university to have an internal quality assurance system in which student evaluation is integrated. These quality assurance systems shall assure and improve educational quality (Lovdata 2005). Within the confines of the law, each university has autonomy to create their local quality assurance system and make decisions on its form, content and delivery. The internal quality assurance system at this university allows for different evaluation methods, and aims to capture both the perspectives of both academics and students. Although the legal act (Lovdata 2005) mandates student evaluation, evaluation of teaching and courses by the academics are not regulated by law.

Programme managers or course leaders are, according to the local quality assurance system at the university, responsible for designing and carrying out student evaluations. They can choose between dialogue-based or written evaluation methods, or they can combine these. The internal quality assurance system, however, recommends the use of

formative evaluation which ensures user involvement and invites students to give feedback relevant to educational quality. Moreover, the local quality assurance system describes that student evaluation should contribute to giving students an active role in quality assurance. It is underlined that student evaluation is considered as both part of students' learning processes and the university's self-evaluation (Universitetet i Tromsø 2012).

2.2 Methods

Eight health profession education programmes were included in this research. The programmes are not identified by profession, as this may affect the anonymity of the informants. Students from the programmes were interviewed in focus groups, and teaching academics were interviewed in semi-structured interviews. The interviews are the primary data in the study.

Students were recruited for focus group interviews with assistance from student representatives from each educational programme. The student representatives were also invited to participate in the interviews themselves, with one inclusion criterion being prior experience and participation in student evaluation. For the academics, the inclusion criteria were responsibility for a minimum of one academic course, including teaching, and experience with designing, distributing and/or summarising student evaluations. Two of the informants were programme leaders and therefore more involved in programme evaluation than teachers who were solely responsible for one or more courses. In this paper, teaching academics are referred to as academics or teachers.

Educational documents from 2013 to 2015 were included as supplementary data sources. These documents were surveys from the eight programmes, descriptions about dialogue-based evaluation methods and educational reports. The documents were studied before each interview for contextual background.

Leaders of programmes, departments and faculty were informed about the project, and ethical approval was granted from the university and The Norwegian Centre for Research Data (NSD).

The focus group interviews with students were conducted from 2016 to 2017; the groups ranged in size from three to seven students ($n = 30$), and interviews lasted approximately 90 min. Using an interview guide, students were encouraged to engage in dialogue about the different topics. When student informants are quoted in this paper, they are referred to as focus groups A–H. No individuals are named.

The semi-structured interviews with the teachers were conducted in 2016–2017. An interview guide was developed for this project with topics and open-ended questions to uncover different aspects of the evaluation practice. The interviews lasted between 75 and 90 min. When the academics are quoted, they are referred to as informants A–H.

At the beginning of each interview, the informants were asked about their background, their role in the evaluation practice, the purposes of evaluation in higher education and about national and local regulation of student evaluations. Next, the interviews focused on local evaluation practices in the represented programme, including characteristics of different methods. The interviews concluded with questions about use of student evaluation in relation to educational quality and course development.

2.3 Analysis

Interview data were audio recorded and transcribed verbatim. A thematic analysis of the interview data in different stages was performed in NVivo. In the first stage, the thematic analysis was inductive, and the empirical data were sorted by codes that described the data, created by the first author. Descriptive and process coding were the dominant code types (Saldaña 2013). Descriptive and process codes, like evaluation methods, purpose of evaluation, student involvement and implementation of evaluation findings, were used to illustrate the evaluation practice and its characteristics for the eight programmes. The first author created in total 14 codes in the first stage, each of which had subcodes.

In order to understand phenomena and create meaning, categories were developed from the initial codes in the second stage of the analysis. In this stage, some codes were merged, others were split and subcategories were developed. The categories developed in this stage were less descriptive than the codes in the first stage, and the thematic analysis was more abductive as the coding also was informed by theory, like process use, learning-oriented evaluation and feedback expectations. Using the categories created in the second stage, themes were developed in the last stage of the process. These themes were overall themes that are presented in this paper and two upcoming papers. Throughout the analytical process, interview data and documents from the same programme were compared to create a broader picture of the evaluation practice for each programme. Although three main stages describe the analytical process, these stages also overlap in an iterative process.

3 Results

Both students and academics stated that evaluation practices varied greatly, even within the same programme. However, the most common methods for obtaining student feedback for the included programmes were different formats of surveys and dialogue-based evaluation methods. Whereas the surveys were distributed by administrative staff at the end of the course or after the final assessment, the dialogue-based evaluations often took place before final exams. In addition to the evaluation methods included in this study, the academics received student feedback in numerous other ways, e.g., via student representation on institutional bodies, by e-mail, orally or through student representative meetings with academic or administrative staff. During the interviews, students spent more time elaborating upon their experiences with dialogue-based evaluation methods than with surveys. Consequently, dialogue-based evaluation is given more attention in this paper. Academics, on the other hand, gave almost equal attention to dialogue-based evaluation methods and surveys.

The students considered quality improvement to be the main aim of student evaluations, whereas the academics explained that student evaluation is multifunctional in that it aims to both improve and assure educational quality.

3.1 Written evaluation surveys

Document analysis of the written surveys at the institution showed that the number of questions varied from 5 to 75. The programme that used the shortest survey asked the

students to rate their satisfaction level with smiley faces for one of the courses. The longest survey contained 75 questions using a Likert scale (1–5) and had no open-ended questions. Most of the course evaluation surveys had similarly worded open-ended questions such as 'What worked well in the course?' and 'Which factors about the course could be improved?'

The types of questions in written evaluations varied, but with some similarities across the eight programmes. None of the surveys concentrated solely on the course's curriculum, and most of the surveys included questions about resources and facilities, including the library, learning management systems, computer labs and general information about the administration. Only two programmes included the teachers' names in surveys, but not for every course. None of the programmes had separate surveys for teacher evaluation and course evaluation.

Learning outcome descriptions from the syllabus did not feature in any of the surveys. However, two programmes included questions about the achieved learning of the main course topics. Students from these programmes spoke positively about the questions and design of the surveys, but shared examples about course descriptions with diffuse learning outcomes and how this made focused feedback difficult. Students from programmes that did not include questions on course topics also found it challenging to fill out these surveys because of how the questions were phrased. Questions were often considered non-specific, i.e., asking about how satisfied they were with the teaching or the learning activities. They described filling out surveys as challenging, and often frustrating and meaningless. One informant elaborated and said: "What does it mean when I am satisfied with the course? Is it the instructor's ability to make the course exciting, the pedagogical approach, the course literature or learning activities? They should be more specific in the questions" (focus group H).

The students explained that non-specific questions focused on how teaching or learning activities in general helped them achieve the learning outcomes, rather than on how specific learning activities helped them achieve specific expected learning outcomes. The students did not bring paper examples to the interview, but analysis of the documents supported students' views and showed that the templates from the departments included general questions such as

- How would you rate the teaching in the course?
- How was the outcome of the teaching in the course?
- How would you rate the learning outcome in group learning activities?
- How would you rate the learning outcome in the lectures?
- To which degree did the teacher spark your interest for the course topic?

The students found these questions impossible to rate because courses often involved several teachers and learning activities. Therefore, their answers were often an average score of all the activities and did not provide any specific information about the different learning activities or teachers. Open-ended questions and open spaces for general feedback were considered by students to be very important, especially in questionnaires with non-specific questions. They wanted to provide feedback about factors that really mattered to them, or to explain low ratings.

The academics described written evaluation methods as time-efficient and easy ways to compare data over time. None of the programmes used standardised surveys;

however, some of the programmes had a list of questions to consider using for course evaluation surveys. The academics did note limitations with existing surveys and evaluation practices. Like the students, academics expressed that many of the questions in the templates were non-specific. In course evaluations where such questions were used, the academics found that the answers were unusable for course and teaching development because they did not know which aspects of the course the students had evaluated. Hence, the informants claimed that they created their own questions or surveys, stressing the importance of well-phrased and specific questions. When asked to give examples of pre-defined questions they considered unsuitable for course improvement, one informant answered:

> Yes, they ask how satisfied you are with the teaching, kind of a general and overall question, how do you answer that? Which learning activity are you evaluating? It refers to the whole course, maybe some activities were good and others really bad, then you have to rate it in the middle, it does not tell me anything about how they valued different parts of the course. (informant H)

Both students and academics stressed the importance of evaluating whether learning activities, reading lists and practical placement helped students achieve expected learning outcomes rather than the level of satisfaction with teaching performance. As mentioned above, none of the surveys included questions about whether the learning activities or teaching in the courses helped students achieve expected learning outcomes.

3.2 Dialogue-based evaluation methods

Dialogue-based evaluations are conducted with selected students or the entire student group and one or more staff members (Universitetet i Tromsø 2012). The format of dialogue-based evaluation varies between café-dialogue evaluations, focus group discussions, student-led discussions and meetings with student representatives and academics; however, we refer to all types as dialogue-based evaluation methods. These evaluation methods had more open formats and fewer pre-defined questions than written surveys. Students appreciated how these dialogues allowed them to set the agenda and express their opinions about aspects of the courses that mattered to them.

Academics who used dialogue-based evaluation methods emphasised the use of an open format in discussions and encouraged students to facilitate the discussions. They considered students' feedback to be valuable for formative course adjustments and course planning.

The students valued the immediate responses they received to their feedback in dialogue-based evaluations. This two-way dialogue was highly appreciated by the students, even when their feedback was not used in course development. Moreover, they said that teachers in these dialogues often explained why the curriculum or teaching was designed the way it was, sometimes in relation to the expected learning outcomes. All the students expressed that if teachers showed interest in their opinion and enhanced dialogues about their learning processes, it positively affected their motivation to provide feedback.

Academics believed that it is important to establish a culture of continuous dialogue with the students throughout the course. They considered an open-door policy and

dialogues after lectures as important informal evaluation activities. Academics shared examples about students' feedback that required immediate follow-up and underlined that it is important to create a culture of dialogue in order to capture different issues with the course. However, they said that such culture takes time to establish and it is based upon trust and a safe learning environment.

One informant emphasised that it is the teachers' responsibility to create a safe environment that invites students to give feedback: "I try to meet the students with an attitude that learning is something we do together, but it is the teachers' responsibility to facilitate that students have a good learning environment" (informant C). He considered dialogue with students about their learning processes as valuable for course planning, his teaching development and the students' learning environment. However, he believed that power asymmetry between students and academics could be a hindrance to honest feedback.

3.3 Awareness of students' learning processes

Students and academics experienced an increased awareness about learning as a result of their reflections during the evaluation process. Referring to dialogue-based evaluation, one student said: "It helps you to reflect upon what you have learned in a course; you start to reflect upon it and reflection has proven to be useful if you are learning something new" (focus group F).

In terms of learning processes, students emphasised that learning could occur in different ways for themselves and future students, as a result of evaluation data and participation in the evaluation process. Firstly, when evaluation data are followed up and subsequently lead to improvements in teaching, future students will have better learning conditions. Secondly, during these dialogues, the informants themselves developed professional competencies, such as communication and reflection skills. This is exemplified by one student who said that dialogue-based evaluations helped her learn how to give constructive feedback and communicate clearly. Another student emphasised how necessary it is within the health professions to be analytical and have good reflective skills; he believed that the dialogue-based evaluations helped him develop these skills. The students regarded these skills as important for their learning processes and professional development:

> You learn how to be a good teacher yourself; not all of us are going to be teachers but you learn how to talk to people and that is especially important if you are explaining something to somebody or teaching them something (focus group F)

3.4 Student evaluation and improved teaching

Academics stated that students' feedback was used for minor adjustments during and after the courses. When asked to give examples of adjustments in courses as a result of student evaluations, they shared examples of changes related to student placement and practice instructors often as a result of negative feedback: "We have replaced practice instructors with others based upon negative feedback over time" (informant G). Students had similar stories about changes that took place based on their feedback, often related to issues with student placements or practice instructors.

Student feedback that inspired course plans or changes in the curriculum was seldom from student evaluation alone. Course leaders pointed out that if changes were made to assignments, exams or teaching methods, these changes were based on systematic feedback from students over time and on discussions among academic colleagues. They underscored that their pedagogical knowledge and available administrative and curricular resources also affected how they followed up on student feedback.

All the academics agreed that there were several reasons for caution when using student feedback for course development. Four academics expressed that students did not have the same knowledge as teachers about pedagogics nor the required skills for the profession. These four argued that students may be experts on their learning processes but not on teaching. Moreover, five of the academics questioned the validity of the surveys: they questioned if the right questions were asked at the right time. Some said that they believed students often rated active learning activities negatively, because these activities required a great deal of effort and involvement. Furthermore, they believed that funny and entertaining teachers got better evaluations than their peers who were not so entertaining. They also doubted that achieved learning was related to the satisfaction rating. One academic elaborated:

> We have talked about it at work... that you are not only an 'educator' but also an 'edutainer' in your teaching. You have to be able to engage the students as well as be fun. It can be a challenge for many. Then you have to evaluate the teaching, and maybe they rate the teaching positively because you were able to engage the students, but the learning outcome was probably not that high. (informant F)

Informant F therefore suggested that it might be a good idea to differ between performance and learning outcome in the written evaluations.

The academics expressed that evaluation practices had been given little attention from university management and were seldom discussed among colleagues. This is in contrast to the implementation of a learning outcome-based curriculum that was given significant attention due to the Norwegian national qualification framework of 2012 (Meld. St. 27 2000–2001). One academic said: "We have been working a lot on creating learning outcomes.... but we didn't include the evaluation in this work" (informant F).

4 Discussion

Student evaluation was originally introduced in education as a pedagogical tool to provide a valuable impetus for improving teaching practices. Several decades later, this function of student evaluation has been backgrounded as a stronger discourse on quality assurance and control functions proliferate within academic systems (Darwin 2016, p. 3). With the originally intended use in mind, we will discuss how the evaluation methods themselves invite students to provide feedback about aspects relevant for their learning processes and how academics and students portray the use of this data through evaluation processes.

4.1 Dialogue-based vs. written evaluation

In this study, the focus on students' learning processes was more apparent in dialogue-based evaluation methods than in surveys. When the students talked about dialogue-based evaluation methods, they explained how these dialogues, in contrast to surveys, invited them to give feedback about aspects they regarded as relevant to their learning processes and what really mattered for them. This is probably strongly related to their expectations of the intended use of evaluation. Regardless of the types of dialogue-based methods, the students felt that their experiences and perspectives were listened to and seriously considered in these discussions. In these dialogues, they could focus on the course aspects that they regarded as contributing to their learning.

Dialogue-based evaluation methods invited students to provide feedback about the courses, particularly about what hinders or improves learning. If the intention is to use the evaluation as a pedagogical tool for course and teaching improvement, more effort seems to be needed to increase dialogue with students about factors that affect student learning during courses (Darwin 2017; Huxham et al. 2008). This is aligned with what the informants in this study expressed. Our study also indicates that feedback from dialogue-based evaluation methods are already used more frequently than survey data to shape course changes.

Learning-oriented evaluation approaches are characterised by involvement of the practitioners in the evaluation process (Donnelly and Searle 2016), and are also central in UFE (Patton 2008). This study shows that the students' role in written evaluation practices at this university was solely to respond to surveys and that they had no influence on which questions were asked about the courses and their learning processes. This is in contrast to student involvement in dialogue-based evaluations, where they were invited to set the agenda for what should be evaluated. It is a principle of UFE to invite participants or intended users to participate in the planning of evaluation, in order to increase use of findings (Patton 2008). Participant involvement in the evaluation process can contribute to establishing a sense of ownership over the evaluation and increasing the relevance of evaluation questions, which in turn might affect use of the subsequent data. More dialogue between students and academics and user involvement in evaluation processes can be keys to achieve the objective of evaluation stated in the internal quality assurance system where student evaluation is regarded as part of students' learning processes.

4.2 Students expectations

The students were eager to share their opinions about how they believed teaching and courses could be improved in order to enhance their learning, which they regarded as the purpose and intended use of evaluation. Nevertheless, not all evaluations invited them to provide feedback about how the courses facilitated learning. They asked why many of the questions, particularly in written evaluations, were requesting responses about satisfaction level, not achieved learning. Students considered themselves to be experts on their own learning processes and therefore expected to be invited into dialogues about whether the learning activities in a course were successful or not. After all, they are the primary users of the educational system, wherein student learning is the goal. Additionally, this new generation of students has probably been involved in

dialogue about their learning processes since elementary school, as user involvement in education is an objective in the Norwegian Education Act (Lovdata 1998) and learning-outcome based education has been standard as long as they have been enrolled in school (Prøitz 2015). Consequently, they expected to provide feedback about what really matters for them: their learning processes, not the teachers nor the teaching.

A shift in the view of quality assurance, moving from a teaching to a learning focus, has changed higher education in many European countries (Smidt 2015). Research on the sector in Norwegian higher education has shown that learning-outcome based curriculum has become standard (Havnes and Prøitz 2016; Prøitz 2015) and the learning focus has increased in teaching and assessment (Michelsen and Aamodt 2007). However, the findings in this study, aligned with other studies, indicate that the emphasis on learning in written student evaluation methods appears to be low (Bergsmann et al. 2015; Edström 2008; Ramsden 2003). Based upon our findings and research on student evaluation, we believe SET has the potential to facilitate reflections on students' learning processes among intended users, though this potential has not yet been realised. The need to focus on students' learning processes in evaluation of teaching in higher education is also emphasised by Bergsmann et al. (2015, p. 6), who states that: "…once students and their competencies are put center stage, this evaluation aspect is of high importance."

4.3 Evaluation as process use

In order to evaluate complex contemporary learning environments, qualitative evaluation approaches are recommended (Darwin 2012, 2017; Haji et al. 2013; Huxham et al. 2008; Nygaard and Belluigi 2011). This is because qualitative evaluation approaches seem to capture aspects of how and why learning best takes place and how learning contexts might affect learning processes (Haji et al. 2013; Nygaard and Belluigi 2011). In this study, the students expressed how the dialogue-based evaluations gave them an opportunity to reflect upon their learning processes, and that these reflections improved their awareness of achieved learning and helped them develop professional skills. These are examples of learning that takes place during the evaluation process, defined as *process use* in evaluation theory. Furthermore, this learning opportunity can be strengthened if used consciously. The student description of how they developed professional competencies during dialogue-based evaluations can be understood as meta-perspectives of learning, and illustrates how student evaluations also can be opportunities for reflective learning. Other researchers have suggested reframing student evaluation by focusing more on dialogue and reflections about students' learning processes (Bovill 2011; Darwin 2012, 2016; Ryan 2015).

In the interviews, the academics did not mention process learning for students in course evaluation. When they were asked about how evaluation might affect students' learning, they referred to the learning of future students that could be enhanced by course improvements based on previous student feedback. The academics also underlined that dialogue with students during courses increased their awareness of student learning processes. These discussions, both in an informal setting and as part of a scheduled dialogue-based course evaluation, made the academics more attentive to the views of others and changed how they thought about learning. UFE states that learning during the evaluation process—process use—has often been overlooked in

evaluation reports, which instead focus heavily on findings and summaries from surveys (Cousins et al. 2014; Forss et al. 2002; Patton 1998). Aligned with the statements above, process use and learning during the evaluation process were not described in the internal evaluation reports at this university. However, the academics valued these evaluative dialogues during courses, and elaborated on how they informed their teaching. In UFE, it is desirable to learn from and use both the findings and what happens in the evaluation process. When process learning is made intentional and purposeful, the overall utility of an evaluation can increase (Patton 2008, p. 189). The students in this study shared examples of process use during evaluation and explained how this practice increased their awareness of achieved learning and helped them develop reflective skills important for their health professions.

4.4 Evaluation questions and their fitness for the intended purpose

Document analysis of templates and surveys showed that many of the questions in written evaluations asked students about their satisfaction level, and not of how aspects about the course and teaching affected their learning processes. Both students and academics referred to non-specific and unclear questions as meaningless. They cautioned that data generated from such questions should be used with great care because they did not know what they intended to measure. When questions are open to interpretations by respondents, it might affect the validity of the results. Additionally, those who develop questions for the templates might have different interpretations of the questions than the respondents do (Desimone and Le Floch 2004; Faddar et al. 2017). Two criteria for UFE questions are that intended users can specify the relevance of an answer for future action and that primary intended users *want* to answer the evaluation questions (Patton 2008, p. 52). Unclear and non-specific questions open for interpretation are not regarded as relevant for the future action of improving teaching or learning among the intended users and informants in this study. Consequently, the students' motivation to respond to evaluation like this was low.

Dialogue-based evaluation at the university was led and developed by academics, whereas the surveys were often designed by the administrative staff. The role of administrative staff is obviously different from the role of students and academics, and it may affect how staff define the purposes of evaluation and design the evaluation methods. Student learning and educational quality are shared goals for all stakeholders. Nevertheless, administrative staff, teaching academics, educational leaders and students have different perspectives and understandings about what constitutes high-quality learning and teaching (Dolmans et al. 2011).

The decision of academics not to use templates provided by the administration and instead create their own surveys may be related to what they regarded as the intended use of evaluation. In order to improve their teaching, they need qualitative, rich and in-depth knowledge about what in the teaching hindered and facilitated learning, rather than satisfaction rates of the students. The administrative staff, on the other hand, are often responsible for monitoring the quality assurance system and ensuring that the legal regulation is followed. They need different kinds of data than the academics for this purpose. It may therefore not be surprising that they develop evaluation questions that are well suited for quality assurance, control and accountability, but not for teaching improvement. The accountability focus in the legal regulation might

overshadow the development purpose for the administrative staff. When the intended users disagree on the aim of an evaluation, they will, according to UFE, also judge or value the utility of the same evaluation differently.

Traditionally, quality assurance systems have been developed by administrative staff (Newton 2000; Stensaker 2002). This is also the case in Norway (Michelsen and Aamodt 2007). The links between those who have developed the quality assurance system and the surveys, their roles in higher education and the types of questions asked are important.

Although the written evaluation methods at this university are rather teaching-oriented, surveys can also be learning-oriented. By putting more effort into the design of evaluation questions, written evaluations can also be pedagogical tools useful for improvement of teaching and learning. One of the many ways to accommodate this, in line with principles from UFE (Patton 2008, p. 38), is to involve the academics and students more in evaluation planning.

4.5 Concluding remarks

The stakeholders and intended users of evaluations interviewed in this study (students and academics) expressed that the types of questions were different in the two evaluation methods. Students in particular considered the questions in dialogue-based evaluation methods to be learning-oriented and those in surveys as teaching-oriented.

The types of questions in today's written evaluation methods do not seem to invite students to give feedback on aspects relevant to their learning processes to the same extent as dialogue-based methods do. Moreover, the informants elaborated that both teachers and students benefited from evaluative dialogue; the students reflected upon their own learning processes, and the teachers received valuable feedback about achieved learning outcomes and the success of specific learning activities. If this feedback is used, it will benefit future students by improving their learning environments and their learning processes. Furthermore, the students shared examples from dialogue-based evaluation activities wherein professional competencies were developed. This is an unintended but a positive effect of evaluation that needs further exploration.

Most of the written evaluation surveys were found to be rather superficial, with questions focusing on overall satisfaction with teaching rather than on aspects that facilitated or inhibited student learning. Academics and students found that such evaluation data from surveys were not relevant for the intended purpose of student evaluation, which is teaching development for the sake of students' learning processes. Moreover, many of the questions in surveys were not requesting feedback from students that would be suitable for educational enhancement. The responses from academics indicated that the administrative staff had an important role in the development of written evaluations. This finding calls for further study. In UFE, evaluation is judged by its utility for its intended users. In this study, we have defined intended use as improved teaching and learning and intended users as academics and students. Both groups expected evaluation to collect relevant data for these purposes but agreed that there is an unrealised potential to focus more on students' learning processes in student evaluation.

If the intended users in this study, however, were politicians, administrators or the university management, and the intended purposes were quality assurance and

accountability, the utility of the data must be judged by these parties. When evaluation is described as a balancing act between quality assurance and quality enhancement, we relate this to the diverse stakeholder group of actors who have different roles in the education system and different interests in evaluation. It is therefore important to have the intended users and the intended use of evaluation in mind when designing them.

With a learning outcome-based approach becoming the standard in higher education, it is time to reconsider evaluation practices and revise teaching-focused evaluation questions. The students expected student evaluation to focus on their learning and were surprised by teaching-oriented questions in the surveys. If student evaluation data should be used as intended, to improve teaching and learning, and be included in student learning processes, it is time to stop asking students questions about satisfaction and rather request feedback about what hindered and facilitated learning. Evaluation methods are a pedagogical tool with the potential to strengthen student learning processes in the future.

References

Arthur, L. (2009). From performativity to professionalism: lecturers' responses to student feedback. *Teaching in Higher Education, 14*(4), 441–454. https://doi.org/10.1080/13562510903050228.

Ballantyne, R., Borthwick, J., & Packer, J. (2000). Beyond student evaluation of teaching: identifying and addressing academic staff development needs. *Assessment & Evaluation in Higher Education, 25*(3), 221–236.

Bedggood, R. E., & Donovan, J. D. (2012). University performance evaluations: what are we really measuring? *Studies in Higher Education, 37*(7), 825–842.

Beran, T., & Rokosh, J. (2009). Instructors' perspectives on the utility of student ratings of instruction. *Instructional Science: An International Journal of the Learning Sciences, 37*(2), 171–184. https://doi.org/10.1007/s11251-007-9045-2.

Beran, T., & Violato, C. (2005). Ratings of university teacher instruction: how much do student and course characteristics really matter? *Assessment and Evaluation in Higher Education, 30*(6), 593–601. https://doi.org/10.1080/02602930500260688.

Beran, T., Violato, C., Kline, D., & Frideres, J. (2005). The utility of student ratings of instruction for students, faculty, and administrators: a "consequential validity" study. *Canadian Journal of Higher Education, 35*(2), 49–70.

Bergsmann, E., Schultes, M.-T., Winter, P., Schober, B., & Spiel, C. (2015). Evaluation of competence-based teaching in higher education: from theory to practice. *Evaluation and Program Planning, 52*, 1–9. https://doi.org/10.1016/j.evalprogplan.2015.03.001.

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics, 145*, 27–41. https://doi.org/10.1016/j.jpubeco.2016.11.006.

Bovill, C. (2011). Sharing responsibility for learning through formative evaluation: moving to evaluation as learning. *Practice and Evidence of Scholarship of Teaching and Learning in Higher Education, 6*(2), 96–109 Retrieved from eprints.gla.ac.uk/57707/1/57707.pdf.

Cathcart, A., Greer, D., & Neale, L. (2014). Learner-focused evaluation cycles: facilitating learning using feedforward, concurrent and feedback evaluation. *Assessment & Evaluation in Higher Education, 39*(7), 790–802. https://doi.org/10.1080/02602938.2013.870969.

Chen, Y., & Hoshower, L. B. (2003). Student evaluation of teaching effectiveness: an assessment of student perception and motivation. *Assessment & Evaluation in Higher Education, 28*(1), 71–88. https://doi.org/10.1080/02602930301683.

Christie, C. A. (2007). Reported influence of evaluation data on decision makers' actions: an empirical examination. *American Journal of Evaluation, 28*(1), 8–25. https://doi.org/10.1177/1098214006298065.

Cousins, J. B. (2003). Utilization effects of participatory evaluation. In D. Kellaghan, L. Stufflebeam, & L. A. Wingate (Eds.), *International handbook of educational evaluation* (pp. 245–265). Boston: Springer.

Cousins, J. B., Goh, S. C., Elliott, C. J., & Bourgeois, I. (2014). Framing the capacity to do and use evaluation. *New Directions for Evaluation, 2014*(141), 7–23. https://doi.org/10.1002/ev.20076.

Dahler-Larsen, P. (2009). Learning oriented educational evaluation in contemporary society. In K. Ryan & B. Cousins (Eds.), *The SAGE International Handbook of Educational Evaluation* (pp. 307–322). https://doi.org/10.4135/9781452226606.

Danø, T., & Stensaker, B. (2007). Still balancing improvement and accountability? Developments in external quality assurance in the Nordic countries 1996–2006. *Quality in Higher Education, 13*(1), 81–93. https://doi.org/10.1080/13538320701272839.

Darwin, S. (2012). Moving beyond face value: re-envisioning higher education evaluation as a generator of professional knowledge. *Assessment & Evaluation in Higher Education, 37*(6), 733–745. https://doi.org/10.1080/02602938.2011.565114.

Darwin, S. (2016). *Student evaluation in higher education: reconceptualising the student voice*. Switzerland: Springer International Publishing.

Darwin, S. (2017). What contemporary work are student ratings actually doing in higher education? *Studies in Educational Evaluation, 54*, 13–21. https://doi.org/10.1016/j.stueduc.2016.08.002.

Desimone, L. M., & Le Floch, K. C. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Educational Evaluation and Policy Analysis, 26*(1), 1–22. https://doi.org/10.3102/01623737026001001.

Dolmans, D., Stalmeijer, R., Van Berkel, H., & Wolfhagen, H. (2011). Quality assurance of teaching and learning: enhancing the quality culture. In *Medical education: Theory and practice* (pp. 257–264). Edinburgh: Churchill Livingston, Elsevier Ltd.

Donnelly, C., & Searle, M. (2016). Optimizing use in the field of program evaluation by integrating learning from the knowledge field. *Canadian Journal of Program Evaluation, 31*(3), 305–327.

Douglas, J., & Douglas, A. (2006). Evaluating teaching quality. *Quality in Higher Education, 12*(1), 3–13. https://doi.org/10.1080/13538320600685024.

Edström, K. (2008). Doing course evaluation as if learning matters most. *Higher Education Research and Development, 27*(2), 95–106. https://doi.org/10.1080/07294360701805234.

Erikson, M., Erikson, M. G., & Punzi, E. (2016). Student responses to a reflexive course evaluation. *Reflective Practice, 17*(6), 663–675.

European University Association. (2007). *Trends V: universities shaping the higher education area*. Retrieved from Brussels: http://www.eua.be/Libraries/publications-homepage-list/EUA_Trends_V_for_web.pdf?sfvrsn=6. Accessed 7 September 2018

Faddar, J., Vanhoof, J., & De Maeyer, S. (2017). Instruments for school self-evaluation: lost in translation? A study on respondents' cognitive processing. *Educational Assessment, Evaluation and Accountability, 29*(4), 397–420. https://doi.org/10.1007/s11092-017-9270-4.

Fan, Y., Shepherd, L. J., Slavich, E., Waters, D., Stone, M., Abel, R., & Johnston, E. L. (2019). Gender and cultural bias in student evaluations: why representation matters.(research article). *PloS one, 14*(2), e0209749. https://doi.org/10.1371/journal.pone.0209749.

Forss, K., Rebien, C. C., & Carlsson, J. (2002). Process use of evaluations: types of use that precede lessons learned and feedback. *Evaluation, 8*(1), 29–45. https://doi.org/10.1177/1358902002008001515.

Freeman, R., & Dobbins, K. (2013). Are we serious about enhancing courses? Using the principles of assessment for learning to enhance course evaluation. *Assessment & Evaluation in Higher Education, 38*(2), 142–151. https://doi.org/10.1080/02602938.2011.611589.

Golding, C., & Adam, L. (2016). Evaluate to improve: useful approaches to student evaluation. *Assessment & Evaluation in Higher Education, 41*(1), 1–14. https://doi.org/10.1080/02602938.2014.976810.

Grebennikov, L., & Shah, M. (2013). Student voice: using qualitative feedback from students to enhance their university experience. *Teaching in Higher Education, 18*(6), 606–618. https://doi.org/10.1080/13562517.2013.774353.

Haji, F., Morin, M. P., & Parker, K. (2013). Rethinking programme evaluation in health professions education: beyond "did it work?". *Medical Education, 47*(4), 342–351. https://doi.org/10.1111/medu.12091.

Hansen, H. F. (2009). Educational evaluation in Scandinavian countries: converging or diverging practices? *Scandinavian Journal of Educational Research, 53*(1), 71–87. https://doi.org/10.1080/00313830802628349.

Harvey, L. (2002). Evaluation for what? *Teaching in Higher Education, 7*(3), 245–263. https://doi.org/10.1080/13562510220144761.

Havnes, A., & Prøitz, T. (2016). Why use learning outcomes in higher education? Exploring the grounds for academic resistance and reclaiming the value of unexpected learning. *Educational Assessment, Evaluation and Accountability, 28*(3), 205–223. https://doi.org/10.1007/s11092-016-9243-z.

Hendry, G. D., Lyon, P. M., & Henderson-Smart, C. (2007). Teachers' approaches to teaching and responses to student evaluation in a problem-based medical program. *Assessment & Evaluation in Higher Education, 32*(2), 143–157.

Huxham, M., Laybourn, P., Cairncross, S., Gray, M., Brown, N., Goldfinch, J., & Earl, S. (2008). Collecting student feedback: a comparison of questionnaire and other methods. *Assessment & Evaluation in Higher Education, 33*(6), 675–686. https://doi.org/10.1080/02602930701773000.

Johnson, K., Greenseid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on evaluation use: a review of the empirical literature from 1986 to 2005. *American Journal of Evaluation, 30*(3), 377–410.

Kember, D., Leung, D. Y. P., & Kwan, K. P. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment & Evaluation in Higher Education, 27*(5), 411–425. https://doi.org/10.1080/0260293022000009294.

Liaw, S.-H., & Goh, K.-L. (2003). Evidence and control of biases in student evaluations of teaching. *International Journal of Educational Management, 17*(1), 37–43. https://doi.org/10.1108/09513540310456383.

Lovdata (1998). *Opplæringslova [Education Act]*. Lovdata Retrieved from https://lovdata.no/dokument/NL/lov/1998-07-17-61. Accessed 24 June 2019

Lovdata (2005). *Lov om universiteter og høyskoler [Act relating to Universities and University Colleges Section 1–6]*. Lovdata Retrieved from https://lovdata.no/dokument/NL/lov/2005-04-01-15/KAPITTEL_1-1#%C2%A71-1. Accessed 28 April 2018

Marsh, H. W., & Roche, L. (1993). The use of students' evaluations and an individually structured intervention to Enhance University teaching effectiveness. *American Educational Research Journal, 30*(1), 217–251. https://doi.org/10.2307/1163195.

Meld. St. 16 (2016-2017). *Kultur for kvalitet i høyere utdanning*. Oslo: Kunnskapsdepartement.

Meld. St. 27 (2000-2001). *Kvalitetsreform av høyere utdanning, Gjør din plikt, krev din rett*. Oslo: Kirke-, utdannings- og forskningsdepartementet Retrieved from https://www.regjeringen.no/no/dokumenter/stmeld-nr-27-2000-2001-/id194247/. Accessed 10 October 2018

Meld. St. 7 (2007-2008). *Statusrapport for Kvalitetsreformen i høgre utdanning*. Oslo: Kunnskapsdepartementet Retrieved from https://www.regjeringen.no/no/dokumenter/Stmeld-nr-7-2007-2008-/id492556/. Accessed 10 November 2018

Michelsen, S., & Aamodt, P. O. (2007). *Evaluering av Kvalitetsreformen*. Sluttrapport. Retrieved from http://hdl.handle.net/11250/279245. Accessed 7 September 2018

Nasser, F., & Fresko, B. (2002). Faculty views of student evaluation of college teaching. *Assessment & Evaluation in Higher Education, 27*(2), 187–198.

Neumann, R. (2000). Communicating student evaluation of teaching results: rating interpretation guides (RIGs). *Assessment & Evaluation in Higher Education, 25*(2), 121–134. https://doi.org/10.1080/02602930050031289.

Newton, J. (2000). Feeding the beast or improving quality? Academics' perceptions of quality assurance and quality monitoring. *Quality in Higher Education, 6*(2), 153–163. https://doi.org/10.1080/713692740.

Niessen, T. J. H., Abma, T. A., Widdershoven, G. A. M., & van der Vleuten, C. P. M. (2009). *The SAGE International handbook of educational evaluation*. https://doi.org/10.4135/9781452226606.

Nygaard, C., & Belluigi, D. Z. (2011). A proposed methodology for contextualised evaluation in higher education. *Assessment & Evaluation in Higher Education, 36*(6), 657–671. https://doi.org/10.1080/02602931003650037.

Patrick, C. L. (2011). Student evaluations of teaching: effects of the big five personality traits, grades and the validity hypothesis. *Assessment & Evaluation in Higher Education, 36*(2), 239–249. https://doi.org/10.1080/02602930903308258.

Patton, M. Q. (1997). *Utilization-focused evaluation : the new century text* (3rd ed.). Thousand Oaks: Sage.

Patton, M. Q. (1998). Discovering process use. *Evaluation, 4*(2), 225–233. https://doi.org/10.1177/13563899822208437.

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks: Sage publications.

Penny, A. R., & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: a meta-analysis. *Review of Educational Research, 74*(2), 215–253. https://doi.org/10.3102/00346543074002215.

Piccinin, S., Cristi, C., & McCoy, M. (1999). The impact of individual consultation on student ratings of teaching. *The International Journal for Academic Development, 4*(2), 75–88. https://doi.org/10.1080/1360144032000071323.

Prøitz, T. S. (2015). Learning outcomes as a key concept in policy documents throughout policy changes. *Scandinavian Journal of Educational Research, 59*, 275–296.

Raban, C. (2007). Assurance "versus" enhancement: less is more? *Journal of Further and Higher Education, 31*(1), 77–85. https://doi.org/10.1080/03098770601167948.

Ramsden, P. (2003). *Learning to teach in higher education*. London: RoutledgeFalmer.

Richardson, J. T. E. (2005). Instruments for obtaining student feedback: a review of the literature. *Assessment & Evaluation in Higher Education, 30*(4), 387–415. https://doi.org/10.1080/02602930500099193.

Roche, L. A., & Marsh, H. W. (2002). Teaching self-concept in higher education. In *Teacher thinking, beliefs and knowledge in higher education* (pp. 179–218). Dordrecht: Springer.

Ryan, M. (2015). Framing student evaluations of university learning and teaching: discursive strategies and textual outcomes. *Assessment & Evaluation in Higher Education, 40*(8), 1142–1158. https://doi.org/10.1080/02602938.2014.974503 Retrieved from https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1078721&site=ehost-live.

Ryan, K., & Cousins, J. B. (2009). *The SAGE International handbook of educational evaluation*. California: Sage Publications.

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). Los Angeles: Sage.

Smidt, H. (2015). European quality assurance—a European higher education area success story [overview paper]. In A. Curaj, L. Matei, R. Pricopie, J. Salmi, & P. Scott (Eds.), *The European higher education area: Between critical reflections and future policies* (pp. 625–637). Cham: Springer International Publishing.

Spencer, K. J., & Schmelkin, L. P. (2002). Student perspectives on teaching and its evaluation. *Assessment & Evaluation in Higher Education, 27*(5), 397–409. https://doi.org/10.1080/0260293022000009285.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching. *Review of Educational Research, 83*(4), 598–642. https://doi.org/10.3102/0034654313496870.

Stein, S. J., Spiller, D., Terry, S., Harris, T., Deaker, L., & Kennedy, J. (2012). *Unlocking the impact of tertiary teachers' perceptions of student evaluations of teaching*. Wellington: Ako Aotearoa National Centre for Tertiary Teaching Excellence. Retrieved from. https://www.researchgate.net/profile/Dorothy_Spiller2/publication/265439470_Unlocking_the_impact_of_tertiary_teachers'_perceptions_of_student_evaluations_of_teaching/links/54bdbf430cf218d4a16a3331.pdf.

Stein, S. J., Spiller, D., Terry, S., Harris, T., Deaker, L., & Kennedy, J. (2013). Tertiary teachers and student evaluations: never the twain shall meet? *Assessment & Evaluation in Higher Education, 38*(7), 892–904. https://doi.org/10.1080/02602938.2013.767876.

Stensaker, B. (2002). *Strategiske valg og institusjonelle konsekvenser: Organisasjonsutvikling ved Høgskolen i Sør-Trøndelag*. (2002:18) Oslo: NIFU. Retrieved https://nifu.brage.unit.no/nifu-xmlui/bitstream/handle/11250/278218/NIFUskriftserie2002-18.pdf?sequence=1.

Stensaker, B., & Leiber, T. (2015). Assessing the organisational impact of external quality assurance: hypothesising key dimensions and mechanisms. *Quality in Higher Education, 21*(3), 328–342. https://doi.org/10.1080/13538322.2015.1111009.

Steyn, C., Davies, C., & Sambo, A. (2019). Eliciting student feedback for course development: the application of a qualitative course evaluation tool among business research students. *Assessment & Evaluation in Higher Education, 44*(1), 11–24. https://doi.org/10.1080/02602938.2018.1466266.

Universitetet i Tromsø. (2012). *Kvalitetssystem for utdanningsvirksomheten ved Universitetet i Tromsø*. Tromsø: Universitetet i Tromsø Retrieved from https://uit.no/om/enhet/artikkel?p_document_id=167965&p_dimension_id=88126&men=65815.

Uttl, B., & Smibert, D. (2017). Student evaluations of teaching: teaching quantitative courses can be hazardous to one's career. *PeerJ, 5*(5), e3299. https://doi.org/10.7717/peerj.3299.

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation, 54*, 22–42. https://doi.org/10.1016/j.stueduc.2016.08.007.

Williams, J. (2016). Quality assurance and quality enhancement: is there a relationship? *Quality in Higher Education, 22*(2), 97–102. https://doi.org/10.1080/13538322.2016.1227207.

Affiliations

Iris Borch[1] · Ragnhild Sandvoll[1] · Torsten Risør[2,3]

Ragnhild Sandvoll
ragnhild.sandvoll@uit.no

Torsten Risør
torsten.risor@uit.no

[1]   Centre for Teaching, Learning and Technology, UiT The Arctic University of Norway, Postboks 6050 Langnes, N-9037 Tromso, Norway

[2]   The Norwegian Centre for E-Health, UiT The Arctic University of Norway, Tromso, Norway

[3]   Faculty of Health Sciences, UiT The Arctic University of Norway, Tromso, Norway